

## Chapter 2

# Ultralow-Power Pseudospintronic Devices via Exciton Condensation in Coupled Two-Dimensional Material Systems

Xuehao Mou, Leonard Franklin Register, and Sanjay Kumar Banerjee

**Abstract** “Pseudospintronic” device concepts, novel “beyond-CMOS” device proposals targeted toward revolutionizing the current semiconductor technology based on MOSFETs and CMOS logic, are addressed in detail. These pseudospin devices include the voltage-controlled *Bilayer pseudoSpin Field-Effect Transistor* (BiSFET) and the current-controlled *Bilayer pseudoSpin Junction Transistor* (BiSJT). MOSFETs are confronted by the intractable physics of thermionic emission and resulting source-to-drain leakage that limits voltage scaling. As a result, CMOS faces an “energy crisis” much as the one faced by bipolar junction transistor-based logic that led to CMOS. As for many other beyond-CMOS concepts, these pseudospintronic devices are based on a completely different physics of switching, potentially allowing much lower voltage and power operation. These pseudospintronic device concepts employ possible room-temperature interlayer electron–hole exciton condensates between two dielectrically separated layers of two-dimensional (2D) materials for subthermal voltage ( $\text{sub-}k_{\text{B}}T/q$ ) switching, specifically from a nearly shorted interlayer conductance state to a highly resistive interlayer conductance state with increasing interlayer voltage. These collective exciton states with their “which-layer” degree of freedom are somewhat analogous to collective spin states in magnets, which is the origin of the “pseudospintronics” moniker. Device performance *in the presence* of such condensates is the primary focus of this work; the possibility of room-temperature condensates, itself, is addressed by other research still in progress. We begin with a discussion of the underlying physics. Graphene-based pseudospintronic systems then are analyzed using quantum transport simulations incorporating the nonlocal exchange interaction. However, the essential transport physics should be much the same for other 2D material systems, including transition metal dichalcogenides for which the realization of the condensate may be easier. The BiSFET and BiSJT device concepts are presented in detail, and basic logic gate designs are illustrated for each. Compact device models are developed and SPICE-level circuit simulations are performed to

---

X. Mou (✉) • L.F. Register • S.K. Banerjee  
The University of Texas at Austin, Austin, TX, USA  
e-mail: [xmou@utexas.edu](mailto:xmou@utexas.edu); [register@austin.utexas.edu](mailto:register@austin.utexas.edu); [banerjee@ece.utexas.edu](mailto:banerjee@ece.utexas.edu)

demonstrate possible switching energies on the scale of or below a tenth of an attojoule, well below even end-of-the-road-map CMOS. However, like many other beyond-CMOS concepts, these devices remain concepts without solid experimental embodiments. The fabrication concerns of such novel devices are also discussed along with recent experimental progress.

## 2.1 Introduction

The “pseudospintronic” devices considered in this chapter are novel “beyond-CMOS” device proposals targeted toward revolutionizing the current semiconductor technology, which is currently based on MOSFETs and CMOS logic. These pseudospin devices include the *Bilayer pseudoSpin Field-Effect Transistor* (BiSFET), the *Bilayer pseudoSpin Junction Transistor* (BiSJT), and potential applications of near-perfect Coulomb drag. MOSFETs are confronted by the intractable difficulty of thermionic emission and resulting source-to-drain leakage that limits voltage scaling. As a result, CMOS faces an “energy crisis” much as the one faced by bipolar junction transistor-based logic that leads to CMOS. As for many other beyond-CMOS concepts, these pseudospintronic devices are based on a completely different physics of switching. These pseudospintronic device concepts employ possible room-temperature interlayer electron–hole exciton condensates in bilayer two-dimensional (2D) material systems for enhanced interlayer conduction and sub- $k_B T/q$  switching voltages. In this chapter, device performance *in the presence* of such condensates is the main focus, but not the possibility of room-temperature condensates, itself, which is addressed by other research still in progress. Graphene-based pseudospintronic devices are simulated by quantum transport simulations for the essential physical properties, and SPICE-level circuit simulations are performed to demonstrate possible super low-power switching that could greatly exceed even so-called “end of the road map” CMOS. The fabrication concerns of such novel devices are also discussed along with recent experimental progress. However, also like many other beyond-CMOS concepts, these devices remain concepts without solid experimental embodiments. Though this device research has been carried out based on graphene systems until now, the essential physics should be translated to other 2D materials that may yet be more favorable than graphene for the realization of such devices.

## 2.2 Pseudospintronics

In this section, we will address the physics that underlies the proposed pseudospin devices. These devices rely on intrinsically quantum mechanical behaviors, in contrast to the CMOS device whose basic function can be explained classically. They also depend on many-body quantum effects that contrast them even to some other beyond-CMOS device concepts, such as tunnel FETs, which depend on single-particle tunneling.

### 2.2.1 Quantum Mechanics Prequel/Refresher

This subsection is specifically prepared for novices to quantum mechanics or those who have not dealt with the subject for some time, to introduce the reader to or remind them of necessary concepts going forward in this work; those well versed in the subject may skip this part. (Indeed, the material constitutes common knowledge and one may turn to one of many textbooks on quantum mechanics, solid-state physics, etc., for this information, so we will refrain from using citations here for the most part.) Although it is not necessary to understand the circuit applications of the pseudospintronic devices—knowledge of the simple  $I$ - $V$  behavior of this pseudospintronic device should be enough for readers to browse through Sect. 2.5 addressing circuit applications—quantum mechanics is essential for understanding the why’s and how’s. The exotic physics of pseudospintronic devices is the foundation for their exotic device characteristics as compared to those of conventional electronic devices.

#### 2.2.1.1 Description of the World with Quantum Mechanics

The development of quantum mechanics perhaps began in the late nineteenth century. Ludwig Boltzmann suggested that the energy states of a physical system could be discrete, and soon after Max Planck proposed energy quanta to successfully interpret blackbody radiation. However, quantum mechanics was not designed on the basis of “quanta,” but rather “quanta” were the products of quantum mechanics in which, as De Broglie postulated, there is particle-wave duality of, strictly speaking, everything. The particle-wave duality, which was first used to describe light (the “photon” as the corresponding particle), was experimentally justified in the early twentieth century as in the electron interference pattern exhibited using single-crystal nickel [1]. Nevertheless, this duality is dominated by Planck’s constant  $h$ , a constant so minute ( $6.63 \times 10^{-34}$  J s) that only microscopic particles, such as electrons, protons, neutrons, and atoms/molecules composed thereof, manifest significant particle-wave duality. In this regard, quantum mechanics is mainly applied in the research of microscopic systems. In the macroscopic world, classical mechanics is sufficiently accurate.

In Schrödinger’s wave mechanics, which is widely known, the wave property of particles is represented by a “probability wave” or “wavefunction” in real-space  $\varphi(\mathbf{r})$ , which may have real and imaginary components. Analogous to the sound wave for which how “loud” the sound—how much energy is contained in the sound wave—is proportional to the magnitude squared of the amplitude of the wave as a function of location, in quantum mechanics, the *probability* to find the particle in a certain location  $\mathbf{r}$ , another type of strength, is proportional to  $|\varphi(\mathbf{r})|^2$ , which remains an intrinsically real quantity.

