Ambipolar Transport Compact Models for Two-Dimensional Materials Based Field-Effect Transistors

Zhaoyi Yan, Guangyang Gou, Jie Ren, Fan Wu, Yang Shen, He Tian*, Yi Yang*, and Tian-Ling Ren*

Abstract: Three main ambipolar compact models for Two-Dimensional (2D) materials based Field-Effect Transistors (2D-FETs) are reviewed: (1) Landauer model, (2) 2D Pao-Sah model, and (3) virtual Source Emission-Diffusion (VSED) model. For the Landauer model, the Gauss quadrature method is applied, and it summarizes all kinds of variants, exhibiting its state-of-art. For the 2D Pao-Sah model, the aspects of its theoretical fundamentals are rederived, and the electrostatic potentials of electrons and holes are clarified. A brief development history is compiled for the VSED model. In summary, the Landauer model is naturally appropriate for the ballistic transport of short channels, and the 2D Pao-Sah model is applicable to long-channel devices. By contrast, the VSED model offers a smooth transition between ultimate cases. These three models cover a fairly completed channel length range, which enables researchers to choose the appropriate compact model for their works.

Key words: Field-Effect Transistor (FET); compact model; ambipolar transport; Landauer formula; Pao-Sah model; virtual source

1 Introduction

Ambipolar Two-Dimensional materials based Field-Effect Transistors (2D-FETs), such as Black Phosphorus[1–9] (BP), molybdenum diselenide[10] (MoSe$_2$), and tungsten diselenide[11, 12] (WSe$_2$), have been reported so far. Because of their ambipolar characteristics, 2D-FETs are expected to achieve p-FET and n-FET in the same device, and thus enable various innovative applications, including ambipolar flash memories, artificial synaptic transistors[13], logic devices, and light-emitting transistors[14]. A recent work[15] particularly demonstrated the potential of ambipolar devices in reconfigurable secure circuits, where the ability of *in situ* switching between p-FET and n-FET played an important role. More ambipolar 2D-materials are expected to be revealed, and the wide tunable bandgap range of 2D materials affords further versatile applications (Fig. 1). Here we focus on the compact modeling of these materials to give an appropriate description of their ambipolar drain current. Clearly, it is of fatal importance to their applications in the logic devices in which a clear interpretation of the operational regime of transistors and an efficient computational framework for the integrated circuit design are demanded.

Three laws are mainly adopted for the selections of the reviewed compact models: (1) They should have been benchmarked by experiments. (2) They should be systematically extended. (3) They should not rely on numerically solving differential equation(s). However, in our opinion, compact models can integrate numerically and solve simple transcendental equations.

An extendable model can have several variants rather than being only applicable to isolated cases,

• Zhaoyi Yan, Guangyang Gou, Jie Ren, Fan Wu, and Yang Shen are with the Institute of Microelectronics, Tsinghua University, Beijing 100084, China.

• He Tian, Yi Yang, and Tian-Ling Ren are with the Institute of Microelectronics, and Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: tianhe88@tsinghua.edu.cn; yiyang@tsinghua.edu.cn; RenTL@tsinghua.edu.cn.

∗ To whom correspondence should be addressed.

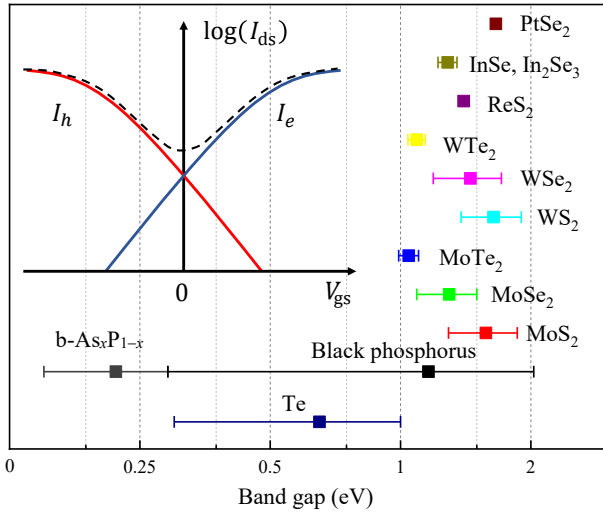 Manuscript received: 2020-11-28; accepted: 2020-12-16

Fig. 1 Tunable bandgap of the ambipolar 2D material family (adapted from Refs. [16] and [17]). The inset shows the typical transfer curve of ambipolar 2D-FETs. Here, I_h and I_e represent electron and hole components of the drain current.

which affords the degree of freedom in choosing ideal assumptions and nonideal corrections to accommodate itself in different conditions. Establishing a compact model for transistors follows a paradigm[18] in which it regularly abstracts the realistic physical factors as effective parameters or an ideal model. For example, the crystalline anisotropy of the channel is reflected by the anisotropic carrier effective masses[19], the layer number-dependent transport characteristic[20] is tackled with the layer-dependent bandgap[21] (and other parameters that will be introduced later), and the carrier energy dispersion is modeled with nearly-free-electrons[22]. However, a compact model cannot always result in a closed-form final expression for current, and indeed, it depends on the specific modeling approach[23]. For instance, for the traditional Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET), the resulting computational complexity varies from the simplest version of the square law approximated from the Pao-Sah double-integral model[24] to the modern complicated Berkeley Short-channel IGFET Model[25] (BSIM) for MOS transistors comprising thousands of parameters. Thus, compact model should not be simply defined as a drain current model that has a friendly analytical form. Otherwise, it is too stringent to include enough representative works.

Applying the laws above, modeling works can be summarized into three types, namely, (1) Landauer formula based model, (2) 2D Pao-Sah model, and (3)

Virtual Source Emission-Diffusion (VSED) model. All of them achieve ambipolar branch (a critical discussion of the term "ambipolar" can be found in Ref. [26], and we follow the conventional nomenclature[27, 28] here) and have been benchmarked by the experimental data. The Landauer method focuses on the junction-limited mechanism at source (drain) Metal/Semiconductor (MS) channel interfaces, whereas the 2D Pao-Sah model is based on the drift-diffusion transport theory. However, the VSED naturally incorporates both effects of Schottky junctions and channel.

To date, these types of modeling methods have formed their own research communities, as reflected on the continuously increasing citations (Fig. 2).

Before we delve into the distinction between the three models, we will present some general remarks on the basic idea of constructing a drain current model for 2D-FETs. First, the raw FET structure is divided into two sub-devices[29]: Part (1) the source-channel-drain resistor, which decides the map $\Gamma: (V_{ds}, \varphi) \mapsto I_{ds}$, i.e., from the drain-source biased voltage V_{ds} and the channel potential φ to the drain current I_{ds}, and part (2) the gate capacitor, which couples the gate-source biased voltage V_{gs} with φ, defined as map $\Lambda: \varphi \mapsto V_{gs}$, as shown in Fig. 3. Thus, formally, the drain current I_{ds} is determined by V_{ds} and V_{gs} as follows:

$$\left. \begin{array}{l} I_{ds} = \Gamma(\varphi, V_{ds}) \\ V_{gs} = \Lambda(\varphi) \end{array} \right\} \Rightarrow I_{ds} = \Gamma\left(\Lambda^{-1}(V_{gs}), V_{ds}\right) \quad (1)$$

As is going to be explained in detail below, for the Landauer formula based compact model, the φ in
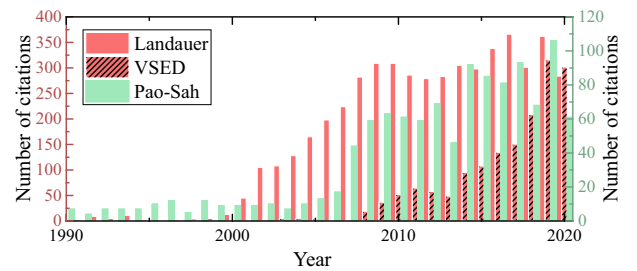

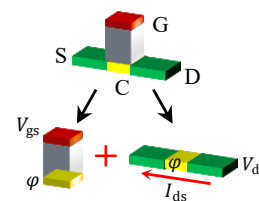**Fig. 2 Publication citations of three types of modeling methods. Source: Web of Science.**

Fig. 3 Device schematic. Here, G: Gate, S: Source, D: Drain, C: Channel.

Formula (1) is more likely (but not necessarily) a scalar number V_ℓ representing the overall channel potential. Thus, map Γ stays as a function, whereas for the 2D Pao-Sah model, φ represents the channel electrostatic potential ψ_s, and it varies with the channel coordinate or quasi-Fermi level, thus making Γ a functional rather than a function.

Model evaluation by the experimental data is usually based on the least squares regression to determine the adjustable model parameters. The residual error function can be expressed as

$$\xi^2\left(a_\mu; V_{ds}\right) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{I_{ds}^{(i)} - \Gamma A_{(i)}^{-1}}{I_{ds}^{(i)}}\right|^2 \tag{2}$$

where a_μ, $I_{ds}^{(i)}$, and N are model parameters, sampled current data points, and sampled point number, respectively. It should be noted that the loss function is normalized for $I_{ds}^{(i)}$ before to be summed over the sampling data points in Eq. (2). It is necessary, because of the current magnitude range over 3–4 orders (usually even more). It is not surprising that with enough model parameters, good fitting always can be acquired. However, plethora of adjustable parameters can spoil all the benefits associated with the improved model accuracy, because the overfitting problem, in this case, is nearly inevitable.

As for the organization of the chosen materials, the general architecture of the compact model will be systematically elucidated, and the subtle idea beneath the model that produces the ambipolar branch and the computational complexity of achieving the model simulation coming up with the idea will be examined in details. We adopt the top-down manner by exhibiting the most general architecture of the theory at first and then digging into other variants of the model or the remarkable approximation proposed by the researchers. Hence, both the general treatments and special tricks are involved.

Nonetheless, these limited discussions are far from the exhausted classifications on this topic, and some interesting works are clearly unable to be attributed to certain classes that we present here or a simple combination of them. However, we manage to exhibit the most intrinsic structures of the three types of models, in order to lay a basis for understanding the models, which cannot be covered by this review.

Sections 2–4 discuss the Landauer model, Pao-Sah model, and VSED model, respectively. Then we conclude the review with a comparison of these methods in Section 5.

2 Landauer Model

Typical Landauer formula based models can be found in Refs. [30, 31]. Although it was initially thought to be rigorous in physics to describe the ballistic transport, it turns out to be applicable to the diffusive regime and can acquire a satisfactory agreement with experimental data. It can be exploited as a kind of a lumped model, in which the device's internal parameter channel potential distribution ϕ is represented by a single value V_ℓ and its interactions with the external biases, i.e., V_{gs} and $-V_{ds} < 0$, are transmitted by the lumped model as introduced below. The advantage of this approach is that it requires only the attention on the MS junction formed at the source/drain, thus being practical and efficient for the quantitative analysis.

2.1 Definition and typical result of the model

The device structure is usually composed of source/drain reservoirs, channel region, and serial connection of the quantum capacitor C_q and gate oxide capacitor C_{ox} as shown in Fig. 4a. The channel electrical potential V_ℓ is assumed to be homogeneous throughout the channel and related to the effective gate voltage bias $V_{gs} - V_{fb}$ by the serial capacitor's formula[31], that is, V_ℓ and V_{gs} are connected with each other by Eqs. (3)–(6),

$$V_\ell = \left(V_{gs} - V_{fb}\right) C_{ox}/\left(C_q + C_{ox}\right) \tag{3}$$

$$C_{ox} = \varepsilon_{ox}/t_{ox} \tag{4}$$

$$C_q = \int_{-\infty}^{+\infty} \frac{qD\left(E\right)}{4V_T}\mathrm{sech}^2\left(\frac{E/q - V_\ell}{2V_T}\right)\mathrm{d}E \tag{5}$$

$$V_{fb} = \frac{\phi_m - \phi_{BP}}{q} - \frac{q}{C_{ox}}N_f -$$
$$\frac{q}{C_{ox}}\int_{-\infty}^{+\infty}\frac{D_{it,a}}{1 + \exp\left[(E + qV_\ell)/qV_T\right]}\mathrm{d}E \,+$$
$$\frac{q}{C_{ox}}\int_{-\infty}^{+\infty}\frac{D_{it,d}}{1 + \exp\left[-(E + qV_\ell)/qV_T\right]}\mathrm{d}E \tag{6}$$

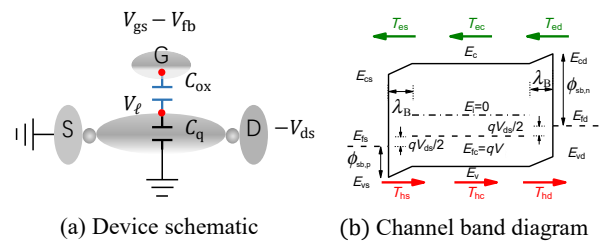(a) Device schematic   (b) Channel band diagram

Fig. 4   Device model and channel band diagram.

where

$$D(E) = \begin{cases} m_h^*/(\pi\hbar^2), & E \leqslant E_v; \\ 0, & E_v < E \leqslant E_c; \\ m_e^*/(\pi\hbar^2), & E \geqslant E_c \end{cases} \quad (7)$$

$$D_{\text{it,a(d)}} = \frac{N_{\text{it,a(d)}}}{\sqrt{2\pi}\sigma_{\text{a(d)}}} e^{-\frac{\left(E - E_{c(v)}\right)^2}{2\sigma_{\text{a(d)}}^2}} \quad (8)$$

where $\hbar$ is reduced Planck constant, V_T denotes the thermal voltage equivalent, $\phi_{\text{m(BP)}}$ is the work function of the gate metal (BP), C_{ox} is the oxide capacitance per unit area, t_{ox} is the thickness of the oxide, ε_{ox} is the permittivity of the oxide layer, $D(E)$ is the number of states per unit area per unit energy of the 2D periodic system with nearly free electrons approximation, $m_{e(h)}^*$ is the effective mass of the electrons and holes, $E_{c(v)}$ is the minimum (maximum) of the conduction (valence) band, N_f is the fixed charge density in the oxide layer per unit area, $D_{\text{it,a(d)}}$ is the interface acceptors (donors) trapped in the occupied density of states per unit area per unit energy, $N_{\text{it,a(d)}}$ is the total amount of $D_{\text{it,a(d)}}$, and $\sigma_{\text{a(d)}}$ is the standard deviation of $D_{\text{it,a(d)}}$.

