2021

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

Zhaoyi Yan  
*Institute of Microelectronics, Tsinghua University, Beijing 100084, China*

Guangyang Gou  
*Institute of Microelectronics, Tsinghua University, Beijing 100084, China*

Jie Ren  
*Institute of Microelectronics, Tsinghua University, Beijing 100084, China*

Fan Wu  
*Institute of Microelectronics, Tsinghua University, Beijing 100084, China*

Yang Shen  
*Institute of Microelectronics, Tsinghua University, Beijing 100084, China*

See next page for additional authors

Follow this and additional works at: [https://dc.tsinghuajournals.com/tsinghua-science-and-technology](https://dc.tsinghuajournals.com/tsinghua-science-and-technology)  
Part of the Computer Sciences Commons, and the Electrical and Computer Engineering Commons

**Recommended Citation**  

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in Tsinghua Science and Technology by an authorized editor of Tsinghua University Press: Journals Publishing.
Ambipolar Transport Compact Models for Two-Dimensional Materials Based Field-Effect Transistors

Authors
Zhaoyi Yan, Guangyang Gou, Jie Ren, Fan Wu, Yang Shen, He Tian, and Yi Yang

This research article is available in Tsinghua Science and Technology: https://dc.tsinghuajournals.com/tsinghua-science-and-technology/vol26/iss5/2
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] (MoSe2), and tungsten diselenide[11, 12] (WSe2), 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,
Ambipolar Transport Compact Models for Two-Dimensional Materials Based Field-Effect Transistors

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 capacitor, which couples the gate-source biased voltage $V_{gs}$ and the channel potential $\varphi$ to the drain current $I_{ds}$, and part (2) the gate resistor, which decides the map $V_{ds}$; $I_{ds}$ as shown in Fig. 3. Thus, formally, the drain current $I_{ds}$ is determined by $V_{ds}$ and $V_{gs}$ as follows:

$$I_{ds} = \Gamma (\varphi, V_{ds}) \quad V_{gs} = \Lambda (\varphi) \quad \Rightarrow \quad I_{ds} = \Gamma (\Lambda^{-1} (V_{gs}), V_{ds}) \quad (1)$$

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

![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, $I_e$ and $I_h$ represent electron and hole components of the drain current.](image)

![Fig. 2 Publication citations of three types of modeling methods. Source: Web of Science.](image)

![Fig. 3 Device schematic. Here, G: Gate, S: Source, D: Drain, C: Channel.](image)
Formula (1) is more likely (but not necessarily) a scalar number $V_\ell$ representing the overall channel potential. Thus, map $\Gamma$ stays as a function, whereas for the 2D Pao-Sah model, $\phi$ represents the channel electrostatic potential $\psi_z$, and it varies with the channel coordinate or quasi-Fermi level, thus making $\Gamma$ 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

$$\xi^2 (a_{\mu}; V_{ds}) = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{I_{ds}^{(i)}(V_{gs}, V_{fb}, V_{V}) - \Gamma A^{-1}(V_{gs}, V_{fb}, V_{V})}{I_{ds}^{(i)}} \right|^2$$

where $a_{\mu}$, $I_{ds}^{(i)}$, 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 $\phi$ is represented by a single value $V_\ell$ 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 $V_\ell$ 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, $V_\ell$ and $V_{gs}$ are connected with each other by Eqs. (3)–(6),

$$V_\ell = (V_{gs} - V_{fb}) \frac{C_{ox}}{C_q + C_{ox}}$$

(3)

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

(4)

$$C_q = \int_{-\infty}^{+\infty} \frac{q D(E)}{4V_T} \operatorname{sech}^2 \left( \frac{E/q - V_\ell}{2V_T} \right) dE$$

(5)

$$V_{fb} = \frac{\phi_{m} - \phi_{BP}}{q} - \frac{q}{C_{ox}} N_{\Gamma} - \frac{q}{C_{ox}} \int_{-\infty}^{+\infty} \frac{D_{n,a}}{1 + \exp \left( (E + qV_\ell)/qV_T \right)} dE + \frac{q}{C_{ox}} \int_{-\infty}^{+\infty} \frac{D_{n,d}}{1 + \exp \left( -(E + qV_\ell)/qV_T \right)} dE$$

(6)