Just as sound waves are delocalized in space, quantum mechanical wavefunctions generally are also delocalized, which is equivalent to saying that the location of the particle is *uncertain*. Similarly, there can be uncertainty in other

physical observables, such as momentum, energy, angular momentum, etc., associated with the quantum mechanical wavefunction. However, there is a set of wavefunctions for each physical observable, the “eigenfunctions,”  $\alpha_n(\mathbf{r})$  for which there is no uncertainty in that observable. The associated precisely known value of the observable is the “eigenvalue,”  $\gamma_n$ . However, for many, if not most, pairs of observables, being in an eigenstate of one observable will be mutually exclusive of being in the eigenstate of the other, as is the case for real-space position and momentum. This general principle is known as the *Heisenberg uncertainty principle* (which comes with a simple prescription for determining whether such uncertainty exists, which need not be addressed here).

Critically, the set of eigenfunctions of any observable  $\{\alpha_n(\mathbf{r})\}$  provides a mathematically complete and orthogonal (or at least can be made orthogonal in the case of different eigenfunctions with the same eigenvalue) basis set in which any quantum mechanical wavefunction  $\varphi(\mathbf{r})$  can be expanded, such that one may write  $\varphi(\mathbf{r}) = \sum_n c_n \alpha_n(\mathbf{r})$ . Thus, we now also may talk about the expansion coefficient  $c_n$  or the (complex) amplitude for finding the particle at any “location” in “momentum space,” “energy space” or “angular momentum space.” In other words, the  $c_n$  represent the momentum, energy or angular momentum space wavefunction, respectively. The use of such different basis sets is somewhat analogous to describing real-space position in Cartesian coordinates  $(x, y, z)$ , with cylindrical coordinates  $(\rho, \theta, z)$  or with spherical coordinates  $(r, \theta, \varphi)$ . However, the number of axes/dimensions can be infinite—and the sum over  $n$  above becomes an integral—in these spaces (which are examples of Hilbert spaces). A closer analogy would be transforming a real-space function to Fourier space or vice versa. Indeed, the transformation of a quantum mechanical wavefunction from real-space  $\mathbf{r}$  to momentum space  $\mathbf{p}$ , or vice versa, is precisely equivalent to a Fourier transform.

In contrast to the wavefunction, the *state* of a particle, however, must be independent of its coordinate system/representation. In the coordinate-system-independent Dirac “bra-ket” notation, the state can be represented by the “ket”  $|\varphi\rangle$ . The also necessary coordinate-system-independent integral,  $\int \varphi_b^*(\mathbf{r}) \varphi_a(\mathbf{r}) d^3 r = \int \varphi_b^*(\mathbf{p}) \varphi_a(\mathbf{p}) d^3 p = \dots$ , is written as  $\langle \varphi_b | \varphi_a \rangle$ , where the “bra”  $\langle \varphi |$  is the “Hermitian adjoint” of the ket  $|\varphi\rangle$ . We emphasize that expression does not represent just the product of two wavefunctions; the integration is implicit in the notation. For example, the orthogonality condition between two basis states  $|\alpha_n\rangle$  and  $|\alpha_{n'}\rangle$  takes the form  $\langle \alpha_{n'} | \alpha_n \rangle = \delta_{n,n'}$ .

To convert from the coordinate-system-independent bra-ket notation to any desired coordinate system, one may simply use the “projection operator”  $\sum_n |\alpha_n\rangle \langle \alpha_n| \equiv \mathbf{1}$ , where  $\mathbf{1}$  is the identity matrix. The equivalence to the identity matrix means that applying the projection operator to a state does not change the state in any way; however, it does “project” it onto the basis of the eigen-kets,  $|\alpha_n\rangle$ . (We note that there is a common convention to suppress the summation notation such that the expression  $|\alpha_n\rangle \langle \alpha_n|$  implies the summation, but we shall refrain from doing so here

for clarity.) That is,  $|\varphi\rangle \equiv 1|\varphi\rangle \equiv \sum_n |\alpha_n\rangle \langle \alpha_n | \varphi \rangle$ , where  $\langle \alpha_n | \varphi \rangle \equiv c_n$  are the expansion/amplitude coefficients in the basis  $|\alpha_n\rangle$  of the chosen physical observable. For example, to find the coordinate in real space, we write, now with a continuum of eigenstates,  $|\varphi\rangle = \int d^3r |\mathbf{r}\rangle \langle \mathbf{r} | \varphi \rangle$ , where the  $|\mathbf{r}\rangle$  are eigenstate of position  $\mathbf{r}$ . The expansion coefficients  $c(\mathbf{r}) = \langle \mathbf{r} | \varphi \rangle$  can be evaluated in any coordinate system, but we shall choose real space where the eigenstates of position are just delta functions  $\delta(\mathbf{r} - \mathbf{r}')$ . Therefore,  $c(\mathbf{r}) \equiv \langle \mathbf{r} | \varphi \rangle \equiv \int d^3r' \delta(\mathbf{r} - \mathbf{r}') \varphi(\mathbf{r}') \equiv \varphi(\mathbf{r})$ . That is, as a specific case of what was noted above, the expansion coefficients in the basis of the real-space eigen-kets are the real-space wavefunction.

The time evolution of the state is described by the *time-dependent Schrödinger's equation*,

$$i\hbar \frac{\partial}{\partial t} |\varphi\rangle = H |\varphi\rangle, \quad (2.1)$$

where  $H$  is the *Hamiltonian*, in analogy to the classical picture of *total energy* but in *operator* form. In the real-space representation, this equation becomes

$$i\hbar \frac{\partial}{\partial t} \varphi(\mathbf{r}, t) = \left( -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}, t) \right) \varphi(\mathbf{r}, t), \quad (2.2)$$

which is a more well-known form. If the Hamiltonian does not explicitly depend on time, i.e.,  $V(\mathbf{r}, t) = V(\mathbf{r})$ , Eq. (2.1) can be decoupled as

$$|\varphi\rangle = \sum_n c_n |\alpha_n\rangle e^{-iE_n t/\hbar} \quad (2.3a)$$

where

$$H |\alpha_n\rangle = E_n |\alpha_n\rangle \quad (2.3b)$$

is the total energy eigenvalue equation for the set of the energy eigenstates. This latter equation, whether written in real space or bra-ket notation as in Eq. (2.3b), is better known as the *time-independent* or *stationary Schrödinger's equation*. With the time-dependent Schrödinger's equation, the evolution in time of the state is uniquely described and, therefore, so is with the “movement” or “motion” of particles, at least in terms of their probability distributions for whatever observable.

In quantum mechanics, using Eq. (2.2), one may readily obtain (in real-space coordinates) the quantum mechanical version of the (in this case probability) current continuity equation for any given state  $\varphi(\mathbf{r}, t)$  as

$$\frac{\partial}{\partial t} |\varphi(\mathbf{r}, t)|^2 = \frac{i\hbar}{2m} \nabla \cdot [\varphi^*(\mathbf{r}, t) \nabla \varphi(\mathbf{r}, t) - \varphi(\mathbf{r}, t) \nabla \varphi^*(\mathbf{r}, t)] = -\nabla \cdot \mathbf{j}. \quad (2.4)$$

Since  $|\varphi(\mathbf{r}, t)|^2$  is the probability density,  $\mathbf{j}$  is the probability flux/current which determines, with the addition of a factor of the electron charge  $-q$ , the charge current for electrons. Even for “stationary states,” a nonzero current will flow if the product  $\varphi(\mathbf{r})\nabla\varphi^*(\mathbf{r})$  has a nonzero imaginary part, such as the case for the (unnormalized) real-space wavefunction of a free electron  $\varphi(\mathbf{r}) = e^{ik\cdot\mathbf{r}}$ . The term “stationary” indicates only that the observable properties of the energy eigenstates do not evolve in time, but not necessarily that the particle is not moving. Think of the example of steady-state current flow in electrical circuits; the charge distribution and the current flow remain static, but electrons are moving.