To address the 2D-material channel, 2D quantum capacitance C_q is exploited. Particularly, the integral of Eq. (5) has a closed form because of the piecewise energy-independent 2D density of states, as shown in Formula (7).

Equation (6) shows that the ingredients of V_{fb} contain not only the fixed charge of the oxide layer and work function differences, but also channel potential-dependent factors[32], i.e., the interface trap charge in the channel as depicted in Eq. (8), which has been assumed to obey the Gaussian distribution centered at $E_{c(v)}$. The existence of the channel potential-dependent component entails the numerical solution of $V_\ell - V_{\text{gs}}$ in Eq. (3). The flat-band voltage V_{fb} plays a role in aligning the minimum drain current points predicted by the model to that of the benchmark, which might be experimental data or other advanced numerical results.

Based on the channel potential V_ℓ, the vertex of the energy band diagram as shown in Fig. 4b, can be determined as follows:

$$E_{\text{fs}} = -qV_\ell - (1 - \alpha) \cdot qV_{\text{ds}} \quad (9)$$

$$E_{\text{fd}} = -qV_\ell + \alpha \cdot qV_{\text{ds}} \quad (10)$$

$$E_{c(v),s} = E_{\text{fs}} \pm \phi_{\text{sb,n(p)}} \quad (11)$$

$$E_{c(v),d} = E_{\text{fd}} \pm \phi_{\text{sb,n(p)}} \quad (12)$$

where α is the distribution ratio ranging from 0 to 1 of the drain-source bias voltage V_{ds}. Because of the

weak switch-on state, assigning a constant to define the partition of the voltage drop makes sense. For example, $\alpha = 1$ in Ref. [33], whereas it was set to be 0.5 in the Ref. [34]. The saturation characteristic of the output curves of the two schemes differs[35]. $\phi_{\text{sb,n(p)}}$ is the Schottky barrier for electrons (holes), $E_{\text{fs(d)}}$ is the Fermi level of the source (drain) metal, $E_{c(v),s}$ is the conduction (valence) band minimum (maximum) at the channel/source electrode interface, and $E_{c(v),d}$ is that of the drain. Here the symmetry source and drain metals are assumed, but the generalization of the asymmetry case is straightforward. λ_B, the width of the barrier formed at the channel/source (drain) interface is approximated using the screening length as[36]

$$\lambda_B = \sqrt{t_{\text{ch}}t_{\text{ox}}} \cdot \left(\frac{\varepsilon_{\text{ox}}}{\varepsilon_{\text{ch}}} + \frac{t_{\text{ox}}}{t_{\text{ch}}}\right)^{-1/2} \quad (13)$$

where ε_{ch} and t_{ch} are the permittivity and thickness of the channel, respectively. The correctness of λ_B evaluated with Eq. (13) should be noted because the expression of the screen length depends on the boundary conditions adopted[36] (electrical-potential-type boundary conditions are adopted here as an example, but the use of the electrical-field-type boundary conditions in modeling the screen length can be found in Refs. [30, 33, 37, 38]).

Then the drain current can be computed by the Landauer formula as follows:

$$I_{\text{ds}} = \frac{2qW}{h} \int_{-\infty}^{+\infty} T(E) M(E) [f_d - f_s] \, dE \quad (14)$$

where h is Planck constant, W is the channel width, $f_{\text{s(d)}}(E)$ represents the Fermi-Dirac distribution of the carrier at the source (drain) electrode, $M(E)$ is the number of modes per unit width[35], and $T(E)$ is the transmission coefficients[39] calculated with the WKB approximation[30, 33, 34, 37, 38]. Specifically, we have

$$f_{\text{s(d)}} = \frac{1}{1 + \exp\left(\frac{E - E_{\text{fs(d)}}}{qV_T}\right)} \quad (15)$$

$$M(E) = g_v \cdot \frac{\theta\left[\pm\left(E - E_{c(v)}\right)\right]}{\frac{1}{2}h/\sqrt{2m_{c(v)}^*|E - E_{c(v)}|}} \quad (16)$$

$$T(E) = T_{\text{e(h)}}(E)\,\theta\left[\pm\left(E - E_{c(v)}\right)\right] \quad (17)$$

$$T_{\text{e(h)}}^{-1} = T_{\text{e(h),s}}^{-1} + T_{\text{e(h),c}}^{-1} + T_{\text{e(h),d}}^{-1} - 2 \quad (18)$$

$$T_{\text{e(h),s}}^{-1} = \exp\left[\frac{4\sqrt{2m_{\text{e(h)}}^*}}{3\hbar/\lambda_B} \cdot \frac{\text{Re}\sqrt{\left(\pm\left(E_{c(v)s} - E\right)\right)^3}}{\pm\left(E_{c(v)s} - E_{c(v)}\right)}\right] \quad (19)$$

$$T_{e(h),d}^{-1} = \exp\left[\frac{4\sqrt{2m_{e(h)}^*}}{3\hbar/\lambda_B} \cdot \frac{\mathrm{Re}\sqrt{\left(\pm\left(E_{c(v)d}-E\right)\right)^3}}{\pm\left(E_{c(v)d}-E_{c(v)}\right)}\right] \tag{20}$$

$$T_{e(h),c}^{-1} = 1 + L/\lambda_{e(h)}(E) \tag{21}$$

where $\theta[\cdot]$ denotes the Heaviside step function; L is the channel length, g_v is the valley degeneracy, which is regularly set as 2; and $\lambda_{e(h)}(E)$ is the energy-dependent mean free path. It is assumed that $\lambda_{e(h)}(E)$ consists of two components expressed as[29, 34, 40]

$$\frac{1}{\lambda_{e(h)}} = \frac{N_{it}}{\lambda_{0,e(h)}}\left(\frac{|E-E_{c(v)}|}{qV_T}\right)^{-\gamma_0} +$$

$$\frac{1}{\lambda_{1,e(h)}}\left(\frac{|E-E_{c(v)}|}{qV_T}\right)^{-\gamma_1} \tag{22}$$

where N_{it} denotes the density of the ionized trap. Thus,

$$N_{it} = \int_{-\infty}^{+\infty} \frac{D_{it,a}}{1+\exp\left[(E+qV_\ell)/qV_T\right]}\mathrm{d}E +$$

$$\int_{-\infty}^{+\infty} \frac{D_{it,d}}{1+\exp\left[-(E+qV_\ell)/qV_T\right]}\mathrm{d}E \tag{23}$$

The first term on the right-hand side of Eq. (22) with $\gamma_0 = 3/2$ represents the contribution from the charged impurity scattering and $\gamma_1 = 1/2$ represents that from the phonon scattering. The adjustable parameters $\lambda_{0,e(h)}$ and $\lambda_{1,e(h)}$ are determined from the least squares method as mentioned before. In addition, the right-hand side of Eqs. (16) and (17) both require a summation over carrier types.

Equations (3)–(22) define a complete compact model, and the transfer and output curves of 2D-FETs can be computed once the parameters are specified. The typical transfer curves and variation of the band diagram are shown in Figs. 5a and 5b, respectively. The current component of thermal emission electrons (holes) is defined by $T_{e(h)} = 1$, which is opposite to the tunneling component, in which the tunneling barrier is non-trivial. On regions I and III, as is depicted in Figs. 5a and 5b, the channel potential V_ℓ is pushed close to the edge of the band when the gate voltage is strong enough. Thus, the dominant carrier tunneling component is significant. On region II, the thermal emission
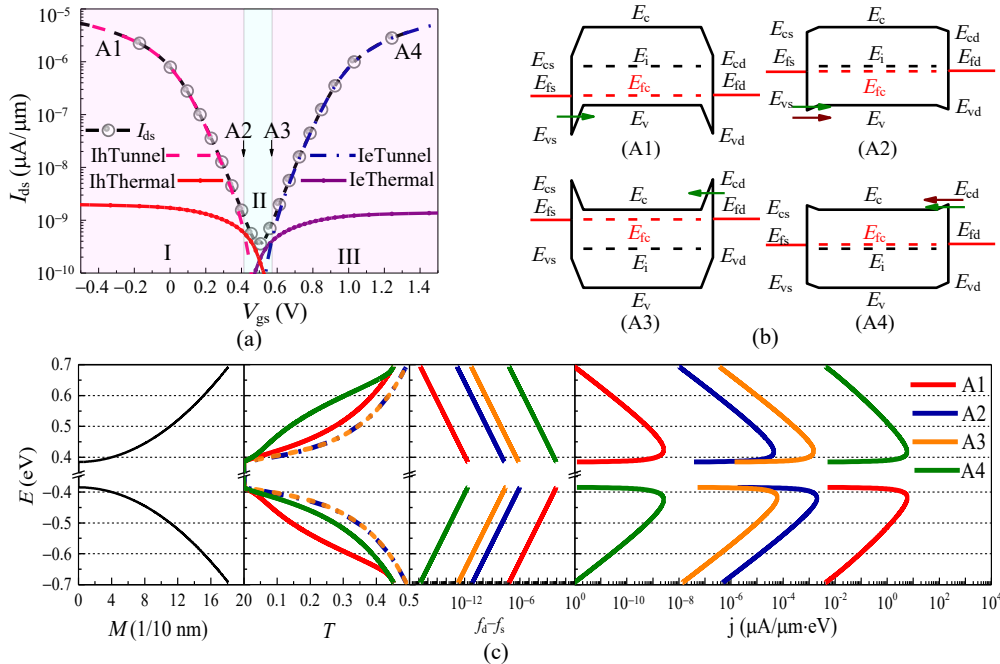

Fig. 5 (a) Transfer curve, (b) variation of the band diagram, and (c) product factor distributions in the energy scalar. Parameters are set as $\phi_{sb,n(p)} = 0.5E_g = 0.39$ eV, $m_{e(h)}^* = m_0$, $N_{it,e(h)} = 8.8\times10^{12}$ cm$^{-2}\cdot$eV^{-1}, $\sigma_{a(d)} = 10^{12}$, $\lambda_{i,e(h)} = 300$ nm, $\varepsilon_{ox} = 3.9\varepsilon_0$, $\varepsilon_{ch} = 8.3\varepsilon_0$, $t_{ox} = t_{ch} = 2$ nm, $L = 10^3$ nm, $V_T = 0.026$ V, and $V_{ds} = 10$ mV. For (a), based on the model, four current components, namely, electron thermal emission ("IeThermal", purple solid), electron tunneling ("IeTunnel", blue dashed lines), hole thermal emission ("IhThermal", red solid), and hole tunneling ("IhTunnel", pink dashed lines), are recognized from the total current (I_{ds}, circle-marked black dashed lines). The voltage regime bounded by letters "A1" and "A2" denotes holes' dominant branch (I); letters "A3" and "A4" denote the electrons' dominant branch (III), and letters "A2" and "A3" denote the conversion interval (II). The four band diagrams corresponding to "A1" – "A4" are shown in (b), where the thermal emission mechanism and barrier tunneling are represented by the brown curved arrows and green straight arrows, respectively. The reference point of energy is set as E_i. The energy distribution of $M(E)$, $T(E)$, $f_d(E) - f_s(E)$, and $j(E)$ at "A1"–"A4" is illustrated in (c).

component dominated for the tunneling of the carrier is nearly banned. To give an intuitive impression of the energy distribution of the product factors of the current density in Eq. (14), $T(E)$, $M(E)$, $f_d(E) - f_s(E)$, and the current density $j(E)$ are plotted in Fig. 5c. When the energy of the carrier becomes more remote from the band edge, $M(E)$ and $T(E)$ increase, whereas the Fermi window $f_d(E) - f_s(E)$ exponentially decreases, so the peak of $j(E)$ is clamped near the band edge.