![Fig. 4 Device model and channel band diagram.](image)
where

\[ D(E) = \begin{cases} 
\frac{m^*}{(\pi \hbar^2)}, & E \leq E_c; \\
0, & E_c < E \leq E_f; \\
\frac{m^*}{(\pi \hbar^2)}, & E \geq E_f. 
\end{cases} \tag{7} \]

\[ D_n,a(d) = \frac{N_{n,a}(d)}{\sqrt{2\pi \sigma_{n,a}(d)}} e^{-\frac{(E-E_{c(v)})^2}{2\sigma_{n,a}(d)}}, \tag{8} \]

where \( \hbar \) is reduced Planck constant, \( V_T \) denotes the thermal voltage equivalent, \( \phi_{m(BP)} \) 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, \( \varepsilon_{ox} \) 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, \( m^*(b) \) is the effective mass of the electrons and holes, \( E_{c(v)} \) is the minimum (maximum) of the conduction (valence) band, \( N_\ell \) is the fixed charge density in the oxide layer trapped in the occupied density of states per unit area per unit energy, \( D_{n,a}(d) \) is the interface acceptors (donors) per unit energy of the 2D periodic system, \( \sigma_{n,a}(d) \) is the standard deviation of \( D_{n,a}(d) \).

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 \( V_{t} - V_{gs} \) 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 \( V_{t} \), the vertex of the energy band diagram as shown in Fig. 4b, can be determined as follows:

\[ E_{fs} = qV_t - (1 - \alpha) \cdot qV_{ds} \tag{9} \]
\[ E_{fd} = qV_t + \alpha \cdot qV_{ds} \tag{10} \]
\[ E_{c(v),s} = E_{c(v)} + \phi_{sb,a(p)} \tag{11} \]
\[ E_{c(v),d} = E_{c(d)} + \phi_{sb,a(p)} \tag{12} \]

where \( \alpha \) 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]. \( \phi_{sb,a(p)} \) is the Schottky barrier for electrons (holes), \( E_{fd(d)} \) 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. \( \lambda_{ch} \), the width of the barrier formed at the channel/source (drain) interface is approximated using the screening length as[36]

\[ \lambda_{ch} = \left( \frac{E_{c(v)} - \varepsilon_{ox}}{\varepsilon_{ch} t_{ch} + \varepsilon_{ox} t_{ch}} \right)^{-1/2} \tag{13} \]

where \( \varepsilon_{ch} \) and \( t_{ch} \) are the permittivity and thickness of the channel, respectively. The correctness of \( \lambda_{ch} \) 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:

\[ I_{ds} = 2q\frac{W}{h} \int_{-\infty}^{+\infty} T(E) M(E) [f_d - f_s] dE \tag{14} \]

where \( h \) is Planck constant, \( W \) is the channel width, \( f_{d(s)}(E) \) 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

\[ f_{d(s)} = \frac{1}{1 + \exp \left( \frac{E - E_{c(v)}}{qV_T} \right)} \tag{15} \]

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

\[ T(E) = T_{c(h)} \cdot \left[ \pm (E - E_{c(v)}) \right] \tag{17} \]

\[ T_{c(h)}^{-1} = T_{c(h),s}^{-1} + T_{c(h),d}^{-1} - 2 \tag{18} \]

\[ T_{c(h),s}^{-1} = \exp \left[ \frac{4 \sqrt{2m^*_v}}{3h/\lambda_{ch}} \cdot \sqrt{\left( \pm (E_{c(v)} - E_{c(v)}) \left[ \pm (E_{c(v)} - E_{c(v)}) \right] \right)^3} \right] \tag{19} \]
The first term on the right-hand side of Eq. (22) with 
\( y_0 = 3/2 \) represents the contribution from the charged
impurity scattering and \( y_1 = 1/2 \) represents that from the
phonon scattering. The adjustable parameters \( \lambda_{e,(b)} \)
and \( \lambda_{1,e,(b)} \) 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 \( T_{e(b)} = 1 \), 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 \( V_\ell \) 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

\[
T_{e(b),d}^{-1} = \exp \left[ \frac{4 \sqrt{2m_e^*}}{3\hbar / \lambda_B} \left( \frac{\pm (E_{e(c),d} - E)}{qV_T} \right)^3 \right] 
\]