### 2.2.1.2 Many Particle Systems, Indistinguishability, Fermions and Bosons, and the Density Matrix

*Spin* is another concept in quantum mechanics that distinguishes different states. Although an accurate description of spin requires a lot more mathematics, in this chapter, *spin* can be viewed as an *independent coordinate* describing the particle’s intrinsic angular momentum in units of  $\hbar$ . There are only two values of spin for electrons,  $\pm\hbar/2$ , which are commonly referred to as “spin up” and “spin down.” For other “particles,” there can be half-integer or integer spin values in terms of  $\hbar$ , and the number of allowed values can range from 1 upward. “Spinning” charge along with the orbital angular motion generates magnetic fields, the accumulation of which is the origin of macroscopic magnetic fields, with up spins producing fields in the opposite direction as down spins.

One essential property of fundamental particles such as electrons is their indistinguishability. For example, if you exchange the locations of any two electrons, there can be no observable difference. Mathematically, if  $\varphi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$  describes a two-particle system with one particle at location  $\boldsymbol{\mu}_1$  and the other at  $\boldsymbol{\mu}_2$ , within any coordinate system  $\boldsymbol{\mu}$ , there are only two allowed forms for this wavefunction to account for the indistinguishability under exchange,  $\varphi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \pm\varphi(\boldsymbol{\mu}_2, \boldsymbol{\mu}_1)$ , where the possible sign change has no effect on the observable properties of the state. Whether or not the sign changes under exchange, however, leads to two fundamental classes of particles. With the plus sign, the particles are considered to have “even parity” and are called *bosons*, for example, photons or hydrogen<sub>1</sub> atoms. With the negative sign on the other hand, the particles are considered to have “odd parity” and called *fermions*. As it turns out, half-integer spin particles such as electrons and protons are fermions, and integer spin particles such as hydrogen<sub>1</sub> atoms, not coincidentally composed of two fermions, as well as photons are bosons.

A critical property of fermions is that for  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}$  the two-particle wavefunction, which must satisfy  $\varphi(\boldsymbol{\mu}, \boldsymbol{\mu}) = -\varphi(\boldsymbol{\mu}, \boldsymbol{\mu})$  under exchange, must therefore vanish. This means that no identical fermions can occupy the same quantum mechanical state at the same time, in whatever coordinate system  $\boldsymbol{\mu}$  one may consider, which is the well-known *Pauli exclusion principle*. This results lead to, e.g., Fermi statistics for which the maximum occupation probability of quasi-single-particle states is unity.

Moreover, to the extent that the many-body fermion wavefunction is a continuous function of the coordinates  $\mathbf{r}$ , the probability of being found close together also will be less than otherwise expected. In other words, identical fermions such as electrons tend to avoid one another.

On the other hand, bosons are subject to no such restriction. Indeed, they actually tend to be found closer together than one would otherwise expect classically due to exchange considerations. That bosons, with their much different take on being close to each other, can be created from paired fermions is the basis for superconductivity.

The *density matrix* is a particularly useful tool to help describe the state of a statistical or probabilistic ensemble of particles. For many particles in quasi-single-particle states  $\alpha_n(\mathbf{r})$  but with no coherence among the single-particle states, the density matrix takes the form  $\rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \alpha_n(\mathbf{r}) \alpha_n^*(\mathbf{r}')$  in real-space coordinates, where  $f_n$  is the occupation probability of the state. For example,  $f_n$  would be obtained from Fermi statistics for electron energy eigenstates  $\alpha_n(\mathbf{r})$  in equilibrium. The diagonal elements  $\rho(\mathbf{r}, \mathbf{r})$  are just the total probability density as a function of  $\mathbf{r}$ . The off-diagonal elements represent the degree of overall coherence in the many-body quantum mechanical state between the points  $\mathbf{r}$  and  $\mathbf{r}'$  and will generally decrease with increasing magnitude of  $\mathbf{r} - \mathbf{r}'$ . In the classical limit, the off-diagonal elements vanish and the density matrix is diagonal. In bra-ket notation, the density matrix *operator* takes the form,  $\rho = \sum_n f_n |\alpha_n\rangle \langle \alpha_n|$ , which is the form of the projection operator except for the occupation probability weightings. (Consistent with this latter relation, if the occupation probabilities are identical,  $f_n = f$ , the density operator reduces to  $f$  times the identity matrix).

### 2.2.1.3 Crystals, Bloch's Theorem, and Band Structure

Most of the solids we are interested in are crystalline, in which the arrangement of atoms is highly ordered. Mathematically, they exhibit *translation symmetry* such that one can always assign three vectors  $\mathbf{R}_1$ ,  $\mathbf{R}_2$ , and  $\mathbf{R}_3$  for a real three-dimensional crystal such that the atom arrangement is identical if observed at arbitrary positions  $\mathbf{r}$  and  $\mathbf{r} + n_1 \mathbf{R}_1 + n_2 \mathbf{R}_2 + n_3 \mathbf{R}_3$ , where the  $n$  are integers and the  $\mathbf{R}$  are called the translation vectors. The set of points  $\{\mathbf{R} = n_1 \mathbf{R}_1 + n_2 \mathbf{R}_2 + n_3 \mathbf{R}_3\}$  defines a “Bravais lattice.” The parallelepiped constructed by  $\mathbf{R}_1$ ,  $\mathbf{R}_2$ , and  $\mathbf{R}_3$  is one representation of a lattice “unit cell,” which can be translated by the vectors  $\mathbf{R}_1$ ,  $\mathbf{R}_2$ , and  $\mathbf{R}_3$  to fill the entire space. The atoms contained within each unit cell and their positions are the “atomic basis.” The smallest possible atomic basis (e.g., containing two silicon atoms in silicon crystals), the associated lattices and translation vectors are called “primitive.” However, even the primitive unit cell of a given crystal is not unique. The unit cell that best exhibits the *point group symmetry* (symmetry under rotation, reflection, and inversion) about a lattice point is the “Wigner-Seitz” unit cell. For every real-space lattice, there is a corresponding “reciprocal space lattice” defined

by the reciprocal space vectors  $\mathbf{G}_1$ ,  $\mathbf{G}_2$ , and  $\mathbf{G}_3$  determined by the requirement that  $\mathbf{G}_i \cdot \mathbf{R}_j \equiv 2\pi\delta_{ij}$ . The Wigner–Seitz unit cell of the reciprocal lattice is known as the “(first) Brillouin zone.”

These symmetries are reflected in the quantum mechanical Hamiltonian  $H$ . As a result, after some work, it can be shown that the electron single-particle energy eigenfunctions of this system must be of the form,  $\alpha_{\mathbf{k},\sigma}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k},\sigma}(\mathbf{r})$ , where  $\mathbf{k}$  is a wavevector within the first Brillouin zone,  $\sigma$  labels the “energy band” as there will be multiple energy eigenstates with the same value of  $\mathbf{k}$  but with different energies, and  $u_{\mathbf{k},\sigma}(\mathbf{r})$  has the periodicity of the real-space primitive unit cell. This result is Bloch’s theorem, and the functions  $e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k},\sigma}(\mathbf{r})$  are called “Bloch functions.” The eigen-energy  $E_{\mathbf{k},\sigma}$  vs.  $\mathbf{k}$  relationship, which is continuous within each energy band, is known as the “band structure.” *Various properties of electrons in crystals can be obtained entirely from the band structure, or at least depend on the band structure*, including carrier velocity and rate of acceleration due to applied fields (and associated “effective mass”) as a function of “crystal momentum”  $\hbar\mathbf{k}$ , and “densities of states” per unit energy per unit volume for the Bloch states.