It particular, the Landauer formula itself can be the basis of the tight-binding transportation calculation method[41]. One of the major differences of the compact models reflects on the transmission modeling. Thus, the success in the compact model based on Landauer theory shows that tunneling transmission is approximated reasonably. Moreover, the computational framework holds only for the small V_{ds}, so the quasi-Fermi level in the channel does not deviate much from the equilibrium location, because the approximated Landauer theory applies only to nearly equilibrium transportation.

2.2 Numerical recipe

Typically, there are two scenarios where the efficient numerical algorithm of the model is required to minimize the time consumption of each calculation: (1) planting the model into transistor-level circuits' simulation program for industrial application, and (2) extraction of adjustable parameters by least squares regression defined as Eq. (2). Except for parameters, such as physical constants, environment temperature, and measurable geometry quantities, which can be clearly known, the parameters E_g, m_c^*, m_v^*, $\varphi_{sb,n}$, N_{it}, N_f, $\sigma_{a(d)}$, $\lambda_{0,e(h)}$, and $\lambda_{1,e(h)}$ are relatively tough to decide. Ideally, only the last three parameters of the model are the adjustable parameters that are fitted by the model, and the rests possess specific physical meanings; however, difficulties exist. The bandgap E_g and effective masses $m_{e(h)}^*$ are tunable by several factors, such as the layer of BP[19, 42], impurity defect[43], and even electrostatic field[44]. Thus, except for *in situ* measurements, other data sources are susceptible. The accurate value of the Schottky barrier, which is determined from an independent experiment or theoretical calculation, will fade into insignificance compared with the Fermi-level's pining effect existing in reality[45–47]. The value of N_{it} and N_f also suffer from device-to-device variations. All these parameters could be extracted from the experimental current-voltage data in principle if the number of sample points is far more than that of the fitted parameters. As for the

technique aspects, employing the advanced optimization methods to minimized ξ defined by Eq. (2) is feasible, such as simplex algorithm[48], which is insensitive to the non-differentiable drawback of the model. Clearly, the generalized integrals required by Eqs. (14) and (16) are time-consuming. Provided that the speed of calculating the integral improved by N' times, it will cost $(k+1)N'N$ less time to calculate ξ once, where k denotes the iteration time necessary for the self-consistently solution of Eq. (3). Thus, there is a great potential to accelerate the model calculation.

A numerical method to accelerate the integral speed is to first substitute the Fermi-Dirac distribution by Boltzmann distribution, which holds for the non-degenerated case. Then the $V_\ell - V_{gs}$ equation-related integral Eq. (6) and the drain current integral Eq. (14) can be evaluated by the Gauss-Hermite and Gauss-Laguerre quadrature technique[49], respectively. Compared to the built-in standard function of commercial software, such as MATLAB, the trick of employing Gauss quadrature can enhance the computation speed by nearly two magnitudes with a negligible relative error to MATLAB's built-in function "integral", as indicated in Fig. 6, which takes Gauss-Hermite quadrature as an example.

However, the plethora of adjustable parameters improves the fitting accuracy of the model at the risk of overfitting.

2.3 Basic blocks of the model

There are at least seven degrees of freedom to construct a BP-compact model. Besides (1) the divide-ratio α of V_{ds} and (2) boundary conditions of barrier width λ_B, which have been mentioned before, the rest five are analyzed
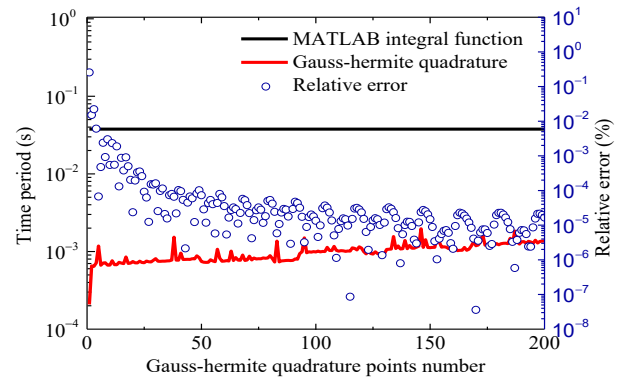

Fig. 6 Time period and relative error with increasing Gauss-Hermite quadrature points number. Configuration of the platform: Intel(R) Core (TM) i5-4210U CPU (1.7 GHz × 4) processor with 8 GB of RAM.

as follows:

The term "mode" $M(E)$ has a homology with the concept of the mode number per unit width in a 2D waveguide[35], equal to the reciprocal of the de Broglie half-wavelength (up to a valley degeneracy constant g_v), which can be calculated at a given energy level E, as shown in Eq. (16). Thus, it is an energy dependent quantity in contrast to the 2D density of states, which is energy independent, and more importantly, it depends on the specific energy dispersion relation selected to describe the carrier in BP.

The total transmission coefficient $T(E)$ is recognized as the serial connection[39] of three components, i.e., source/channel interface barrier, channel scattering region, and drain/channel interface barrier. Tunneling through the forbidden gap (i.e., E falls in the energy gap) is banned, as shown in Eqs. (17) and (18). The coefficients at the interface barrier are evaluated using the semiclassical WKB approximations, shown by Eqs. (19) and (20). However, the approximation only holds[50] at the short-wavelength limit and fails at the turning point, i.e., $E = V(x)$, where

$$\lambda_{\text{de Broglie}} = \frac{h}{\sqrt{2m^*_{c(v)}|E - V(x)|}} \to \infty,$$

such that the electron entirely behaves like matter-waves rather than a localized particle and the semiclassical picture breaks down. Improvement can be achieved by employing the Airy function and transfering matrixes to get rid of the deficiency of the semiclassical picture[51, 52].

Besides the WKB approximation, Eqs. (19) and (20) rely on the assumptions that (1) the band bending of $E_{c(v)}$ at the interface barrier linearly varies with the distance, i.e., $\Delta E \propto \Delta x$; (2) the band dispersion is approximated by that of the nearly free electrons, i.e., parabolic dispersion relation; and (3) the quasi-Fermi level at the drain (source) electrode should not be higher (lower) than the $E_{c(v)}$ in the channel, so that no band-to-band tunneling happens, which is nearly E-independent. However, Ref. [37] adopted the distinct assumptions of an elliptic complex band structure[53] and exponential screening potential[36] (i.e., $\propto \exp(-x/\lambda_B)$), which is very similar to the Yukawa screening potential[54] in particle physics except for the Coulomb-interaction term. Reference [30] explored the involvement of band-to-band tunneling into the transmission expression, while keeping the other assumptions the same with Eqs. (19) and (20). However, Ref. [38] adopted the

exponential screening potential and included the band-to-band tunneling mechanism.

As for the channel scattering, Eq. (21) can be found in Ref. [39] and is clearly the main symbol to extend the Landauer formula into the diffusive regime. For the subthreshold case, such as the Ref. [33], the channel scattering term can be omitted, thus the total transmission reads as

$$\frac{1}{T} = \frac{1}{T_s} + \frac{1}{T_d} - 1 \tag{24}$$

Thus, at present, building a compact model for 2D-FETs based on the Landauer formula is partly similar to building Lego blocks, as shown in Fig. 7.

Based on the discussions above, seven categories of building blocks are used to define a model: (1) "V_{ds}" blocks decide how to divide V_{ds} between the source and drain Fermi level; (2) "λ_B" blocks decide the kind of boundary condition used to determine the interface barrier width; (3) "T_B" blocks indicate which method between WKB approximation and Airy function and related transfer matrix will be employed to calculate the transmission coefficients of barrier tunneling; (4) "B.B." blocks distinguish different methods of band bending; (5) "$E(k)$" blocks give the energy dispersion relation used in $M(E)$ and $T(E)$; (6) "BtBT" blocks specify whether the band-to-band tunneling is enabled; and (7) "T_ℓ" blocks decide if channel scattering counts in the $T(E)$.

All possible combinations of the blocks have been far from being exhausted; more importantly, symbols "A&T" and "ψ" defined in Fig. 8 are not included in Fig. 7, which means that almost all of the published works at present rely on the field-type boundary conditions and WKB approximation. Thus, there is still enough room for enhancement in the Landauer-based

Ref.	V_{ds}	λ_B	T_B	B.B.	$E(k)$	BtBT	T_ℓ
[30]	1	ε	W	x	P	BtB	Bal.
[33, 55]	1	ε	W	x	$P\&E$	noB	Bal.
[37]	1	ε	W	exp	E	noB	Bal.
[34]	1/2	ε	W	x	P	noB	Dif.
[38]	1	ε	W	exp	P	BtB	Bal.

Fig. 7　Structures of BP-compact models based on the Landauer formula constituted by distinct building blocks. The symbol designations are referred to Fig. 8.

1	$\alpha = 1$	1/2	$\alpha = 1.2$
ε	Field-type B.C.	ψ	Potential-type B.C.
W	WKB approx.	A&T	Airy function
x	Linear B.B.	exp	Exponential B.B.
P	Parabolic dispersion	E	Ellipse dispersion
BtB	BTBT is allowed	noB	BTBT is banned
Bal.	C.S. neglected	Dif.	C. S. involved

Fig. 8 Symbol designation in Fig. 7. The following abbreviations are used: B.C.= Boundary Condition, B.B.= Band Bending, and C.S.= Channel Scattering.

compact model that 2D-FETs can be achieved.

3 Pao-Sah Model

3.1 General framework for the 2D Pao-Sah model

Recently published works[55–60] have drawn our attention back to the traditional Pao-Sah model[24, 61, 62]. As is well known, its original form cannot be directly applied to simulate the ambipolar branch case of 2D-FETs. Inpired by these works, we first review the general ambipolar transport theory based on the drift-diffusion equation, and then derive the 2D Pao-Sah model proposed by Ref. [58] in a more reasonable way.

The device schematic and channel band diagram are shown in Fig. 9, and for the MOS band diagram, referring to the classical work by Sah[62] is recommended. Particularly, the voltage reference of all kinds of absolute energies/voltages should be pointed out. Here, the reference point is selected as the source metal Fermi energy level E_{fs}, which is the same with the channel Fermi energy E_{fc}^0 when $V_{\text{ds}} = 0$. This convention is necessary for MS-junction-based 2D-FETs.

The drift-diffusion net electrical drain current is described in Eqs. (25)–(28):

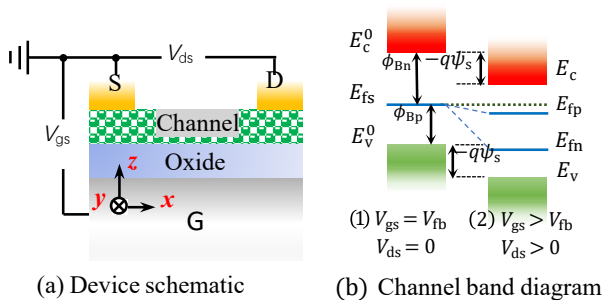$$j_{\text{e(h)}} = -\mu_{\text{e(h)}} n_{\text{e(h)}} \nabla E_{\text{fn(p)}} \quad (25)$$

(a) Device schematic   (b) Channel band diagram

Fig. 9 Schematic of the 2D-FET and channel band diagram.

$$\nabla (j_{\text{e}} + j_{\text{h}}) = 0 \quad (26)$$

$$I_{\text{ds}} = -\frac{W}{L} \int_0^{V_{\text{ds}}} \mu_{\text{e}} n_{\text{e}} dE_{\text{fn}} + \mu_{\text{h}} n_{\text{h}} dE_{\text{fp}} \quad (27)$$

$$V_{\text{fb}} + V_{\text{ox}} + \psi_{\text{s}} = V_{\text{gs}} \quad (28)$$

In Eq. (25), $n_{\text{e(h)}}$ denotes the electron (hole) density, $E_{\text{fn(p)}} = -qV_{\text{n(p)}}$, and $j_{\text{e(h)}}$ is the current density of electrons (holes) in the plane (D→S) is set as the positive direction for the current, and it is the reverse for the coordinate). Although it seems more like a drift current expression, the drift and diffusion components are captured via the gradient of the quasi-Fermi level[24], with a generalized Einstein relation[63]. Equation (26) indicates that only steady current is considered here, and neither net generation nor recombination remains, so the current was kept constant through the channel.