(20)

where \( \theta[\cdot] \) denotes the Heaviside step function; \( L \) is
the channel length, \( g_\ell \) is the valley degeneracy, which is
regularly set as 2; and \( \lambda_{e,(b)}(E) \) is the energy-dependent
mean free path. It is assumed that \( \lambda_{e,(b)}(E) \) consists of
two components expressed as\(^{[20, 34, 40]}
\[
\frac{1}{\lambda_{e,(b)}} = \frac{N_{it}}{\lambda_{0,e,(b)}} \left( \frac{|E - E_{c,v}|}{qV_T} \right)^{-y_0} + \frac{1}{\lambda_{1,e,(b)}} \left( \frac{|E - E_{c,v}|}{qV_T} \right)^{-y_1}
\]

(22)

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

\[
N_{it} = \int_{-\infty}^{+\infty} \frac{D_{n,a}}{1 + \exp \left[ (E + qV_t) / qV_T \right]} dE + \int_{-\infty}^{+\infty} \frac{D_{n,d}}{1 + \exp \left[ - (E + qV_t) / qV_T \right]} dE
\]

(23)

The parameters are set as \( \phi_{a,app} = 0.5 \), \( E_g = 0.39 \) eV, \( m_e^* = m_0 \), \( N_{it,app} = 8.8 \times 10^{12} \) cm\(^{-2} \), \( \sigma_{eff} = 10^2 \), \( \lambda_{0,e,(b)} = 300 \) nm, \( e_{e,0} = 3.9 \),
\( e_{e,0} = 8.3 \), \( e_{t,0} = 2 \) nm, \( L = 10^2 \) nm, \( \gamma = 0.026 \) V, and \( V_{th} = 10 \) 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 (Id),
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 \( M(E), T(E), f_{d}(E) - f_{e}(E) \), 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), $T(E)$, $M(E)$, $f_0(E) - f_s(E)$, 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 $f_0(E) - f_s(E)$ 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 circuit’s 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 $E_g$, $m_{e/h}^*$, $\phi_{b/n}$, $N_i$, $N_r$, $\sigma_{d/d}$, $\lambda_{0,e/b}$, and $\lambda_{1,e/b}$ 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 $N_r$ 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 $\xi$ 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 $\xi$ 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 $V_{d} - V_{gs}$ 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 $\alpha$ of $V_{ds}$ and (2) boundary conditions of barrier width $\lambda_B$, which have been mentioned before, the rest five are analyzed.
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 \( g \)), 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\(^{[30]} \) at the short-wavelength limit and fails at the turning point, i.e., \( E = V(x) \), where

\[
\lambda_{de Broglie} = \frac{h}{\sqrt{2m E_{c(x)}|[E - V(x)]}} \to \infty,
\]

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 transferring 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 \( E_{c(x)} \) 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 \( E_{c(x)} \) 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

\[
\frac{1}{T} = \frac{1}{T_s} + \frac{1}{T_d} - 1
\]

(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) “\( \lambda_B \)” 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) “\( T_e \)” 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 “Ad&T” and “\( \psi \)” 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

\[
\text{Ref. } & \quad V_{ds} \quad \lambda_B \quad T_B \quad \text{B.B. } E(k) \quad \text{BtBT} \quad T_e \\
[30] & \quad 1 \quad \epsilon \quad W \quad x \quad P \quad \text{Bal.} \\
[33, 55] & \quad 1 \quad \epsilon \quad W \quad x \quad P&E \quad \text{noB} \quad \text{Bal.} \\
[37] & \quad 1 \quad \epsilon \quad W \quad \exp \quad E \quad \text{noB} \quad \text{Bal.} \\
[34] & \quad 1/2 \quad \epsilon \quad W \quad x \quad P \quad \text{noB} \quad \text{Dif.} \\
[38] & \quad 1 \quad \epsilon \quad W \quad \exp \quad P \quad \text{noB} \quad \text{Bal.}
\]

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.
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. Inspired 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 $E^0_k$ when $V_{ds} = 0$. This convention is necessary for MS-junction-based 2D-FETs.