In terms of the types of crystals, an insulator results when there is a substantial energy “band gap” between an essentially fully occupied energy band—a “valence band”—where the electrons, loosely speaking, cannot move due to Pauli exclusion, and an essentially empty band—a “conduction band.” A metal results when there is one or more partially full (or equivalently, partially empty) bands with many electrons free to move. A semiconductor is a type of insulator, such that it remains a poor conductor when intrinsic, but with a relatively small band gap such that it is perhaps no longer a particularly good insulator either. However, the materials we commonly refer to as semiconductors can be converted to good conductors by adding certain types of impurity atoms called “donors” and “acceptors,” by attracting charge carriers to the surface by application of electric fields, or by absorption of light. Of importance, instead of keeping track of the occupied electron states in the valence band, it is generally more productive to keep track of any empty ones. In a formally defined way, one can associate positively charged quasi-particle “holes” with these empty states. Thus, we can consider two types of charge carriers, electrons and holes. Often electrons and holes are represented simply as “e” and “h,” respectively. For example, electron–hole pair generation due to absorption of light becomes “e–h pair generation.” Similarly, negatively and positively charged regions can be referred to as “n type” and “p type,” respectively.

### 2.2.1.4 Tight-Binding Approximation and Current Flow

In practice, the Schrödinger’s equation can be somewhat daunting to solve and even more so when considering many-body exchange effects. Therefore, there are many approximations developed to simplify the mathematics while still yielding reliable physics. The tight-binding approximation is one of them, which typically takes an outer-shell electron energy eigenstate of freestanding atoms (their “valence”



“orbitals”), or a reasonable facsimile thereof, as a basis set for describing the actual emergent electronic states of the crystal. Using this basis, Schrödinger’s equation can be transformed from a differential equation in real space, Eq. (2.2), to a matrix equation in orbital space with interorbital “hopping potentials.” Although intrinsically incomplete with a finite number of orbitals, this approximation can be sufficient to yield accurate results, while also providing an efficient discretization of Schrödinger’s equation that is compatible with numerical analysis and retaining intuitive clarity.

For example, assume that the atoms are positioned at  $\{\mathbf{R}\}$ , each with  $n_{\mathbf{R}}$  orbitals  $|\varphi_{\mathbf{R},i}\rangle$  where  $i = 1, 2, \dots, n_{\mathbf{R}}$ , which (for convenience although not a necessity) have been modified enough to be orthogonal among the atoms. Any state spanned from this basis can be written as

$$|\Phi\rangle = \sum_{\mathbf{R}} \sum_{i=1\dots n_{\mathbf{R}}} C_{\mathbf{R}i} |\varphi_{\mathbf{R}i}\rangle \quad (2.5)$$

Substituting Eq. (2.5) into Eq. (2.3b) and operating on both sides with  $\langle \varphi_{\mathbf{R}',i'} |$  produce

$$\sum_{\mathbf{R}} \sum_{i=1\dots n_{\mathbf{R}}} C_{\mathbf{R}i} \langle \varphi_{\mathbf{R}',i'} | H | \varphi_{\mathbf{R}i} \rangle = E C_{\mathbf{R}'i'}, \quad (2.6)$$

which defines an eigenvalue problem of the *Hamiltonian matrix* defined by the matrix elements or hopping potentials,  $H_{\mathbf{R}'i',\mathbf{R}i} = \langle \varphi_{\mathbf{R}',i'} | H | \varphi_{\mathbf{R}i} \rangle$ . The coefficients  $C_{\mathbf{R}i}$  are the eigenvectors/eigenfunctions or “tight-binding wavefunctions” in this tight-binding space. These matrix elements can be obtained, e.g., from the first principles using the density functional theory (DFT) or by adjusting the hopping potentials  $H_{\mathbf{R}'i',\mathbf{R}i}$  to match the results of other band structure calculation methods and/or experimental band structure characteristics. Once the hopping potentials are known, complex structures can be created and real-space variations in potentials can be modeled via additions to the onsite ( $\mathbf{R}'i = \mathbf{R}i$ ) and, in the case of exchange interactions (see below), interatom and interorbital hopping potentials.

Following the same approach used in deriving Eq. (2.4), the tight-binding version of the probability current continuity equation

$$\sum_{\mathbf{R}} \sum_{i=1\dots n_{\mathbf{R}}} \frac{2}{\hbar} \text{Im}(C_{\mathbf{R}'i'}^* H_{\mathbf{R}'i',\mathbf{R}i} C_{\mathbf{R}i}) = \frac{\partial}{\partial t} |C_{\mathbf{R}'i'}|^2 \quad (2.7)$$

can be derived from Eq. (2.6). This result implies that the current flowing between orbitals  $\mathbf{R}i$  and  $\mathbf{R}'i'$  is proportional to the imaginary part of the product  $C_{\mathbf{R}'i'}^* H_{\mathbf{R}'i',\mathbf{R}i} C_{\mathbf{R}i}$ .

### 2.2.2 Interlayer Electron–Hole Exciton Condensation

Excitons are simply an electron and a hole bound together by their Coulomb interaction [2, 3]. However, the exciton, composed of two half-integer spin particles, has integer spin and, thus, is a boson, much like that Cooper pairs of electrons in superconductors are bosons [4]. This bosonic nature allows for the possibility of “condensation” where the bosons occupy the same quantum mechanical state, which is not possible for the individual fermions. Such condensation is sometimes considered as a state of matter in parallel with the commonly known solid, liquid and gas states [5, 6]. Condensation, in turn, can lead to novel macroscopic physical effects such as superfluidity and superconductivity [7, 8]. However, the formation of condensates is something generally reserved for cryogenic temperatures [9–11]. Such pairing is known as Bose–Einstein condensation in the limit of tightly bound pairs and Bardeen–Cooper–Schrieffer (BCS) pairing in the weak pairing limit. In the exciton system, perhaps the strongest effects may be found near the BEC–BCS crossover.

Commonly, excitons are composed of electrons and holes in the same material (although different energy bands, of course), such as created by the absorption of light (photons) in bulk semiconductors. These excitons can be described as “spatially direct.” The excitons considered in this work are “spatially indirect” interlayer excitons, where the electrons and the holes *nominally* are spatially separated in different material layers. Generally the electrons and holes of spatially indirect excitons reside in quasi-two-dimensional systems, such as double quantum wells or, of particular interest here, various 2D materials [12–16]. Condensation of indirect excitons has been observed experimentally in double quantum wells of III–V materials at cryogenic temperatures ( $\sim$ mK) and high magnetic fields ( $\sim$ T) [17–20]. The condensation was explained theoretically via a coherent interlayer many-body exchange interaction, which is a purely quantum effect arising from the indistinguishability of electrons as noted above, with the holes modeled as just the absence of electrons, as is the case in reality [13]. Multiple expected novel interlayer transport behaviors for such condensation were observed experimentally, including greatly enhanced interlayer transmission and near-perfect Coulomb drag between layers [20].

Graphene was the first truly 2D material to be considered, the unique gapless and linear dispersion relation of which allowed physicists to carry on experimental research which had only been theoretically predicted [21, 22]. Applying the interlayer many-body exchange interaction to the bilayer graphene system, the group of MacDonald et al. estimated that the interlayer e–h exciton condensation may be possible at room temperature [14], due to not only the unique physical properties of graphene including symmetric electron and hole band structures [23, 24] (which otherwise are obtained only under high magnetic fields with Landau levels in the conduction band of III–V materials where a partially empty Landau level serves as a hole band [18]) but also the possibility of placing the physically 2D layers much closer together and within a low-dielectric constant environment that is

primarily extrinsic to the graphene and, thus, separately optimizable. Recent calculations by the same group suggest that room-temperature condensate formation in such graphene bilayers will be more difficult than initially expected, but may still be possible [25]. Because graphene was the first material system considered, it currently hosts most of the theoretical and experimental efforts toward “pseudospintronic devices.”