Equation (27) shows an integral technique of exploiting the continuity equation to derive the expression for the drain current without being caught in the detailed calculations of the spatial distribution of $j_{\text{e(h)}}$, which finally transforms the drain current integral with the coordinate to that with the quasi-Fermi level. Particularly, Eq. (27) has identified $\mu_{\text{e(h)}}$ as the average surface mobility[62, 64–66], thus being independent with the spatial coordinate, and $n_{\text{e(h)}}$ is described by the Fermi-Dirac distribution under the control of the difference of $E_{\text{fn(p)}}$, i.e., $-qV_{\text{n(p)}}$, and the channel surface electrostatic potential barrier $-q\psi_{\text{s}}$, where the subscript letter "s" denotes the first letter of the word "surface" originating from the inherited concept of the bulk material. Equation (28), the electrostatic balance equation through the gate MOS capacitor[58, 62], is introduced to determine the formulas of ψ_{s} and $E_{\text{fn(p)}}$, which are comprised of four parts: gate-source voltage V_{gs}, flat-band voltage V_{fb}, oxide voltage drop V_{ox}, and surface potential ψ_{s}. Then the four components of Eq. (28) are analyzed as follows.

(1) Gate-source voltage V_{gs} manifests itself in the diagram as the difference of the metal Fermi level E_{fm} and source Fermi level E_{fs}, i.e., $V_{\text{gs}} = (E_{\text{fm}} - E_{\text{fs}})/(-q)$.

(2) Flat-band voltage V_{fb} is set as the difference between the work functions of the gate metal and BP, i.e., $V_{\text{fb}} = (\phi_{\text{m}} - \phi_{\text{BP}})/q$.

(3) Surface potential energy is defined as $\psi_{\text{s}} = (E_{\text{c}} - E_{\text{c}}^0)/(-q)$, where $E_{\text{c}}^0 = E_{\text{fc}}^0 + \phi_{\text{Bn}} = E_{\text{fs}} + \phi_{\text{Bn}} = \phi_{\text{Bn}}$. ϕ_{Bn} is the electrons' Fermi potential barrier. Its counterpart $\phi_{\text{Bp}} = E_{\text{g}} - \phi_{\text{Bn}}$ is the holes' Fermi potential barrier. For convenience, $q\psi_{\text{Bn(p)}} \equiv \phi_{\text{Bn(p)}}$ is assumed. E_{c}^0 and E_{fc}^0 represent the equilibrium location of the conduction band edge and Fermi energy of

the BP-channel, respectively. Once the drain-source electrodes are biased, E_{fc}^0 splits into E_{fn} and E_{fp}; thus $\pm(E_{\text{fn(p)}} - E_{\text{c(v)}}) = \pm E_{\text{fn(p)}} \pm q\psi_{\text{s}} - \phi_{\text{Bn(p)}}$ measures the electron (hole) density and indicates a degenerate transport system if it is positive.

(4) $V_{\text{ox}} = Q_{\text{s}}/C_{\text{ox}}$, where Q_{s} is the areal charge density accumulating on the oxide/semiconductor interface. Two kinds of routines can exist to tackle Q_{s}. If one treats the 2D-channel as the ultrathin bulk material, then the integral of the bulk charge density ρ_{b} built-up with Poisson's equation and Gauss's law along the depth-direction is necessary to obtain the density $Q_{\text{s}}^{\text{bulk}}$ as indicated in Eq. (29)[24, 67]:

$$Q_{\text{s}}^{\text{bulk}} = \text{sign}(\psi_{\text{s}})\sqrt{2\epsilon_{\text{s}}}\left|\int_0^{\psi_{\text{s}}}\rho_{\text{b}}\,\mathrm{d}\psi\right|^{\frac{1}{2}} \quad (29)$$

where the coordinate variable is substituted with the electrostatic potential ψ along the depth-direction. On the contrary, if the BP-channel is purely assumed as a 2D-system where the density of states with effective mass and parabolic dispersion approximation is just constant as depicted in Eq. (30), then no integral along the third dimension exists and the Fermi-Dirac integral $F_j(\xi)$, i.e., Eq. (31) can benefit from the exponential zero condition ($j = 0$), resulting in a closed form expression[68–72] for the areal density of the carrier $Q_{\text{e(h)}}^{\text{2D}} = \pm q\,n_{\text{e(h)}}$, as shown in Eq. (32), which is different with Eq. (29) where an analytical expression could only be obtained unless a Boltzmann distribution approximation is used and inconsistent boundary condition is tolerated[24, 67]. Thus, the total density of charge is $Q_{\text{s}}^{\text{2D}} = Q_{\text{e(h)}}^{\text{2D}} + Q_{\text{e(h)}}^{\text{2D}} + qN_- - qN_+$, where $N_{-(+)}$ is the ionized charge from the completely ionized doped donors (acceptors), which will be omitted in the following discussion, because they both can be regarded as a part of the flat-band voltage.

$$D_{\text{e(h)}} = m_{\text{e(h)}}^*/\left(\pi\hbar^2\right) \quad (30)$$

$$n_{\text{e(h)}} = \int_{E_{\text{c(v)}}}^{\pm\infty} \frac{\pm D_{\text{e(h)}}}{\mathrm{e}^{\pm\frac{E - E_{\text{fn(p)}}}{kT}} + 1}\,\mathrm{d}E \quad (31)$$

$$Q_{\text{e(h)}}^{\text{2D}} = \pm qkT D_{\text{e(h)}}\ln\left(1 + \mathrm{e}^{\pm\frac{E_{\text{fn(p)}} - E_{\text{c(v)}}}{kT}}\right) \quad (32)$$

Because of the inherent simplicity (analytical and strict Fermi-Dirac distribution) and the reasonable physical picture, Eqs. (30)–(32) are frequently exploited at present in the compact modeling research. Thus, we focus on the 2D-description of Q_{s}. An intriguing history of Eq. (32) is that it initially served as a smooth modelling function that adjoints the strong inversion status and subthreshold regime in bulk materials[73].

The proponents of the equation may not have realized that their mathematical modeling trick will become a physical fact decades later[74].

The regular treatment on the relation of ψ_{s} and $E_{\text{fn(p)}}$ applies with unipolar conduction FETs, where only one type of carrier's quasi-Fermi level, i.e., E_{fn} or E_{fp} is assumed to vary in the channel and the other type of carrier's Fermi level is considered to be constant throughout the channel. By assuming that the ambipolar transfer curve can be reproduced by the compact model, both variations of E_{fn} and E_{fp} should be involved.

Considering the resulting equation of ψ_{s}, E_{fn}, and E_{fp} from Eq. (28), with V_{gs} taken as the parameter,

$$n_{\text{e}}(\psi_{\text{s}}, E_{\text{fn}}) - n_{\text{h}}(\psi_{\text{s}}, E_{\text{fp}}) = \frac{C_{\text{ox}}}{q}\left(V_{\text{gs}} - V_{\text{fb}} - \psi_{\text{s}}\right) \quad (33)$$

Equation (33) indicates an implicit function of $\psi_{\text{s}} = \Psi(V_{\text{n}}, V_{\text{p}}; V_{\text{gs}})$, or a suppression of V_{gs} as $\psi_{\text{s}} = \Psi(V_{\text{n}}, V_{\text{p}})$ so that no confusion is caused.

3.2 Compact model

We rearrange Eq. (33) as follows:

$$V_{\text{n(p)}} = \psi_{\text{s}} \mp \psi_{\text{Bn(p)}} \mp \\ V_T\ln\left[\gamma_{\text{e(h)}}\left(V_{\text{p(n)}}\right)\mathrm{e}^{\pm\frac{V_{\text{gs}} - V_{\text{fb}} - \psi_{\text{s}}}{\zeta_{\text{e(h)}}}} - 1\right] \quad (34)$$

where $\zeta_{\text{e(h)}} = m_{\text{e(h)}}\zeta_0$,

$$\gamma_{\text{e(h)}}(x) = \left(\mathrm{e}^{\frac{-\psi_{\text{Bp(n)}} \mp \psi_{\text{s}} \pm x}{V_T}} + 1\right)^{m_{\text{h(e)}}/m_{\text{e(h)}}} \quad (35)$$

Moreover, when $V_{\text{p}} = 0$, and $0 < V_{\text{n}} < V_{\text{ds}}$,

$$V_{\text{n}} = \psi_{\text{s}} - \psi_{\text{Bn}} - V_T\ln\left[\gamma_{\text{e}}(x)\mathrm{e}^{+\frac{V_{\text{gs}} - V_{\text{fb}} - \psi_{\text{s}}}{\zeta_{\text{e}}}} - 1\right] \quad (36)$$

whereas, when $V_{\text{n}} = V_{\text{ds}}$, and $0 < V_{\text{p}} < V_{\text{ds}}$,

$$V_{\text{p}} = \psi_{\text{s}} + \psi_{\text{Bp}} + V_T\ln\left[\gamma_{\text{h}}(x)\mathrm{e}^{-\frac{V_{\text{gs}} - V_{\text{fb}} - \psi_{\text{s}}}{\zeta_{\text{h}}}} - 1\right] \quad (37)$$

Clearly, Eqs. (36) and (37) enable us to numerically compute the drain current. However, we take some simplification here instead of laboring in the time-consuming integral computing. Because the Boltzmann approximation should be applied, the value of $\exp(-\psi_{\text{Bn(p)}} \mp \psi_{\text{s}} \pm V_{\text{p(n)}}/V_T)$ is certainly not much bigger than one. Thus, Eq. (34) can be properly approximated as

$$V_{\text{n(p)}} = \psi_{\text{s}} \mp \psi_{\text{Bn(p)}} \mp V_T\ln\left[\mathrm{e}^{\pm\frac{V_{\text{gs}} - V_{\text{fb}} - \psi_{\text{s}}}{\zeta_{\text{e(h)}}}} - 1\right] \quad (38)$$

Then Eqs. (36) and (37) follow the approximation. When $V_{\text{p}} = 0$, and $0 < V_{\text{n}} < V_{\text{ds}}$,

$$V_{\text{n}} = \psi_{\text{s}} - \psi_{\text{Bn}} - V_T\ln\left[\mathrm{e}^{\frac{V_{\text{gs}} - V_{\text{fb}} - \psi_{\text{s}}}{\zeta_{\text{e}}}} - 1\right] \quad (39)$$

whereas when $V_n = V_{ds}$, and $0 < V_p < V_{ds}$,

$$V_p = \psi_s + \psi_{Bp} + V_T \ln \left[e^{\frac{V_{gs} - V_{fb} - \psi_s}{-\zeta_h}} - 1 \right] \quad (40)$$

Equations (39) and (40) lead to the analytical expression of electron and hole component of the ambipolar current. Starting from Eq. (38), we have

$$\pm q n_{e(h)} = C_{ox} \left(V_{gs} - V_{fb} - \psi_{s,e(h)} \right) \quad (41)$$

Thus, the right-hand side of Eq. (27) can be expressed as

$$\int_0^{V_{ds}} \frac{\pm q}{C_{ox}} n_{e(h)} dV_{n(p)} = \int_0^{V_{ds}} \left(V_{gs} - V_{fb} - \psi_{s,e(h)} \right) dV_{n(p)} \quad (42)$$

We use $\int \psi \, dV = \psi V - \int V \, d\psi$ to integrate the third term in the right-hand side expression of Eq. (42) by part, and then apply Eqs. (39) and (40), which express $V_{n(p)}$ as an explicit function of $\psi_{s,e(h)}$. Employing the formula:

$$\int \ln \left[e^u - 1 \right] du = \frac{u^2}{2} + \mathrm{Li}_2 \left(e^{-u} \right), \quad u > 0 \quad (43)$$

leads to a transcendental function $\mathrm{Li}_2(\xi)$, which denotes the polylogarithm function of order two. By applying Eq. (43) to the integrate-by-part result of Eq. (42), we arrive at

$$I_{e(h)} = \pm q \frac{W}{L} C_{ox} \mu_{e(h)} [(V_{gs} - V_{fb} - \psi_{2,e(h)}) V_{ds} + \left(F \left(\psi_{2,e(h)} \right) - F \left(\psi_{1,e(h)} \right) \right)],$$

$$F(u) = \frac{u^2}{2} \mp \psi_{Bn(p)} u + \frac{1}{2} \zeta_{e(h)} V_T \left(\frac{V_{gs} - V_{fb} - u}{\zeta_{e(h)}} \right)^2 +$$

$$\zeta_{e(h)} V_T \mathrm{Li}_2 \left(e^{\mp \left(\frac{V_{gs} - V_{fb} - u}{\zeta_{e(h)}} \right)} \right) \quad (44)$$

where the integral upper/lower limits $\psi_{2,e(h)}/\psi_{1,e(h)}$ can be determined from Eqs. (39) or (40) by setting $V_{n(p)} = V_{ds}(0)$.