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

\[
J_{eb}(b) = -\mu_{eb} e_n E_{fn(p)} \nabla E_{fn(p)} 
\]  

\[
I_{ds} = -\frac{W}{L} \int_0^{V_{ds}} \mu_e n_e dE_{fn} + \mu_h n_h dE_{fp} 
\]  

\[
V_{fb} + V_{ox} + \psi_s = V_{gs} 
\]  

In Eq. (25), $n_{eb}$ denotes the electron (hole) density, $E_{(n(p))} = -qV_{(n(p))}$, and $j_{eb}$ is the current density of electrons (holes) in the plane ($D\rightarrow S$) 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 $j_{eb}$, which finally transforms the drain current integral with the coordinate to that with the quasi-Fermi level. Particularly, Eq. (27) has identified $\mu_{eb}$ as the average surface mobility[62, 64–66], thus being independent with the spatial coordinate, and $n_{eb}$ is described by the Fermi-Dirac distribution under the control of the difference of $E_{fn(p)}$, i.e., $-qV_{fn(p)}$, and the channel surface electrostatic potential barrier $-q\psi_s$, 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 $\psi_s$ and $E_{fn(p)}$, which are comprised of four parts: gate-source voltage $V_{gs}$, flat-band voltage $V_{fb}$, oxide voltage drop $V_{ox}$, and surface potential $\psi_s$. 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., $V_{gs} = (E_{fm} - E_{fs})/(-q)$.

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

(3) Surface potential energy is defined as $\psi_s = (E_c - E^0_c)/(q)$, where $E^0_c = E^0_c + \phi_{fn} = E_{fs} + \phi_{fn} = \phi_{fn}$. $\phi_{fn}$ is the electrons’ Fermi potential barrier. Its counterpart $\phi_{fp}$ is the holes’ Fermi potential barrier. For convenience, $q \phi_{fn(p)} \equiv \phi_{BSfn(p)}$ is assumed. $E^0_c$ and $E^0_c$ represent the equilibrium location of the conduction band edge and Fermi energy of
the BP-channel, respectively. Once the drain-source electrodes are biased, $E_f^b$ splits into $E_{fn}$ and $E_{fp}$; thus $\pm (E_{fn(p)} - E_{c(s)}) = \pm E_{fn(p)} \pm q\psi_s - \phi_{fn(p)}$ measures the electron (hole) density and indicates a degenerate transport system if it is positive.

4. $V_{ox} = Q_s/C_{ox}$, 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 $\rho_b$ built-up with Poisson’s equation and Gauss’s law along the depth-direction is necessary to obtain the density $Q_{s}^{bulk}$ as indicated in Eq. (29)\textsuperscript{[24, 67]}:

$$Q_{s}^{bulk} = \text{sign}(\psi_s)\sqrt{2\epsilon_s} \left| \int_{0}^{\psi_s} \rho_b d\psi \right|^\frac{1}{2}$$ (29)

where the coordinate variable is substituted with the electrostatic potential $\psi$ 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 $F_i(\xi)$, i.e., Eq. (31) can benefit from the exponential zero condition ($j = 0$), resulting in a closed form expression\textsuperscript{[68–72]} for the areal density of the carrier $Q_{c}^{2D} = \pm q n_{c(h)}$, 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\textsuperscript{[24, 67]}. Thus, the total density of charge is $Q_s^{2D} = Q_{c}^{2D} = Q_{c}^{2D} + qN_e - qN_h$, where $N_e$ 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.

$$D_{c(h)} = m_{c(h)}^*/(\pi\hbar^2)$$ (30)

$$n_{c(h)} = \int_{E_{c(h)}}^{\pm\infty} \frac{\pm D_{c(h)}}{e^{\pm\frac{E - E_{fn(h)}}{kT}} + 1} dE$$ (31)

$$Q_{c}^{2D} = \pm qKD_{c(h)} \ln \left( 1 + e^{\pm\frac{E_{fn(h)} - E_{c(s)}}{kT}} \right)$$ (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 adjoins the strong inversion status and subthreshold regime in bulk materials\textsuperscript{[73]},

The proponents of the equation may not have realized that their mathematical modeling trick will become a physical fact decades later\textsuperscript{[74]}. The regular treatment on the relation of $\psi_s$ and $E_{fn(p)}$ 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 $\psi_s, E_{fn}$, and $E_{fp}$ from Eq. (28), with $V_{gs}$ taken as the parameter,

$$n_c (\psi_s, E_{fn}) - n_h (\psi_s, E_{fp}) = \frac{C_{ox}}{q} (V_{gs} - V_{fb} - \psi_s)$$ (33)

Equation (33) indicates an implicit function of $\psi_s = \Psi(V_n, V_p; V_{gs})$, or a suppression of $V_{gs}$ as $\psi_s = \Psi(V_n, V_p)$ so that no confusion is caused.