However, room-temperature condensate formation may be more readily possible employing other 2D material systems such as monolayer transition metal dichalcogenides (TMDs) [16]. TMDs are compounds composed of a transition metal plus a column VI material such as  $\text{MoS}_2$ ,  $\text{WTe}_2$ , and  $\text{WSe}_2$ .

If room-temperature condensates can be formed in these novel 2D materials, it is expected that the same novel transport properties observed in III–V materials at very low temperatures and high magnetic fields [20] could be reproduced in these bilayer 2D material systems absent magnetic fields.

### 2.2.3 *Pseudospin and Pseudospintronics*

Pseudospin can be defined for any parameter with only two options. In the bilayer systems of interest here, the “which-layer” degree of freedom for the electrons can be described as a “pseudospin” consisting of “top layer” (T) and “bottom layer” (B) analogous to “spin-up” and “spin-down” electrons. Single-electron states are therefore a mixture of “pseudospin-up” and “pseudospin-down” states coupled by any bare interlayer coupling and, of critical importance here, the interlayer many-body exchange interaction. More generally, in a collective state consisting of many single-particle electron states, the magnitude and phase of top-to-bottom interlayer density matrix  $\rho(\mathbf{R}_T, \mathbf{R}_B) = \sum_{\varphi} f_{\varphi} \varphi(\mathbf{R}_T) \varphi^*(\mathbf{R}_B)$  reflect an overall coherence strength and phase relationship between the top and bottom layers and are described as a collective “pseudospin strength” and “pseudospin phase,” respectively [13, 14]. This coherent pseudospin state is analogous to an actual collective spin state composed of coherent spin up and spin down contributions. An essential property of exciton condensation is greatly enhanced—even entirely self-consistently supported in the conceptual “spontaneous” limit—interlayer coherence, i.e., density matrix/pseudospin strength.

To go further, we introduce the Hartree–Fock potential, which addresses the above-discussed many-body exchange interaction within a single-particle framework using the density matrix and the Coulomb interaction among the electrons. It is employed to explain and characterize the formation of the condensation, as well as associated novel transport effects. For graphene, the conduction and valence band properties can be reasonably represented considering only one spin degenerate  $2p_z$  orbital per carbon atom. The tight-binding version of the Hartree–Fock model, explicitly considering only interlayer exchange, is for the top layer [26],

$$H_{\text{TB}}\varphi_{\beta}(\mathbf{R}_T) + V(\mathbf{R}_T)\varphi_{\beta}(\mathbf{R}_T) + \sum_{\mathbf{R}_B} V_{\text{Fock}}(\mathbf{R}_T, \mathbf{R}_B)\varphi_{\beta}(\mathbf{R}_B) = E_{\beta}\varphi_{\beta}(\mathbf{R}_T), \quad (2.8)$$

where

$$V_{\text{Fock}}(\mathbf{R}_T, \mathbf{R}_B) = -\frac{e^2}{4\pi\epsilon|\mathbf{R}_T - \mathbf{R}_B|} \sum_{\beta'} f_{\beta'}\varphi_{\beta'}(\mathbf{R}_T)\varphi_{\beta'}^*(\mathbf{R}_B), \quad (2.9)$$

with a similar expression for the bottom layer. Here the  $\varphi_{\beta}(\mathbf{R})$  are the quasi-single-particle tight-binding energy eigenstates, where  $\beta$  labels the energy band  $\alpha$ , wavevector  $\mathbf{k}$ , and real spin  $s$ .  $H_{\text{TB}}$  is the nominally single-particle tight-binding Hamiltonian with only intralayer and (bare) interlayer hopping between nearest-neighbor atoms (although many-body exchange and correlation effects are implicitly included in the modeled band structure). The second term on the left-hand side is the “Hartree” term, where  $V(\mathbf{R})$  is the electrostatic potential energy at  $\mathbf{R}$  including potential energy contributed by both externally applied electric fields and the charge distribution within the layers (and even implicitly variations in the intralayer exchange interaction, which would need to be more carefully considered for self-consistent calculations of layer carrier densities as a function of applied gate voltages). The third term, or the “Fock” term, on the left-hand side of Eq. (2.8) is the exchange interaction, which is proportional to the interlayer density matrix as shown in Eq. (2.9), where we have approximated the Coulomb interaction using an effective dielectric constant that represents the interlayer dielectric and—even dominated by—the dielectrics above and below the 2D layers and self-consistent dynamic free-carrier screening [25]. Physically, the exchange interaction reduces the otherwise expected Coulomb interaction, consistent with the electrons being further away from each other than expected classically.

For this multistate  $\beta$  system, extending Eq. (2.7) for single states, the *total* interlayer current is calculated as in the presence of the exchange interaction as

$$j(\mathbf{R}_T, \mathbf{R}_B) = \frac{2e}{h} \text{Im} \left\{ \sum_{\beta} f_{\beta}\varphi_{\beta}^*(\mathbf{R}_T) [H_{\text{bare}}(\mathbf{R}_T, \mathbf{R}_B) + H_{\text{Fock}}(\mathbf{R}_T, \mathbf{R}_B)] \varphi_{\beta}(\mathbf{R}_B) \right\} \quad (2.10)$$

With a little work, it can be shown both that the total interlayer current due to the interlayer Fock interaction alone is identically zero and that the remaining contribution reduces to

$$j(\mathbf{R}_T, \mathbf{R}_B) = \frac{2e}{h} H_{\text{bare}}(\mathbf{R}_T, \mathbf{R}_B) \text{Im} [\rho^{\dagger}(\mathbf{R}_T, \mathbf{R}_B)]. \quad (2.11)$$

Indeed, this is the expression for current flow in a multistate system in terms of the density matrix regardless of whether or not the Fock exchange interaction is

considered. However, the interlayer exchange interaction can greatly increase the pseudospin/interlayer density matrix  $\rho(\mathbf{R}_T, \mathbf{R}_B)$ . Unlike the terminology *spintronics*, the pseudospintronic devices do not rely on manipulating any real spins. In the bilayer graphene system, different real spins can even be decoupled as just two identical channels for interlayer current. However, such a decomposition cannot be performed in TMD systems, where strong spin–orbit coupling breaks the degeneracy and mixes the spin states within individual energy eigenstates [27, 28].

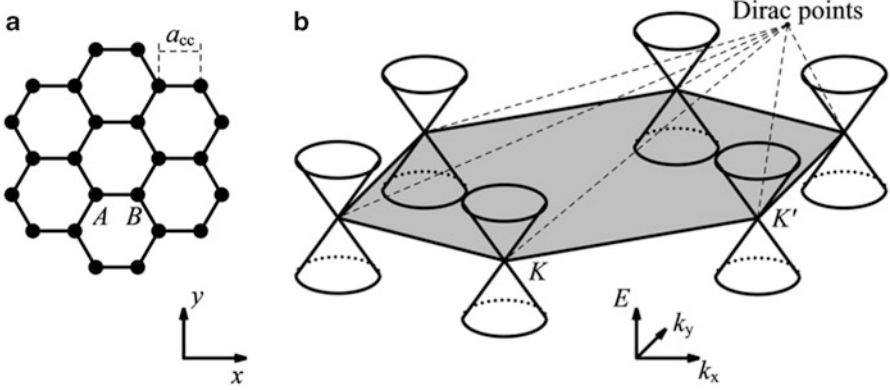
## 2.3 Pseudospintronic Devices and Analysis

In this section, we will consider pseudospintronic devices: BiSFET, BiSJT, and Coulomb drag devices. The essential physics and device concepts should be independent of which bilayer system is considered. However, much of the analysis of this work has been performed for graphene systems for reasons already noted, so we provide detailed information on this system. In addition, we first consider condensate behavior in “bulk” (large area) graphene–dielectric–graphene systems, to provide a reference for device structures and behavior.