The typical results of the transfer and output characteristic calculation are shown in Fig. 10. The ambipolar transfer and saturation drain output characteristic are achieved by the model simulation. The current curves calculated from Eqs. (36) and (37) are also potted in Fig. 10, for comparison, which justifies the approximation. Equation (39) holds $\psi_s < V_{gs} - V_{fb}$ but is contrasting to Eq. (40), which means that V_n governs the variation of ψ_s until $\psi_s > V_{gs} - V_{fb}$ and then V_p takes over ψ_s.

To graphically illustrate the concept, the explicit functions between $V_{n(p)}$ and ψ_s with different V_{gs} and the mapping relation between the trajectory of ψ_s and that of (V_n, V_p) are shown in Fig. 11a. Based on the mapping between ψ_s and $V_n - V_p$, we can sketch (but not quantitatively because we do not tackle the accurate
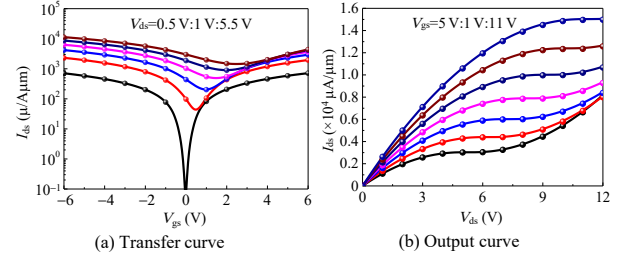

(a) Transfer curve   (b) Output curve

Fig. 10 Ambipolar transfer and output characteristics. $I - V$ curves from Yarmoghaddam's split electron hole model are calculated with the following parameters: $N_{-(+)} = 0$, $m_e^* = m_h^* = m_0$, $E_g = 0.78$ eV, $\phi_{Bn(p)} = 0.5 E_g$, $V_{fb} = -0.27$ V, $t_{ox} = 2$ nm, $\varepsilon_{ox} = 3.9 \, \varepsilon_0$, $W/L = 4$, $\mu_{eff,e(h)} = 150$ cm^2/(V·s), $V_T = 0.026$ V. The round dots represent the results from Yarmoghaddam's model (Eqs. (39) and (40)), and the solid line represents the results from Eqs. (36) and (37). The model reproduces the ambipolar transfer curve. The transfer characteristic curves are calculated using $V_{ds} = 0.5$ V:1 V:5.5 V, and the output characteristic curve $V_{gs} = 5$ V:1 V:11 V from bottom to top, respectively. The output curves do not saturate with $V_{gs} < 10$ V because the hole's concentration benefits from the increasing V_{ds}. Then it is saturated with strongly positive V_{gs}.
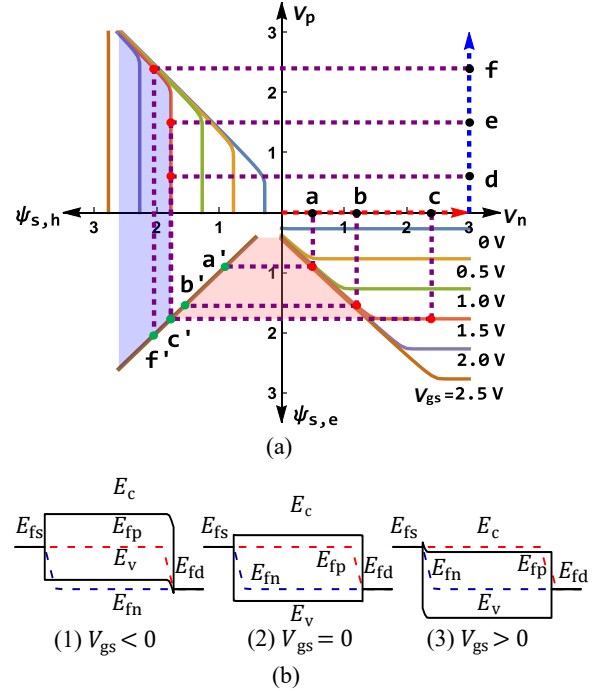

(a)

(1) $V_{gs} < 0$ (2) $V_{gs} = 0$ (3) $V_{gs} > 0$

(b)

Fig. 11 (a) Mapping from (V_n, V_p) to ψ_s and (b) channel energy band diagram for $V_{ds} > 0$. (a) Dots "a" – "f" are the representative preimage of the map in the $V_n - V_p$ space, whose image is composed of dots "a" to "f" in the $\psi_{s,h} - \psi_{s,e}$ space. Dots "a", "b", and "c" move along the V_n-axis and are thus mapped through the $V_n - \psi_{s,e}$ obeying Eq. (39). Dots "d" – "f" follow Eq. (40) correspondingly. The details of the mapping are relevant with V_{gs}, and here, we just show the case for $V_{gs} = 1.5$ V because similar discussions also hold for the other cases.

spatial distribution) the channel energy band diagram under a given bias as shown in Fig. 11b. Comparing the band diagram with that in Fig. 5b, where the channel potential V_ℓ does not split throughout the channel, is intriguing.

Yarmoghaddam et al.[58, 59] proposaled Eqs. (39) and (40) without explanations (we have omitted $N_{-(+)}$ and the trap interface here):

$$V_{\text{fb}} \pm \frac{q n_{\text{e(h)}}}{C_{\text{ox}}} + \psi_{\text{s,e(h)}} = V_{\text{gs,e(h)}} \qquad (45)$$

Clearly, Eq. (45) is equivalent to Eqs. (39) and (40). Thus, we now understand that (1) Eq. (45) indicates that a non-degenerate transport assumption has been made, and (2) the subscript "e" or "h" in ψ_{s} in Eq. (45) claims which quasi-Fermi level controls the surface potential ψ_{s} but not the type of carrier it belongs to (this is an improper concept because there is always one unique electrostatic potential determined by Poisson's equation in an electronic system up to an additive constant). With this approximation, Yarmoghaddam et al.[58, 59] successfully determined the quasi-Fermi level for electrons and holes successfully. Moreover, the p-branch and n-branch of the predicted ambipolar transfer curve by the model can be adjusted independently according to the benchmark, which entails the flexibility of the model. One can just omit the "a"–"c" (or "d"–"f") segments in Fig. 11, if the unipolar transfer curve is expected for the device[57].

3.3 Other non-ideal effects

3.3.1 Schottky barrier junction effect

The discussion above assumes that the contacts formed at the MS interfaces are Ohmic, which has been achieved by the experiments[75]. Hence, the Schottky barrier at the MS interfaces has been omitted. The Schottky barrier was not altered by the bias in Fig. 11b, because the barrier was not plotted out at all. The conduction carrier-type conversion totally relies on the channel electrostatic modulation. Such a treatment cannot tackle the case in which a nearly intrinsic channel exhibits an unipolar conduction, and thus the junction limited effect must be considered. The concerning extension can be achieved by introducing the concept of the ideal source/drain potential,

$$V_{\text{s(d)i}} = V_{\text{s(d)}} \pm \left(I_{\text{e}} R_{\text{e,s(d)}} + I_{\text{h}} R_{\text{h,s(d)}} \right) \qquad (46)$$

where $R_{\text{e,s(d)}}$ and $R_{\text{h,s(d)}}$ represent the resistances of the electrons and holes at the source/drain junction.

3.3.2 Defect charge and non-constant mobility

Vulnerability to moisture results in defects, such as oxide charge and other scattering mechanisms important to BP-FETs[76]. However, the description of the ionized defect charge does share the expression with Eq. (6); thus resulting in a quite similar generalization of Eq. (41) to involve the defect charge effect. Other scattering mechanisms can be easily incorporated into the ideal model by tailing the constant mobility with charge scattering[77, 78].

3.3.3 Geometry correction of channel irregularity

A high-quality BP channel produced in a laboratory is usually fabricated via mechanical exfoliation[79–81], which makes the geometry of the BP flakes highly unpredictable, and thus the top view of the channel is not strictly rectangular but often approximately a trapezoid unless the post-process of lithography is used to artificially scissor the shape at the risk of impairing BP. The channel-limited compact model enables us to compute the effective channel width W_0 due to the geometry correction of channel irregularity, which reads as

$$W_0 = L / \int_0^L \frac{1}{W(x)} \text{d}x \qquad (47)$$

where $W(x)$ denotes the channel width at the position of x. We attached its proof as follows.

For convenience, we consider the unipolar electron conduction. At position x in the channel, we have

$$I_{\text{ds}} = \mu_{\text{e}} \frac{\text{d}V_{\text{n}}(x)}{\text{d}x} W(x) Q_{\text{n}}(x) \qquad (48)$$

where $Q_{\text{n}}(x)$ denotes the charge density per unit area. We consider the integral with variable upper limit,

$$\int_0^x I_{\text{ds}} \text{d}x' = \int_0^{V_{\text{n}}} \mu W(x') Q_{\text{n}}(x') \text{d}V_{\text{n}}' \qquad (49)$$

and use the condition of constant-I_{ds}, so the left-hand expression of Eq. (49) is equal to $x \cdot I_{\text{ds}}$. Then we differentiate V_{n} on both sides of Eq. (49), which results in

$$\frac{\text{d}x}{\text{d}V_{\text{n}}} \cdot I_{\text{ds}} = \mu W(x) Q_{\text{n}}(V_{\text{n}}) \qquad (50)$$

Rearranging the factors in Eq. (50) as

$$\frac{\text{d}x}{W(x)} \cdot I_{\text{ds}} = \mu Q_{\text{n}}(V_{\text{n}}) \text{d}V_{\text{n}} \qquad (51)$$

Consequently, we find

$$I_{\text{ds}} = \left(\int_0^L \frac{\text{d}x}{W(x)} \right)^{-1} \int_0^{V_{\text{ds}}} \mu Q_{\text{n}}(V_{\text{n}}) \text{d}V_{\text{n}} \qquad (52)$$

The geometry effect on the drain current is involved

in the pre-factor of Eq. (52), it allowing us to define the effective channel width as indicated by Eq. (47). However, clearly, we cannot derive it from a junction-limited model.

4 VSED Model

The virtual source[82–86] (such as that adopted in Refs. [9] and [87]), or more specifically, the VSED model[88, 89] (such as that adopted in Ref. [90]), is a physical-based method that seamlessly connects the diffusive and ballistic regimes of MOSFETs. It adapts the thermal emission theory of the Schottky diode developed by Crowell and Sze[91] for unipolar MOSFETs. Thus, it naturally includes the effect of the junction into the model. A recent work[90] demonstrated that it can also be extended to describe ambipolar 2D-FETs.

The concept of "virtual source" in FETs was first introduced in the 1980s by Owczarek[92] to study the punch-through effect, which has been defined as the point x_0 where the transverse (along the channel length direction) electrical field vanishes and the electrostatic potential reaches its maximum lower than the original built-in potential barrier (Fig. 12) or, equivalently[96, 97], the top of the barrier seen by the carrier emission from the source. Furthermore, Owczarek[92] pointed out that the "virtual source" is also the start point where carrier injection into the drain space charge region takes place[92].

Nonetheless, we would prefer to argue that the concept is initially defined for traditional FETs[96–98] in which the source and drain junctions are homogeneous. Thus, in principle, it seems to be generalized when applied to Schottky barrier FETs[94, 95]. The resulting formula is invariant because the scattering process implied by the VSED model treats the source barrier as a black box, and does not distinguish its component[98, 99].
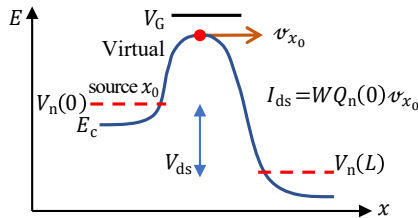
Fig. 12 Sketch of the energy band diagram of the basic VSED model reproduced from Ref. [93]. We emphasize that the details of the diagram at the source and drain need modifications when the theory is applied to Schottky barrier FETs, but it will not break down the original computational framework[94, 95].

4.1 MIT Virtual Source (MVS) model

In the most succinct VSED model[93, 98], the net quasi-ballistic current is recognized as[98]

$$I_{ds} = W Q_n v_{x_0} \tag{53}$$

where Q_n is often modeled by[83, 90]

$$Q_n = C_g n_e \cdot V_T \ln\left(1 + \exp\left(\frac{V_{gs} - V_{th,e}}{n_e \cdot V_T}\right)\right) \tag{54}$$

where n_e is the introduced nonideal factor of electrons related with nonideal subthreshold swing[83]. Now, we have known Eq. (54) is the 2D fermions statistics augmented with parameter n_e.