3.2 Compact model

We rearrange Eq. (33) as follows:

$$V_{n(p)} = \psi_s \mp \psi_{fn(p)} \mp V_T \ln \left[ \gamma_c (\psi_s) e^{\frac{\pm V_{gs} - V_{fb} - \psi_s}{\xi_c} - 1} \right]$$ (34)

where $\gamma_c (\psi_s) = m_{n(h)}(\psi_s)\xi_c$,

$$\gamma_c (\psi_s) = \left( e^{\frac{V_{gs} - V_{fb} - \psi_s}{V_T}} + 1 \right)^{m_{n(h)}(\psi_s)/m_{e(h)}}$$ (35)

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

$$V_n = \psi_s - \psi_{fn} - V_T \ln \left[ \gamma_c (\psi_s) e^{\frac{\pm V_{gs} - V_{fb} - \psi_s}{\xi_c} - 1} \right]$$ (36)

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

$$V_p = \psi_s + \psi_{fn} + V_T \ln \left[ \gamma_c (\psi_s) e^{\frac{\pm V_{gs} - V_{fb} - \psi_s}{\xi_c} - 1} \right]$$ (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 $\exp(-\psi_{fn(p)} \mp \psi_s \pm V_{pin(p)}/V_T)$ is certainly not much bigger than one. Thus, Eq. (34) can be properly approximated as

$$V_{n(p)} = \psi_s \mp \psi_{fn(p)} \mp V_T \ln \left[ e^{\frac{\pm V_{gs} - V_{fb} - \psi_s}{\xi_c} - 1} \right]$$ (38)

Then Eqs. (36) and (37) follow the approximation. When $V_p = 0$, and $0 < V_n < V_{ds}$,

$$V_n = \psi_s - \psi_{fn} - V_T \ln \left[ e^{\frac{\pm V_{gs} - V_{fb} - \psi_s}{\xi_c} - 1} \right]$$ (39)
whereas when \( V_d = V_b \), and \( 0 < V_p < V_d \),

\[
V_p = \psi + \psi_{Bp} + V_T \ln \left[ \frac{V_{p} = V_{n} - \psi_s}{V_{n} - \psi_{s,\epsilon(b)}} \right] - 1 \quad (40)
\]

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

\[
\pm q n_{\epsilon(b)} = C_{ox} \left( V_{gs} - V_{fb} - \psi_{s,\epsilon(b)} \right) \quad (41)
\]

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

\[
\int_{0}^{V_d} \frac{\pm q}{C_{ox}} n_{\epsilon(b)} dV_{n(p)} = \int_{0}^{V_d} \left( V_{gs} - V_{fb} - \psi_{s,\epsilon(b)} \right) dV_{n(p)} \quad (42)
\]

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 \( \psi_{s,\epsilon(b)} \). Employing the formula:

\[
\ln [e^u - 1] du = \frac{u^2}{2} + \text{Li}_2 (e^{-u}) \quad u > 0 \quad (43)
\]

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

\[
\begin{align*}
I_{\epsilon(b)} &= \pm \frac{W}{L} C_{ox}\mu_{\epsilon(b)} \left[ \left( V_{gs} - V_{fb} - \psi_{2,\epsilon(b)} \right) V_{ds} + \\
&\quad \left( F \left( \psi_{2,\epsilon(b)} \right) - F \left( \psi_{1,\epsilon(b)} \right) \right) \right] + \\
&\quad F (u) = \frac{u^2}{2} + \psi_{Bp}(u) + \frac{1}{4} \psi_{\epsilon(b)} V_T \left( \frac{V_{gs} - V_{fb} - u}{\psi_{\epsilon(b)}} \right)^2 + \\
&\quad \psi_{\epsilon(b)} V_T \text{Li}_2 \left( e^{-\left( \frac{V_{gs} - V_{fb} - u}{\psi_{\epsilon(b)}} \right)} \right) \quad (44)
\end{align*}
\]

where the integral upper/lower limits \( \psi_{2,\epsilon(b)}/\psi_{1,\epsilon(b)} \) 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 plotted 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 \( \psi_s \) until \( \psi_s > V_{gs} - V_{fb} \) and then \( V_p \) takes over \( \psi_s \).