### 2.3.1 Condensate Formation in Bulk Graphene Bilayers

#### 2.3.1.1 About Graphene

Graphene is a single layer of graphite. Pristine graphene has a honeycomb structure, where carbon–carbon bond length is  $a_{cc} \approx 1.42 \text{ \AA}$  and, within a tight-binding description, a nearest-neighbor nominally bare hopping potential of  $t_0 \approx -2.7 \text{ eV}$ . The primitive cell includes two atoms labeled by A and B, as shown in Fig. 2.1a, and the set of A atoms and the set of B atoms constitute two sublattices. The linear band structure in graphene resembles the dispersion relationship for light and relativistic particles, as shown in Fig. 2.1b and, therefore, electrons in graphene are sometimes called the “massless Dirac fermions” [21]. However, the fixed carrier velocity magnitude in graphene is roughly  $10^8 \text{ cm/s}$ , only one three-hundredth the speed of light in vacuum. Specifically, the band structure of graphene has cone-shaped valence and conduction bands (the “Dirac cones”) touching each other at the pinnacle (the “Dirac point”), located at the corners of the first Brillouin zone. There are two such cone-shaped pairs in total, centered at the  $K$  and  $K'$  points. (The “pseudospin” terminology also has been applied to the “which-atomic-sublattice” and to the “which-valley” degrees of freedom. In monolayer TMDs, while there are band gaps and the carriers are massive, the valleys are also located at the  $K$  and  $K'$  points. However, the valley pseudospin and real spin are coupled. The band edges are also split between real spin up and down, particularly in the valence band, but with the  $K$  and  $K'$  valleys having opposite spins at the valence band edge [29]).

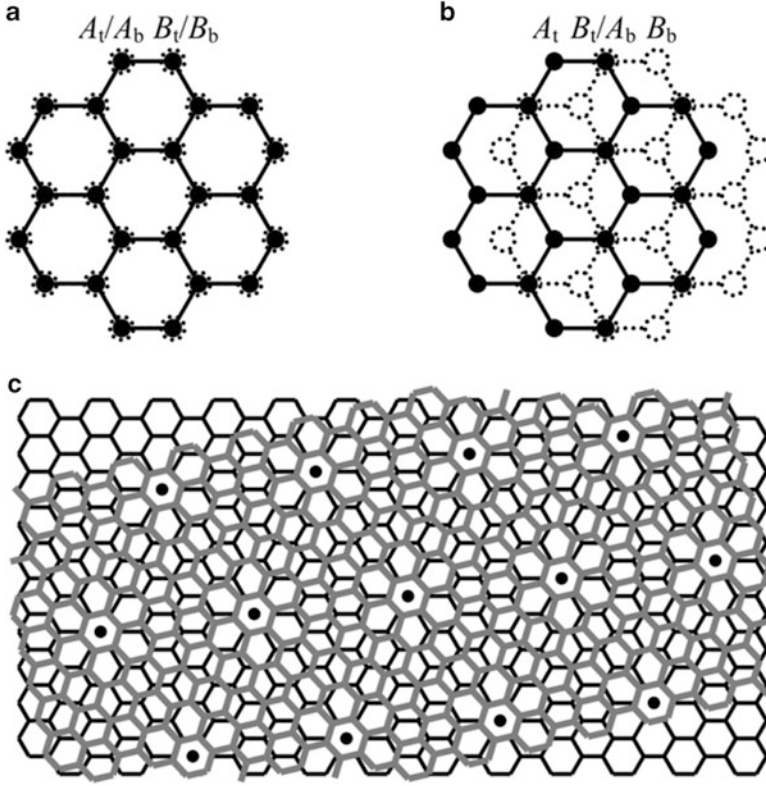


**Fig. 2.1** (a) A graphene flake with a perfect hexagonal honeycomb structure. A and B label the two atoms within the primitive unit cells and the associated A and B sublattices. (b) The conical band structure of graphene around the  $K$  and  $K'$  points in the hexagonal Brillouin zone. Note that the orientations of the Wigner–Seitz unit cells in real space and in momentum space are offset by  $30^\circ$

For our two-graphene-layer system, there are several stacking patterns that can be considered, as shown in Fig. 2.2. Even assuming rotationally aligned graphene for the moment, there are multiple translational alignments that can be considered. “Bernal stacking” as shown in Fig. 2.2b refers to the situation that A sublattice in one layer is aligned with B sublattice in the other, as naturally occurs in graphite. However, particularly with an intervening dielectric—as modeled for our pseudospin device analysis—and in terms of interatomic coupling, one may also consider hexagonal stacking (Fig. 2.2a), which consists of two perfectly aligned graphene layers with A sublattice upon A sublattice and B sublattice upon B sublattice. In terms of bare coupling through a dielectric, one may even consider just A sublattice-to-A sublattice *or* B sublattice-to-B sublattice coupling or linear combinations of all of these. The layers also may be rotated with respect to each other, particularly when one is physically placed upon another or upon an intervening dielectric. For commensurate rotation (Fig. 2.2c), one layer is rotated with respect to the other, but a global periodicity is still preserved for the bilayer system, which is convenient for numerical analysis, although on a larger geometric scale which is less convenient. The global periodicity observed is also named the M $\acute{o}$ ire pattern. The possible commensurate rotation angles are infinite but still can be mathematically enumerated by two integer indices [30].

### 2.3.1.2 Simulation Technique

A full-quantum Hartree–Fock simulator can be developed from Eq. (2.8) to study the formation of exciton condensates in bulk and to simulate transport properties in the presence of the condensates in nanoscale bilayer systems as to be



**Fig. 2.2** Stacking patterns of bilayer graphene: (a) hexagonal stacking, (b) Bernal (A–B) stacking, and (c) commensurate rotational stacking. In (c), the two rotationally misaligned graphene honeycomb lattices are sketched with different colors and line widths, and the resulting honeycomb superlattice is represented by *dots*, which also shows the Moiré pattern resulting from the angular rotation. The relative rotation between layers shown here is  $13.17^\circ$ , producing a total of 76 carbon atoms per unit cell counting both layers

considered later. For graphene layers with periodicity, either rotationally aligned or commensurately rotated, Bloch's theorem can be applied to Eq. (2.8) such that the tight-binding time-independent Schrödinger's equation takes the form,

$$H_{TB}\varphi_\beta(\mathbf{R}_T) + V_T\varphi_\beta(\mathbf{R}_T) + \sum_{\mathbf{R}_B} A_{\text{Fock},\beta}(\mathbf{R}_T, \mathbf{R}_B)\varphi_\beta(\mathbf{R}_B) = E_\beta\varphi_\beta(\mathbf{R}_T), \quad (2.12)$$

$$\begin{aligned} A_{\text{Fock},\beta}(\mathbf{R}_T, \mathbf{R}_B) &= \sum_{\mathbf{R}} V_{\text{Fock}}(\mathbf{R}_T, \mathbf{R}_B + \boldsymbol{\xi})e^{i\mathbf{k}\cdot\boldsymbol{\xi}} \\ &= \sum_{\mathbf{k}'} C(\mathbf{R}_T - \mathbf{R}_B, \mathbf{k} - \mathbf{k}') \sum_{\alpha', s'} f_{\beta'}\varphi_{\beta'}(\mathbf{R}_T)\varphi_{\beta'}^*(\mathbf{R}_B), \end{aligned} \quad (2.13)$$

$$C(\mathbf{R}_T - \mathbf{R}_B, \mathbf{k} - \mathbf{k}') = - \sum_{\mathbf{R}} V_{\text{col}}(\mathbf{R}_T, \mathbf{R}_B + \boldsymbol{\xi}) e^{i(\mathbf{k} - \mathbf{k}') \cdot \boldsymbol{\xi}}, \quad (2.14)$$