According to the one-flux scattering theory in the ballistic channel (Fig. 13) by Lundstrom et al.[99–102], the relation between Q_n and the injection flux from the source j_s is given by

$$Q_n v_{th} = (1 + r_C) j_s \tag{55}$$

where r_C is the reflection coefficient and v_{th} is the thermal velocity of electrons. Thus, the drain current I_{ds} reads as

$$I_{ds} \equiv W j_s t_C = W Q_n v_{th} (1 - r_C)/(1 + r_C) \tag{56}$$

which is compared with Eq. (53) to get the expression for v_{x_0},

$$v_{x_0} = \frac{1 - r_C}{1 + r_C} v_{th} \tag{57}$$

If there holds

$$r_C = \frac{\ell}{\ell + \lambda} \tag{58}$$

$$\ell = V_T / \mathcal{E} \tag{59}$$

$$D_n = v_{th} \lambda / 2 \tag{60}$$

then combining Eqs. (58)–(60), it is easy to find Eq. (57) equivalent to

$$\frac{1}{v_{x_0}} = \frac{1}{\mu_n \mathcal{E}} + \frac{1}{v_{th}} \tag{61}$$

where an ordinary Einstein relation has been used.

Later on, to provide a smooth transition from the saturation regime to the linear regime, an empirical ratio function F_{sat} is added to the model[9, 82, 84], namely,

$$I_{ds} = W Q_n v_{x_0} F_{sat} \tag{62}$$

where $0 \leqslant F_{sat} < 1$ and it reads as

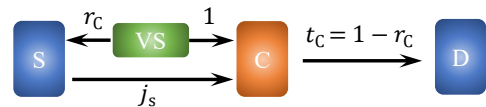$$F_{sat} = \frac{V_{ds}/V_{dsat}}{\left(1 + (V_{ds}/V_{dsat})^\beta\right)^{1/\beta}} \tag{63}$$

Fig. 13 Schematic of the scattering theory by Lundstrom.

Clearly, $F_{sat} \rightarrow 1$ when $V_{ds} \gg V_{dsat}$, and Eq. (62) degrades back to Eq. (53).

Equations (53)–(63) represent the MVS model. Besides the nonlinearity of the parasitic source/drain resistance, the model gives the virtual source velocity v_{x_0} that is restricted to the non-degenerated statistics of the carrier and has assumed a biased-independent gate capacitance C_g to evaluate Q_n. To overcome the shortcomings, a generalization about C_g and v_{x_0}, which is the so-called MVS-2 model, can be found in Refs. [84] and [85].

4.2 VSED theory

The VSED theory stated above is aimed at the quasi-ballistic regime. Long-channel-length devices need modifications to incorporate the diffusion component into the model, which is the motivation for the VSED model.

According to the pioneer work by Crowell and Sze[91], once the Schottky barrier height of the carrier (e.g., electrons) $q\phi_{Bn} = E_c(0) - E_F$ is given, the current density J_n can be expressed as

$$J_n = N_{2D}v_D \left(e^{V_n(L)/V_T} - e^{V_n(0)/V_T} \right) \cdot e^{-E_c(0)/(qV_T)}$$
(64)

where

$$V_n(L) = V_{ds}, \quad V_n(0) = E_F \quad (65)$$

$$v_D = D_n/\ell' \quad (66)$$

$$\ell' = \int_0^L e^{[E_C(x) - E_C(0)]/k_B T} dx \quad (67)$$

Equation (64) reflects the continuity of the emission and diffusion current across the Schottky barrier.

The original form which is based on the Crowell-Sze formula is designed for Schottky diode devices. After incorporating the electrostatic potential, the formula is adapted to describe the current in FETs.

The idea of the model adapted to MOS electrostatics lies in building up the relation between V_{gs} and $q\phi_{B_n}$. This is performed by giving an extra relation between J_n and ϕ_{B_n}, which incorporates V_{gs}-modulated Q_n.

We assume[89] two fluxes F^+ and F^- at the virtual source that determine the net electron density J_n as

$$J_n = -q\left(F^+ - F^-\right) \quad (68)$$

$F^{\pm}$ should satisfy that

$$Q_n \cdot v_{th} = q\left(F^+ + F^-\right) \quad (69)$$

and

$$F^+ = \frac{1}{2} N_{2D} v_{th} e^{-\frac{E_c(0) - E_F}{qV_T}} \quad (70)$$

Then we arrive at[83]

$$J_n = Q_n v_{th} \frac{\frac{\mathcal{T}}{2-\mathcal{T}}\left(1 - e^{-V_{ds}/V_T}\right)}{1 + \frac{\mathcal{T}}{2-\mathcal{T}} e^{-V_{ds}/V_T}} \quad (71)$$

where

$$\mathcal{T} = \frac{\lambda}{\lambda + \ell'} = \frac{1}{1 + v_{th}/(2 - D_n/\ell')} \quad (72)$$

We define

$$F_{sat,VSED} = \frac{\mathcal{T}}{2 - \mathcal{T}} \frac{1 - e^{-V_{ds}/V_T}}{1 + \frac{\mathcal{T}}{2-\mathcal{T}} e^{-V_{ds}/V_T}} \quad (73)$$

Then I_{ds} can be simply expressed as

$$I_{ds} = WQ_n v_{th} F_{sat,VSED} \quad (74)$$

Clearly, we can recognize

$$I_{ds} \rightarrow WQ_n v_{th}\left(\mathcal{T}/2\right)\left(V_{ds}/V_T\right), \quad V_{ds} \rightarrow 0 \quad (75)$$

$$I_{ds} \rightarrow WQ_n v_{th}\mathcal{T}/(2 - \mathcal{T}), \quad V_{ds} \rightarrow +\infty \quad (76)$$

The high and low V_{ds} limitations of the VSED model correspond to thoes of the VS model[88], which implies that the former is a proper variant of the latter.

4.3 Ambipolar model

The ambipolar generalization of VSED was initially developed by Ref. [90], suggesting that the virtual source of electrons and holes arise near different contacts, which is reflected in the expression for the virtual source charge, that is,

$$Q_e = C_g n_e V_T \ln\left(1 + \exp\left(+\frac{V_{gs} - V_{th,e}}{n_e V_T}\right)\right) \quad (77)$$

$$Q_h = C_g n_h V_T \ln\left(1 + \exp\left(-\frac{V_{gd} - V_{th,h}}{n_h V_T}\right)\right) \quad (78)$$

where $V_{gs} \equiv V_{gs} - V_{ds}$, and $V_{th,e(h)}$ denotes the threshold voltage for electrons and holes, respectively.

Their treats on the entanglement of V_n and V_p are straightforward, which continue their philosophy about the splitted electrostatic potential. Thus, ℓ' can be defined for both types of carrier as

$$\ell'_i = \int_0^L \exp\left(-\frac{V_i(x) - V_i(0)}{V_T}\right) dx \quad (79)$$

where $V_i(x)$ is the profile of electrostatic potential and $i = e$ or h, denoting for electrons and holes, respectively. Moreover, λ can be easily generalized from Eq. (60) as

$$\lambda_i = 2\mu_i V_T/v_{th,i} \quad (80)$$

based on which the ambipolar-generalization $\mathcal{T}_i$ from the channel transmission $\mathcal{T}$ of Eq. (72) reads

$$\mathcal{T}_i = \lambda_i/\left(\lambda_i + \ell'_i\right) \quad (81)$$

The ambipolar saturation transition function $F_{sat,i}$ is then defined as

$$F_{sat,i} = \frac{\mathcal{T}_i}{2 - \mathcal{T}_i} \times \frac{1 - e^{-V_{ds}/V_T}}{1 + \frac{\mathcal{T}_i}{2-\mathcal{T}_i} e^{-V_{ds}/V_T}} \quad (82)$$

Thus, the ambipolar current reads

$$I_e = W Q_e v_{T,e} F_{sat,e} \qquad (83)$$

$$I_h = W Q_h v_{T,h} F_{sat,h} \qquad (84)$$

According to Eq. (79) (or Eq. (67)), the model requires that proper approximations[88] about the potential profile should be used for obtaining a satisfactory result from the model, such as gradual channel approximation[77] accommodated for long-channel devices ($L > \lambda$). There are also other choices, such as the exponential form adopted by Ref. [94] for short channel devices. However, one can always turn to the numerical simulation for an accurate profile, even though it usually gives the potential and current simultaneously.

Other corrections such as nonlinearity source/drain resistance, complexity in mobility, and threshold voltage available for the model, show at least the same extendable merit with other approaches[90].

5 Conclusion and Prospect

Published ambipolar transport compact models are classified into three types: (1) Landauer formula based model, which mainly reflects the Schottky junction effect on the drain current; (2) 2D Pao-Sah model, which on the contrary mainly focuses on the channel-limited effect; and (3) VSED model, which potentially contains the effects of the junction and channel.

The junction-limited model treats the channel potential as a lumped parameter. On one hand, it allows to customize a compact model, such as Lego blocks, and to release the computational load. On the other hand, this condition restricts the model's application. For example, as we have discussed, the channel's detailed geometry characteristic cannot be tackled in a junction-limited model and V_{ds} should be restricted to small amplitudes in principle. However, it naturally applies for short-channel lengths, where scattering events are rare.

For the 2D Pao-Sah model, we have explained the compact model proposed by Yarmoghaddam et al.[58, 59] Compared to the Landauer formula based approach, the model is more refined in delving into the physics of devices and is more stretchable in incorporating various nonideal effects. The junction effect and other nonideal effects can then be added to the model easily. Consequently, the Pao-Sah based model is more suitable for use if the channel length is not small enough and the drain-source bias does not vanish at all.

The VSED model provides appropriate descriptions for the critical transition from long to short-channel lengths because it originated from the ballistic transport theory of FETs and was adapted to the diffusive regime. Thus, it serves as a connection between the ultimate channel length cases.

In summary, the ambipolar compact models reported nowadays have covered a fairly complete range of channel lengths and thus formed a relatively systematic simulation methodology. However, with the increase in model complexity, which manifests itself with the number of model parameters, the simulation accuracy is improved at the cost that the underlying device physics becomes more ambiguous. Consequently, improvement of the model without impairing the clear physical picture still remains unsolved. With the development in this field, we can expect that there will be more solid works to cover this aspect in the near future.

References

[1] L. K. Li, Y. J. Yu, G. J. Ye, X. H. Chen, and Y. B. Zhang, Electronic properties of few-layer black phosphorus, *Bull. Am. Phys. Soc.*, vol. 58, no. 1, p. 1, 2013.

[2] L. K. Li, Y. J. Yu, G. J. Ye, Q. Q. Ge, X. D. Ou, H. Wu, D. L. Feng, X. H. Chen, and Y. B. Zhang, Black phosphorus field-effect transistors, *Nat. Nanotechnol.*, vol. 9, no. 5, pp. 372–377, 2014.

[3] H. Liu, A. T. Neal, Z. Zhu, Z. Luo, X. F. Xu, D. Tománek, and P. D. Ye, Phosphorene: An unexplored 2D semiconductor with a high hole mobility, *ACS Nano*, vol. 8, no. 4, pp. 4033–4041, 2014.

[4] Y. C. Du, H. Liu, Y. X. Deng, and P. D. Ye, Device perspective for black phosphorus field-effect transistors: Contact resistance, ambipolar behavior, and scaling, *ACS Nano*, vol. 8, no. 10, pp. 10 035–10 042, 2014.

[5] H. O. H. Churchill and P. Jarillo-Herrero, Phosphorus joins the family, *Nat. Nanotechnol.*, vol. 9, no. 5, pp. 330–331, 2014.

[6] R. Gusmão, Z. Sofer, and M. Pumera, Black phosphorus rediscovered: From bulk material to monolayers, *Angew. Chem., Int. Ed.*, vol. 56, no. 28, pp. 8052–8072, 2017.

[7] Y. Ren and F. Cheng, Ballistic transport through a strained region on monolayer phosphorene, *Chin. Phys. Lett.*, vol. 34, no. 2, p. 027302, 2017.

[8] Z. Nourbakhsh and R. Asgari, Charge transport in doped zigzag phosphorene nanoribbons, *Phys. Rev. B*, vol. 97, p. 235406, 2018.

[9] X. F. Li, Z. Q. Yu, X. Xiong, T. Y. Li, T. T. Gao, R. S. Wang, R. Huang, and Y. Q. Wu, High-speed black phosphorus field-effect transistors approaching ballistic limit, *Sci. Adv.*, vol. 5, no. 6, p. eaau3194, 2019.