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

![Fig. 10 Ambipolar transfer and output characteristics. I - V curves from Yarmoghaddam’s split electron hole model are calculated with the following parameters: \( N_{\epsilon(a)} = 0, m_n^* = m_n, E_g = 0.78 \text{eV}, \psi_{Bp}(0) = 0.5 \text{eV}, V_{fb} = -0.27 \text{V}, t_{ox} = 2 \text{nm}, \, \epsilon_{ox} = 3.9 \epsilon_0, \, W/L = 4, \, \mu_{\epsilon(\text{split})} = 150 \text{cm}^2/\text{V} \cdot \text{s}, \, V_T = 0.026 \text{V}. \) 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 \( V_{ds} = 0.5 \text{V} \): \( 1 \text{V} \): \( 5.5 \text{V} \), and the output characteristic curve \( V_{gs} = 5 \text{V} \): \( 11 \text{V} \) from bottom to top, respectively. The output curves do not saturate with \( V_{gs} < 10 \text{V} \) because the hole’s concentration benefits from the increasing \( V_{ds} \). Then it is saturated with strongly positive \( V_{gs} \).

![Fig. 11 (a) Mapping from \( (V_n,V_p) \) to \( \psi_s \) 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” - “f” in the \( \psi_{s,b} - \psi_{s,e} \) space. Dots “a”, “b”, and “c” move along the \( V_n \)-axis and are thus mapped through the \( V_n = \psi_{s,b} \) obeying Eq. (39). Dots “d” - “f” follow Eq. (40) correspondingly. The details of the mapping are relevant with \( V_{gs} \) and hence, we just show the case for \( V_{gs} = 1.5 \text{V} \) 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 $V_I$ does not split throughout the channel, is intriguing.

Yarmoghaddam et al.\cite{58, 59} proposed Eqs. (39) and (40) without explanations (we have omitted $N_{s(+)}$ and the trap interface here):

$$V_{fb} \pm \frac{qN_{s(e)}(b)}{C_{ox}} + \psi_{s,e}(b) = V_{gs,e}(b) \tag{45}$$

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 $\psi_s$ in Eq. (45) claims which quasi-Fermi level controls the surface potential $\psi_s$, 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.\cite{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\cite{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\cite{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,

$$V_{s(d)i} = V_{s(d)} \pm (I_e R_{e,s(d)} + I_h R_{h,s(d)}) \tag{46}$$

where $R_{e,s(d)}$ and $R_{h,s(d)}$ 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\cite{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\cite{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\cite{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

$$W_0 = L/\int_0^L \frac{1}{W(x)} dx \tag{47}$$

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

$$I_{ds} = \mu_e \frac{dV_e(x)}{dx} W(x) Q_s(x) \tag{48}$$

where $Q_s(x)$ denotes the charge density per unit area. We consider the integral with variable upper limit,

$$\int_0^x I_{ds} dx' = \int_0^x \mu W(x') Q_s(x') dV_n' \tag{49}$$

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

$$\frac{dV_n}{dx} \cdot I_{ds} = \mu W(x) Q_n(V_n) \tag{50}$$

Rearranging the factors in Eq. (50) as

$$\frac{dx}{W(x)} \cdot I_{ds} = \mu Q_n(V_n) dV_n \tag{51}$$

Consequently, we find

$$I_{ds} = \left(\int_0^L \frac{dx}{W(x)}\right)^{-1} \int_0^{V_n} \mu Q_n(V_n) dV_n \tag{52}$$

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\[^{86, 97}\], 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 wherecarrier 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\[^{86-91}\] 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\[^{94, 95}\].](image)

4.1 MIT Virtual Source (MVS) model

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

\[
I_{ds} = WQ_nv_{x_0}
\]

(53)

where \(Q_n\) is often modeled by\[^{83, 90}\]

\[
Q_n = C_e n_e \cdot V_T \ln \left( 1 + \exp \left( \frac{V_{gs} - V_{th,e}}{n_e \cdot V_T} \right) \right)
\]

(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

\[
Q_nv_{th} = (1 + r_C) j_s
\]