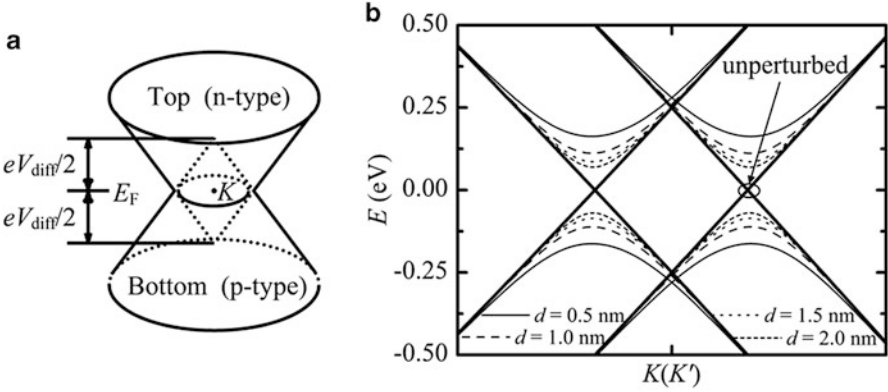
where  $\boldsymbol{\xi}$  are the lattice vectors and  $\mathbf{R}_T$  and  $\mathbf{R}_B$  are the locations of the top- and bottom-layer atoms within the primitive cell [31, 32].  $V_{\text{col}}(\mathbf{R}_T, \mathbf{R}_B + \boldsymbol{\xi})$  describes the Coulomb interaction within (unlike what is shown in Eq. (2.9)) a possibly nonuniform dielectric environment. The time-independent Schrödinger's equation is consequently reduced to the primitive unit cell size. We solve this equation using an iterative method in which the exchange interaction and energy eigenstates  $\beta$  are obtained self-consistently. (If no actual bare coupling is considered, we “seed” the simulation with a small bare interlayer coupling for the first step only.) However, the calculation of density matrix  $\rho$  involves the summation over all states  $\beta$ , and the Hamiltonian contains the long-range (non-nearest-neighbor) interlayer Fock interaction. The resulting matrix equation is dense (has many nonzero elements) and different for each different energy eigenstate  $\beta$ . Equations (2.13) and (2.14) imply that the Fourier transform of the Coulomb interaction can be pre-calculated, making each single  $\mathbf{k}$  state separable and assignable to different CPUs for parallel computing. However, even with supercomputing resources, this calculation can be computationally challenging for large unit cells under commensurate rotations.

### 2.3.1.3 Characterization and Properties of the Bulk Condensate

Consider first two rotationally aligned graphene layers with no bare interlayer coupling. In this case, the translational alignment is of no real consequence. Electrons in one layer and holes in the other are conceptually created via external gates. For simulation purposes, the required carrier concentration are adjusted by on-layer electrostatic potentials  $V(\mathbf{R}_T)$  and  $V(\mathbf{R}_B)$  as in Eq. (2.8) with respect to the equilibrium Fermi level  $E_F$ , which is chosen as the zero energy reference ( $E_F \equiv 0$ ) for simplicity. Setting  $V(\mathbf{R}_T) = -V(\mathbf{R}_B) = V_{\text{diff}}/2$  produces equal electron and hole concentrations, which is ideal for condensate formation. The resulting band structure of such a bilayer graphene system absent significant interlayer coupling is shown in Fig. 2.3a for the interpenetrating cones of the conduction band of the n-type layer and valence band of the p-type layer. The intersection of the cones is at the Fermi level.

The formation of an exciton condensate is accompanied by a band-gap opening/anti-crossing where the bands otherwise would have crossed, as shown in the simulation results of Fig. 2.3b as a function of interlayer separation [14, 26]. For large separation, the band structure is unaffected and therefore remains two essentially decoupled and overlapping Dirac cones representing the top and bottom layers. As the separation decreases and the Coulomb interaction increases, a band gap opens up which can be understood in this way: There can be no interlayer coherence in the absence of interlayer coupling of any type. That is, the interlayer pseudospin is zero. However, if there is any exchange interaction, there will be

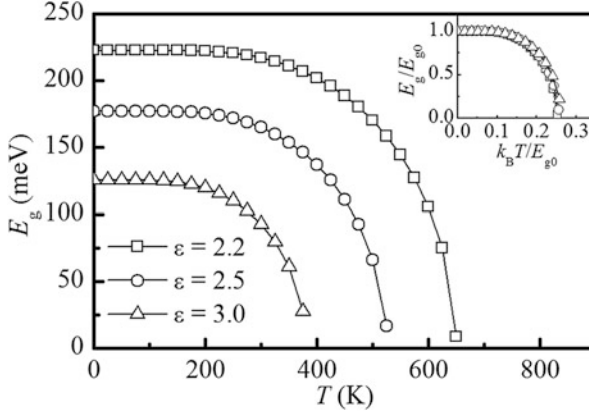




**Fig. 2.3** Bilayer graphene band structures. (a) Without any interlayer interaction, with the n-type layer valence band and p-type layer conduction band suppressed for clarity. (b) As calculated with only the many-body exchange interaction at 0 K as a function of interlayer separation in units of nm, with an *effective* dielectric constant  $\epsilon$  of 2.2. The latter represents the actual dielectric constants of the interlayer dielectric plus—even dominated by—those above and below the graphene layers, plus free-carrier screening which should be calculated self-consistently, but which is beyond the scope of this work [25]. The band structure in (b) is plotted along a line of  $\mathbf{k}$  vectors that crosses the  $K$  point. The full band structure can be envisioned as that of (b) pivoting around a  $z$ -axis that passes through the  $K$  point. While interlayer bare coupling can lead to a band gap opening [33], the opening of such a “condensate gap” does not rely on interlayer bare coupling [14, 26]

interlayer coherence/nonzero pseudospin. In the end, we can find a self-consistent solution for interlayer exchange coupling and interlayer coherence even in the absence of bare interlayer coupling. (Bare coupling still would be required to allow any interlayer current flow, per Eq. (2.11)). This latter solution represents the condensed state, a “spontaneous” one in this case, with coherent interlayer coupling between the n-type and p-type layers, mediated by the exchange interaction. Thus, the excitonic condensate does not result simply from the interlayer Coulomb interaction, but rather from the many-body exchange correction to the Coulomb interactions when, and only when, there is interlayer quantum coherence. The band splitting/anti-crossing about the Fermi level and associated lowering of the occupied states near the anti-crossing reduce the overall energy of the system, making the condensed state the energetically favored one [26].