[10] N. R. Pradhan, D. Rhodes, Y. Xin, S. Memaran, L. Bhaskaran, M. Siddiq, S. Hill, P. M. Ajayan, and L. Balicas, Ambipolar molybdenum diselenide field-effect transistors: Field-effect and hall mobilities, *ACS Nano*, vol. 8, no. 8, pp. 7923–7929, 2014.

[11] S. Das and J. Appenzeller, WSe$_2$ field effect transistors with enhanced ambipolar characteristics, *Appl. Phys. Lett.*, vol. 103, no. 10, p. 103501, 2013.

[12] Z. G. Wang, Q. Li, Y. F. Chen, B. X. Cui, Y. R. Li, F. Besenbacher, and M. D. Dong, The ambipolar transport behavior of WSe$_2$ transistors and its analogue circuits, *NPG Asia Mater.*, vol. 10, no. 8, pp. 703–712, 2018.

[13] C. Q. Yin, Y. X. Li, J. B. Wang, X. F. Wang, Y. Yang, and T. L. Ren, Carbon nanotube transistor with short-term memory, *Tsinghua Science and Technology*, vol. 21, no. 4, pp. 442–448, 2016.

[14] Y. Ren, X. Y. Yang, L. Zhou, J. Y. Mao, S. T. Han, and Y. Zhou, Recent advances in ambipolar transistors for functional applications, *Adv. Funct. Mater.*, vol. 29, no. 40, p. 1902105, 2019.

[15] P. Wu, D. Reis, X. S. Hu, and J. Appenzeller, Two-dimensional transistors with reconfigurable polarities for secure circuits, *Nat. Electron.*, vol. 4, no. 1, pp. 45–53, 2021.

[16] A. Castellanos-Gomez, Black phosphorus: Narrow gap, wide applications, *J. Phys. Chem. Lett.*, vol. 6, no. 21, pp. 4280–4291, 2015.

[17] W. N. Hu, Z. Sheng, X. Hou, H. W. Chen, Z. X. Zhang, D. W. Zhang, and P. Zhou, Ambipolar 2D semiconductors and emerging device applications, *Small Methods*, vol. 5, no. 1, p. 2000837, 2021.

[18] G. Gildenblat, *Compact Modeling*: *Principles*, *Techniques and Applications*. Dordrecht, the Netherlands: Springer, 2010.

[19] J. S. Qiao, X. H. Kong, Z. X. Hu, F. Yang, and W. Ji, High-mobility transport anisotropy and linear dichroism in few-layer black phosphorus, *Nat. Commun.*, vol. 5, no. 1, p. 4475, 2014.

[20] S. J. Liu, N. J. Huo, S. Gan, Y. Li, Z. M. Wei, B. J. Huang, J. Liu, J. B. Li, and H. D. Chen, Thickness-dependent Raman spectra, transport properties and infrared photoresponse of few-layer black phosphorus, *J. Mater. Chem. C*, vol. 3, no. 42, pp. 10 974–10 980, 2015.

[21] A. N. Rudenko and M. I. Katsnelson, Quasiparticle band structure and tight-binding model for single- and bilayer black phosphorus, *Phys. Rev. B*, vol. 89, p. 201408(R), 2014.

[22] X. X. Cui, D. Han, H. L. Guo, L. W. Zhou, J. S. Qiao, Q. Liu, Z. H. Cui, Y. F. Li, C. W. Lin, L. M. Cao, et al., Realizing nearly-free-electron like conduction band in a molecular film through mediating intermolecular van der Waals interactions, *Nat. Commun.*, vol. 10, p. 3374, 2019.

[23] K. Ko, M. Kang, J. Jeon, and H. Shin, Compact model strategy of metal-gate work-function variation for Ultrascaled FinFET and vertical GAA FETs, *IEEE Trans. Electron Devices*, vol. 66, no. 3, pp. 1613–1616, 2019.

[24] H. C. Pao and C. T. Sah, Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors, *Solid-State Electron.*, vol. 9, no. 10, pp. 927–937, 1966.

[25] C. Gupta, R. Goel, H. Agarwal, C. M. Hu, and Y. S. Chauhan, BSIM-BULK: Accurate compact model for analog and RF circuit design, presented at IEEE Custom Integrated Circuits Conf., Austin, TX, USA, 2019, pp. 1–8.

[26] J. G. Champlain, On the use of the term "ambipolar", *Appl. Phys. Lett.*, vol. 99, p. 123502, 2011.

[27] P. Wu, T. Ameen, H. R. Zhang, L. A. Bendersky, H. Ilatikhameneh, G. Klimeck, R. Rahman, A. V. Davydov, and J. Appenzeller, Complementary black phosphorus tunneling field-effect transistors, *ACS Nano*, vol. 13, no. 1, pp. 377–385, 2019.

[28] D. J. Perello, S. H. Chae, S. Song, and Y. H. Lee, High-performance n-type black phosphorus transistors with type control via thickness and contact-metal engineering, *Nat. Commun.*, vol. 6, p. 7809, 2015.

[29] M. Lundstrom, *Fundamentals of Carrier Transport*. 2nd ed. Cambridge, UK: Cambridge University Press, 2000.

[30] S. Das, W. Zhang, M. Demarteau, A. Hoffmann, M. Dubey, and A. Roelofs, Tunable transport gap in phosphorene, *Nano Lett.*, vol. 14, no. 10, pp. 5733–5739, 2014.

[31] S. Das, M. Demarteau, and A. Roelofs, Ambipolar phosphorene field effect transistor, *ACS Nano*, vol. 8, no. 11, pp. 11 730–11 738, 2014.

[32] I. S. Esqueda, C. D. Cress, Y. C. Che, Y. Cao, and C. W. Zhou, Charge trapping in aligned single-walled carbon nanotube arrays induced by ionizing radiation exposure, *J. Appl. Phys.*, vol. 115, p. 054506, 2014.

[33] A. V. Penumatcha, R. B. Salazar, and J. Appenzeller, Analysing black phosphorus transistors using an analytic Schottky barrier MOSFET model, *Nat. Commun.*, vol. 6, p. 8948, 2015.

[34] I. S. Esqueda, H. Tian, X. D. Yan, and H. Wang, Transport properties and device prospects of ultrathin black phosphorus on hexagonal boron nitride, *IEEE Trans. Electron Devices*, vol. 64, no. 12, pp. 5163–5171, 2017.

[35] S. Datta, *Lessons from Nanoelectronics*: *A New Perspective on Transport– Part A*: *Basic Concepts*. 2nd ed. Singapore: World Scientific Publishing, 2017.

[36] R. H. Yan, A. Ourmazd, and K. F. Lee, Scaling the Si MOSFET: From bulk to SOI to bulk, *IEEE Trans. Electron Devices*, vol. 39, no. 7, pp. 1704–1710, 1992.

[37] N. Haratipour, S. Namgung, S. H. Oh, and S. J. Koester, Fundamental limits on the subthreshold slope in Schottky source/drain black phosphorus field-effect transistors, *ACS Nano*, vol. 10, no. 3, pp. 3791–3800, 2016.

[38] M. C. Robbins, N. Haratipour, and S. J. Koester, Band-to-band tunneling limited ambipolar current in black

phosphorus MOSFETs, in *Proc. 75th Annual Device Research Conf.*, South Bend, IN, USA, 2017, pp. 1–2.

[39] S. Datta, *Electronic Transport in Mesoscopic Systems.* Cambridge, UK: Cambridge University Press, 1995.

[40] I. S. Esqueda, C. D. Cress, Y. Cao, Y. Che, M. Fritze, and C. Zhou, The impact of defect scattering on the quasi-ballistic transport of nanoscale conductors, *J. Appl. Phys.*, vol. 117, p. 084319, 2015.

[41] F. Liu, Y. J. Wang, X. Y. Liu, J. Wang, and H. Guo, Ballistic transport in monolayer black phosphorus transistors, *IEEE Trans. Electron Devices*, vol. 61, no. 11, pp. 3871–3876, 2014.

[42] Y. Q. Cai, G. Zhang, and Y. W. Zhang, Layer-dependent band alignment and work function of few-layer phosphorene, *Sci. Rep.*, vol. 4, p. 6677, 2014.

[43] J. Kim, S. S. Baik, S. H. Ryu, Y. Sohn, S. Park, B. G. Park, J. Denlinger, Y. J. Yi, H. J. Choi, and K. S. Kim, Observation of tunable band gap and anisotropic Dirac semimetal state in black phosphorus, *Science*, vol. 349, no. 6249, pp. 723–726, 2015.

[44] B. C. Deng, V. Tran, Y. J. Xie, H. Jiang, C. Li, Q. S. Guo, X. M. Wang, H. Tian, S. J. Koester, H. Wang, et al., Efficient electrical control of thin-film black phosphorus bandgap, *Nat. Commun.*, vol. 8, p. 14474, 2017.

[45] B. Jiang, X. M. Zou, J. Su, J. H. Liang, J. L. Wang, H. J. Liu, L. P. Feng, C. Z. Jiang, F. Wang, J. He, et al., Impact of thickness on contact issues for pinning effect in black phosphorus field-effect transistors, *Adv. Funct. Mater.*, vol. 28, no. 26, p. 1801398, 2018.

[46] S. Y. Lee, W. S. Yun, and J. D. Lee, New method to determine the Schottky barrier in few-layer black phosphorus metal contacts, *ACS Appl. Mater. Interfaces*, vol. 9, no. 8, pp. 7873–7877, 2017.

[47] Y. Z. Guo and J. Robertson, Band offsets and metal contacts in monolayer black phosphorus, *Microelectron. Eng.*, vol. 178, pp. 108–111, 2017.

[48] J. A. Nelder and R. Mead, A simplex method for function minimization, *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[49] A. Gil, J. Segura, and N. M. Temme, *Numerical Methods for Special Functions.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007.

[50] J. J. Sakurai and J. J. Napolitano, *Modern Quantum Mechanics.* 2nd ed. Cambridge, UK: Cambridge University Press, 2017.

[51] K. F. Brennan and C. J. Summers, Theory of resonant tunneling in a variably spaced multiquantum well structure: An Airy function approach, *J. Appl. Phys.*, vol. 61, no. 2, pp. 614–623, 1987.

[52] R. A. Vega, On the modeling and design of Schottky field-effect transistors, *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 866–874, 2006.

[53] X. M. Guan, D. Kim, K. C. Saraswat, and H. S. P. Wong, Complex band structures: From parabolic to elliptic approximation, *IEEE Electron Device Lett.*, vol. 32, no. 9, pp. 1296–1298, 2011.

[54] J. Callaway, *Quantum Theory of the Solid State.* New York, NY, USA: Academic Press, 1976.

[55] X. D. Yan, H. Wang, and I. S. Esqueda, Temperature-dependent transport in ultrathin black phosphorus field-effect transistors, *Nano Lett.*, vol. 19, no. 1, pp. 482–487, 2019.

[56] C. Guo, L. Wang, H. Z. Xing, and X. S. Chen, The study of ambipolar behavior in phosphorene field-effect transistors, *J. Appl. Phys.*, vol. 120, p. 215701, 2016.

[57] E. G. Marin, S. J. Bader, and D. Jena, A new holistic model of 2-D semiconductor FETs, *IEEE Trans. Electron Devices*, vol. 65, no. 3, pp. 1239–1245, 2018.

[58] E. Yarmoghaddam, N. Haratipour, S. J. Koester, and S. Rakheja, A physics-based compact model for ultrathin black phosphorus FETs – Part I: Effect of contacts, temperature, ambipolarity, and traps, *IEEE Trans. Electron Devices*, vol. 67, no. 1, pp. 389–396, 2020.

[59] E. Yarmoghaddam, N. Haratipour, S. J. Koester, and S. Rakheja, A physics-based compact model for ultrathin black phosphorus FETs – Part II: Model validation against numerical and experimental data, *IEEE Trans. Electron Devices*, vol. 67, no. 1, pp. 397–405, 2020.

[60] A. Ueda, Y. J. Zhang, N. Sano, H. Imamura, and Y. Iwasa, Ambipolar device simulation based on the drift-diffusion model in ion-gated transition metal dichalcogenide transistors, *NPJ Comput. Mater.*, vol. 6, p. 24, 2020.

[61] C. T. Sah, The spatial variation of the quasi-Fermi potentials in p-n junctions, *IEEE Trans. Electron Devices*, vol. ED-13, no. 12, pp. 839–846, 1966.

[62] C. T. Sah, Characteristics of the metal-oxide-semiconductor transistors, *IEEE Trans. Electron Devices*, vol. 11, no. 7, pp. 324–345, 1964.

[63] N. W. Ashcroft and N. D. Mermin, *Solid State Physics.* Philadelphia, PA, USA: Saunders College, 1976.