(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

\[
I_{ds} \equiv Wj_s r_C = W Q_n v_{th} (1 - r_C)/(1 + r_C)
\]

(56)

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

\[
v_{x_0} = \frac{1 - r_C}{1 + r_C} v_{th}
\]

(57)

If there holds

\[
r_C = \frac{\ell}{\ell + \lambda}
\]

(58)

\[
\ell = V_T / \varepsilon
\]

(59)

\[
D_n = v_{th} \lambda / 2
\]

(60)

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

\[
\frac{1}{v_{x_0}} = \frac{1}{\mu_e \varepsilon} + \frac{1}{v_{th}}
\]

(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,

\[
I_{ds} = WQ_nv_{x_0} F_{sat}
\]

(62)

where \(0 \leq F_{sat} < 1\) and it reads as

\[
F_{sat} = \left( \frac{V_{ds}/V_{dsat}}{1 + (V_{ds}/V_{dsat})^\beta} \right)^{1/\beta}
\]

(63)

![Fig. 13 Schematic of the scattering theory by Lundstrom.](image)
Clearly, $F_{\text{sat}} \rightarrow 1$ when $V_{ds} \gg V_{\text{th}}$, 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 Eq. (72) reads

$$T \approx 1 + \frac{\lambda}{\lambda + \ell'} \frac{1 - e^{-V_{th}/V_T}}{1 - e^{-V_{th}/V_T}} \frac{1 - e^{-V_{th}/V_T}}{1 - e^{-V_{th}/V_T}}$$

(74)

Then we arrive at

$$J_n = Q_n v_{th} \frac{\tau}{1 + \frac{\tau}{2} e^{-V_{th}/V_T}}$$

(71)

where

$$\tau = \frac{\lambda}{\lambda + \ell'} = 1 + \frac{1}{1 + \frac{\lambda}{\ell'}}$$

(72)

We define

$$F_{\text{sat,VSED}} = \frac{\tau}{2 - \tau} \frac{1 - e^{-V_{th}/V_T}}{1 - e^{-V_{th}/V_T}}$$

(73)

Then $I_{ds}$ can be simply expressed as

$$I_{ds} = WQ_n v_{th} F_{\text{sat,VSED}}$$

(74)

Clearly, we can recognize

$$I_{ds} \rightarrow WQ_n v_{th} (T/2) (V_{ds}/V_T), V_{ds} \rightarrow 0$$

(75)

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

(76)

The high and low $V_{ds}$ limitations of the VSED model correspond to those 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,

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

(77)

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

(78)

where $V_{gs} = V_{gs} - V_{ds}$, and $V_{\text{th,e,h}}$ 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, $\ell'$ can be defined for both types of carrier as

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

(79)

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

$$\lambda_i = 2 \mu_i V_T / v_{th,i}$$

(80)

based on which the ambipolar-generalization $T_i$ from the channel transmission $T$ of Eq. (72) reads

$$T_i = \frac{\lambda_i}{\lambda_i + \ell'_i}$$

(81)

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

$$F_{\text{sat,i}} = \frac{T_i}{2 - T_i} \times \frac{1 - e^{-V_{th}/V_T}}{1 + \frac{T_i}{2} e^{-V_{th}/V_T}}$$

(82)
Thus, the ambipolar current reads

$$I_e = W Q_e e^{VT_e} F_{sat,e}$$  \hspace{1cm} (83)

$$I_h = W Q_h e^{VT_h} F_{sat,h}$$  \hspace{1cm} (84)

According to Eq. (79) (or Eq. (67)), the model requires that proper approximations about the potential profile should be used for obtaining a satisfactory result from the model, such as gradual channel approximation 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.

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. 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.

Acknowledgment

This work was supported by the National Key R&D Program (Nos. 2016YFA0200400 and 2018YFC2001202), and the National Natural Science Foundation of China (Nos. 61434001, 61574083, 61874065, 51861145202, and U20A20168). The authors were also thankful for the support of the Research Fund from Tsinghua University Initiative Scientific Research Program, Beijing Innovation Center for Future Chip, Beijing Natural Science Foundation (No. 4184091), and Tsinghua-Fuzhou Institute for Date Technology (No. TFIDT2018008).

References


[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.


Tsinghua Science and Technology, October 2021, 26(5): 574–591


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 University of Science & Technology 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 University 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.