[64] J. R. Schrieffer, Effective carrier mobility in surface-space charge layers, *Phys. Rev.*, vol. 97, p. 641, 1955.

[65] R. F. Greene, D. R. Frankl, and J. Zemel, Surface transport in semiconductors, *Phys. Rev.*, vol. 118, p. 967, 1960.

[66] R. F. Greene, Nonlocal transport and cuspidal surface mobility in semiconductors, *Phys. Rev.*, vol. 131, p. 592, 1963.

[67] R. H. Kingston and S. F. Neustadter, Calculation of the space charge, electric field, and free carrier concentration at the surface of a semiconductor, *J. Appl. Phys.*, vol. 26, p. 718, 1955.

[68] F. Stern and W. E. Howard, Properties of semiconductor surface inversion layers in the electric quantum limit, *Phys. Rev.*, vol. 163, p. 816, 1967.

[69] F. Stern, Self-consistent results for *n*-type Si inversion layers, *Phys. Rev. B*, vol. 5, p. 4891, 1972.

[70] R. B. Dingle, The Fermi-Dirac integrals $F_p(\eta) = (p!)^{-1} \int_o^\infty \epsilon^p (e^{\epsilon - \eta + 1})^{-1} d\epsilon$, *Appl. Sci. Res., Sect. B*, vol. 6, no. 1, pp. 225–239, 1957.

[71] T. Fukushima, Precise and fast computation of Fermi–Dirac integral of integer and half integer order by piecewise minimax rational approximation, *Appl. Math. Comput.*, vol. 259, pp. 708–729, 2015.

[72] O. N. Koroleva, A. V. Mazhukin, V. I. Mazhukin, and P. V. Breslavskiy, Analytical approximation of the Fermi-Dirac integrals of half-integer and integer orders, *Math. Models Comput. Simul.*, vol. 9, no. 3, pp. 383–389, 2017.

[73] G. T. Wright, Threshold modelling of MOSFETs for CAD of CMOS-VLSI, *Electron. Lett.*, vol. 21, no. 6, pp. 223–224, 1985.

[74] G. Mugnaini and G. Iannaccone, Physics-based compact model of nanoscale MOSFETs–Part I: Transition from drift-diffusion to ballistic transport, *IEEE Trans. Electron Devices*, vol. 52, no. 8, pp. 1795–1801, 2005.

[75] F. Telesio, G. le Gal, M. Serrano-Ruiz, F. Prescimone, S. Toffanin, M. Peruzzini, and S. Heun, Ohmic contact engineering in few-layer black phosphorus: Approaching the quantum limit, *Nanotechnology*, vol. 31, p. 334002, 2020.

[76] Y. J. Xu, Z. Shi, X. Y. Shi, K. Zhang, and H. Zhang, Recent progress in black phosphorus and black-phosphorus-analogue materials: Properties, synthesis and applications, *Nanoscale*, vol. 11, no. 31, pp. 14 491–14 527, 2019.

[77] Y. Tsividis and C. McAndrew, *Operation and Modeling of the MOS Transistor*. 3rd ed. Oxford, UK: Oxford University Press, 2011.

[78] N. Ma and D. Jena, Charge scattering and mobility in atomically thin semiconductors, *Phys. Rev. X*, vol. 4, p. 011043, 2014.

[79] B. C. Yang, B. S. Wan, Q. H. Zhou, Y. Wang, W. T. Hu, W. M. Lv, Q. Chen, Z. M. Zeng, F. S. Wen, J. Y. Xiang, et al., Te-doped black phosphorus field-effect transistors, *Adv. Mater.*, vol. 28, no. 42, pp. 9408–9415, 2016.

[80] D. Xiang, C. Han, J. Wu, S. Zhong, Y. Y. Liu, J. D. Lin, X. A. Zhang, W. P. Hu, B. Özyilmaz, A. H. C. Neto, et al., Surface transfer doping induced effective modulation on ambipolar characteristics of few-layer black phosphorus, *Nat. Commun.*, vol. 6, p. 6485, 2015.

[81] A. Castellanos-Gomez, L. Vicarelli, E. Prada, J. O. Island, K. L. Narasimha-Acharya, S. I. Blanter, D. J. Groenendijk, M. Buscema, G. A. Steele, J. V. Alvarez, et al., Isolation and characterization of few-layer black phosphorus, *2D Mater.*, vol. 1, no. 2, p. 025001, 2014.

[82] L. Wei, O. Mysore, and D. Antoniadis, Virtual-source-based self-consistent current and charge FET models: From ballistic to drift-diffusion velocity-saturation operation, *IEEE Trans. Electron Devices*, vol. 59, no. 5, pp. 1263–1271, 2012

[83] S. Rakheja, Y. Q. Wu, H. Wang, T. Palacios, P. Avouris, and D. A. Antoniadis, An ambipolar virtual-source-based charge-current compact model for nanoscale graphene transistors, *IEEE Trans. Nanotechnol.*, vol. 13, no. 5, pp. 1005–1013, 2014.

[84] S. Rakheja, M. S. Lundstrom, and D. A. Antoniadis, An improved virtual-source-based transport model for quasi-ballistic transistors–Part I: Capturing effects of carrier degeneracy, drain-bias dependence of gate capacitance, and nonlinear channel-access resistance, *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2786–2793, 2015.

[85] S. Rakheja, M. S. Lundstrom, and D. A. Antoniadis, An improved virtual-source-based transport model for quasi-ballistic transistors–Part II: Experimental verification, *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2794–2801, 2015.

[86] D. H. Kim, J. A. del Alamo, D. A. Antoniadis, and B. Brar, Extraction of virtual-source injection velocity in sub-100 nm III-V HFETs, presented at 2009 IEEE Int. Electron Devices Meeting (IEDM), Baltimore, MD, USA, 2009, pp. 1–4.

[87] A. Nourbakhsh, A. Zubair, R. N. Sajjad, K. G. A. Tavakkoli, W. Chen, S. Fang, X. Ling, J. Kong, M. S. Dresselhaus, E. Kaxiras, et al., MoS_2 field-effect transistor with Sub-10 nm channel length, *Nano Lett.*, vol. 16, no. 12, pp. 7798–7806, 2016.

[88] S. Rakheja, M. Lundstrom, and D. Antoniadis, A physics-based compact model for FETs from diffusive to ballistic carrier transport regimes, presented at 2014 IEEE Int. Electron Devices Meeting, San Francisco, CA, USA, 2014, pp. 35.1.1–35.1.4.

[89] M. Lundstrom, S. Datta, and X. S. Sun, Emission-diffusion theory of the MOSFET, *IEEE Trans. Electron Devices*, vol. 62, no. 12, pp. 4174–4178, 2015.

[90] E. Yarmoghaddam, N. Haratipour, S. J. Koester, and S. Rakheja, A virtual-source emission-diffusion I-V model for ultra-thin black phosphorus field-effect transistors, *J. Appl. Phys.*, vol. 125, p. 165706, 2019.

[91] C. R. Crowell and S. M. Sze, Current transport in metal-semiconductor barriers, *Solid State Electron.*, vol. 9, no. 11–12, pp. 1035–1048, 1966.

[92] A. Owczarek, Punch-through phenomenon in MOS transistors, *Electron Technol.*, vol. 13, nos. 1 & 2, pp. 55–65, 1980.

[93] M. S. Lundstrom and D. A. Antoniadis, Compact models and the physics of nanoscale FETs, *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 225–233, 2014.

[94] C. S. Lee, E. Pop, A. D. Franklin, W. Haensch, and H. S. P. Wong, A compact virtual-source model for carbon nanotube FETs in the sub-10-nm regime–Part I: Intrinsic elements, *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 3061–3069, 2015.

[95] C. S. Lee, E. Pop, A. D. Franklin, W. Haensch, and H. S. P. Wong, A compact virtual-source model for carbon nanotube FETs in the Sub-10-nm regime–Part II: Extrinsic elements, performance assessment, and design optimization, *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 3070–3078, 2015.

[96] D. A. Antoniadis, I. Aberg, C. N. Chléirigh, O. M. Nayfeh, A. Khakifirooz, and J. L. Hoyt, Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations, *IBM J. Res. Dev.*, vol. 50, nos. 4 & 5, pp. 363–376, 2006.

[97] A. Khakifirooz and D. A. Antoniadis, Transistor performance scaling: The role of virtual source velocity and its mobility dependence, presented at 2006 Int. Electron Devices Meeting, San Francisco, CA, USA, 2006, pp. 1–4.

[98] A. Khakifirooz, O. M. Nayfeh, and D. Antoniadis, A simple semiempirical short-channel MOSFET current-voltage model continuous across all regions of operation and employing only physical parameters, *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1674–1680, 2009.

[99] M. Lundstrom, Elementary scattering theory of the Si MOSFET, *IEEE Electron Device Lett.*, vol. 18, no. 7, pp. 361–363, 1997.

[100] K. Natori, Ballistic metal-oxide-semiconductor field effect transistor, *J. Appl. Phys.*, vol. 76, no. 8, pp. 4879–4890, 1994.

[101] Y. Liu, M. Luisier, A. Majumdar, D. A. Antoniadis, and M. S. Lundstrom, On the interpretation of ballistic injection velocity in deeply scaled MOSFETs, *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 994–1001, 2012.

[102] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, Theory of ballistic nanotransistors, *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1853–1864, 2003.
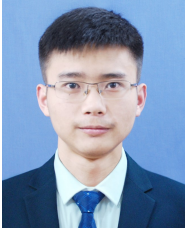
Zhaoyi Yan is currently pursuing the PhD degree at Tsinghua University. He received the bachelor degree from Huazhong University of Science & Technology in 2018. Now, His main research interests are focused on the modeling and simulation of novel micro/nano electronic devices.
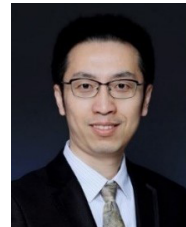
Guangyang Gou received the BS and MS degrees in applied physics from Central South University in 2014 and 2017, respectively. He is currently pursuing the PhD degree at the Institute of Microelectronics, Tsinghua University. His research interests include micro/nano electronic devices and acoustic sensor with 2D material.

Jie Ren is currently pursuing the master degree at Tsinghua University. He received the BE degree in microelectronics and solid state electronics from Northwestern Polytechnical University in 2018. His main research interest is focused on 2D material transistors.

Fan Wu is pursuing the PhD degree at Tsinghua University. He received the bachelor degree from Hefei University of Technology in 2018. Now, he concentrates research topics on novel 2D transistors.

He Tian received the PhD degree from the Institute of Microelectronics, Tsinghua University in 2006. He is currently an associate professor at Institute Microelectronics, Tsinghua University. His research interests focus on novel micro/nano devices based on 2D materials. He is the member of IEEE and EDS and has published more than 100 journal and conference papers.

Yang Shen is currently pursuing the PhD degree at Tsinghua University. He received the bachelor degree from Hefei technology of University in 2019. Now, he concentrates research topics on the study of micro/nano electronic devices based on 2D materials.
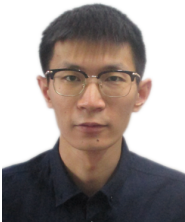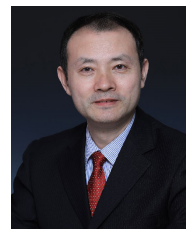
Yi Yang received the BS degree from Fudan University in 2001 and the PhD degree from Tsinghua Universityin in 2006. He is currently an associate professor at Institute of Microelectronics, Tsinghua University. His research interests focus on intelligent micro/nano devices and systems. He has published more than 150 journal and conference papers and has been granted with over 10 China patents. He has served as Technical Program Committee member of IEEE International Electron Devices Meeting (IEDM) (2016, 2017).
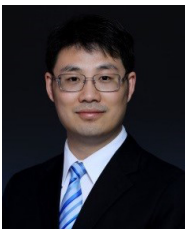
Tian-Ling Ren received the PhD degree from Tsinghua University in 1997. He has been the full professor of Institute of Microelectronics, Tsinghua University since 2003. He was a visiting professor at Electrical Engineering Department of Stanford University in the USA from 2011 to 2012. His research interest covers a broad range of advanced electronics devices and integrated systems. He has published over 300 papers in international journals and conferences. He has served as the administrative committee member of Chinese Society of Micro-Nano Technology, managing director of MEMS & NEMS Society of China Instrument and Control Society, vice president of IEEE Electron Devices Society, administrative committee member of IEEE Electron Devices Society (2015–2016), distinguished lecturer of IEEE Electron Devices Society, executive committee member of IEEE EDTM, and editorial board member of *IEEE Journal of Electron Devices Society*, *IEEE Transaction on Nanotechnology*, and *Scientific Reports*.