The establishment and strength of the condensate/pseudospin is subject to a positive feedback loop: the stronger the exchange interaction, the stronger the interlayer pseudospin; the stronger the exchange interaction, the stronger the pseudospin; and so forth and so on and vice versa. As a result, the condensate is a sensitive function of dielectric environment, layer separation, and, as shown in Fig. 2.4, temperature [26]. The result shows a clear “critical temperature”  $T_c$  for the formation of these condensates. Moreover, the critical temperature is strongly correlated to the zero-temperature band gap,  $4k_B T_c \approx E_{g0}(E_g \text{ at } T=0)$ , where  $k_B$  is Boltzmann’s constant and  $T_c$  is the critical temperature in Kelvin. However, the

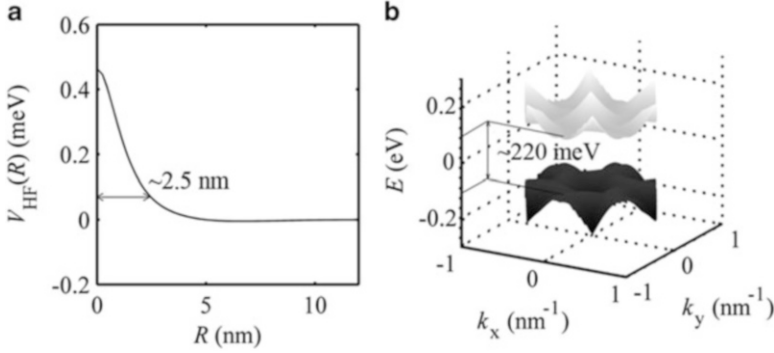


**Fig. 2.4** Temperature dependence of the induced band gap/anti-crossing for spontaneous condensates. The layer separation about the interlayer dielectric is  $d = 1$  nm. Various effective dielectric constants were considered as shown. Balanced electron and hole concentrations are created by  $V_T = -V_B = 0.25$  V. The *inset* is plotted in energies normalized by induced zero-temperature band gap  $E_g(T=0) \equiv E_{g0}$

shape of these band gaps vs. temperature curves is not unique to excitonic condensates but rather is much as exhibited in conventional superconductors.

There is also a Kosterlitz–Thouless transition temperature for 2D systems to consider, which is beyond the scope of this work to explain but which is expected to be approximately  $8k_B T_{K-T} = |E_F - E_D|$  for the graphene system, where  $E_D$  is the Dirac point energy at the pinnacle of the Dirac cones; the condensate can only exist if both conditions are met [14]. However, the zero band gap and low graphene density of states make it relatively easy to move the Fermi level well into the bands with limited carrier concentrations and, thus, manageable gating fields. This condition can be satisfied for electron and hole concentrations of somewhat less than  $5 \times 10^{12}/\text{cm}^2$  for condensation at  $T = 300$  K [34]. The carrier density in graphene layers has been electrostatically modulated to as high as  $10^{13}/\text{cm}^2$  using independent gates [35].

We also have found that the strength of the spontaneous condensate depends little on the pattern of interlayer alignment, including rotational misalignment even when the size of the resulting unit cell is large compared to the in-plane coherence length of the condensate [32], as seen in Fig. 2.5. The latter insensitivity results because if we consider exchange coupling between top and bottom layers whose  $K$  valleys are separated by  $\mathbf{K}_B - \mathbf{K}_T = \Delta\mathbf{K}$ , the resulting uniform phase shift of  $e^{+i\Delta\mathbf{K}\cdot\mathbf{R}_B}$  introduced within the bottom layer to each wavefunction  $\beta$  is canceled by a corresponding uniform phase shift  $e^{-i\Delta\mathbf{K}\cdot\mathbf{R}_B}$  introduced into the exchange interaction operating on the bottom layer as required to determine the effect on the top layer and vice versa for the exchange interaction acting on the bottom layer. However, the rotation of one layer with respect to the other can greatly alter the interlayer bare hopping.



**Fig. 2.5** (a) Magnitude of the Fock exchange interaction  $V_{\text{HF}}(\mathbf{R}_T, \mathbf{R}_B)$  as a function of  $R$ , the magnitude of the in-plane component  $|\mathbf{R}_T - \mathbf{R}_B|_{\text{in-plane}}$ . The interlayer separation is  $d = 1$  nm, the effective dielectric constant is  $\epsilon = 2.2$ , and  $T = 0$  K. The Fock interaction, limited both by in-plane coherence in the density matrix and the Coulomb interaction, is only significant over a radius in  $R$  of approximately 3 nm or less. (b) Band structure of the conduction and valence bands in the first Brillouin zone of commensurately rotated bilayers with a  $2.65^\circ$  rotation angle. The shown condensate-induced band gap is essentially unchanged as a result of the rotation. The real-space unit cell is of approximately 6 nm on an edge, containing 1,876 atoms, and there is a correspondingly small Brillouin zone. All other parameters are the same as in (a)

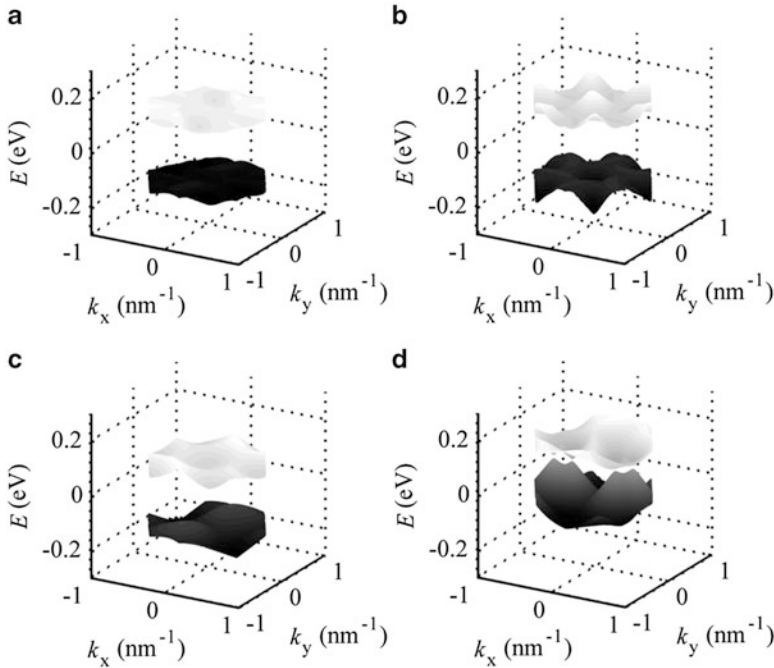
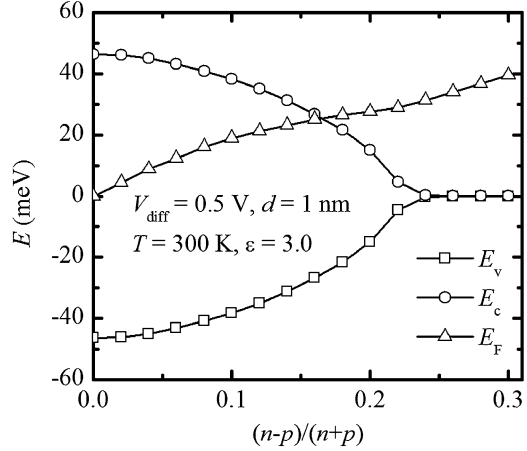
Balanced charge distributions are optimal for exciton condensate formation as already noted. Conversely, charge imbalance can impact condensate formation negatively. The result of Fig. 2.6 shows that small charge imbalances are tolerable, but larger charge imbalances can completely prevent condensation [26]. Thus, strong “charge puddling,” charge imbalance created by electrostatic potential variations resulted from nonideal substrates, impurities, etc. [36], is to be avoided. We note, however, that variations in the electrostatic potentials on scales comparable to or smaller than the coherence length have a more limited effect, as shown in Fig. 2.7 [32].

## 2.3.2 Pseudospin Transport Properties

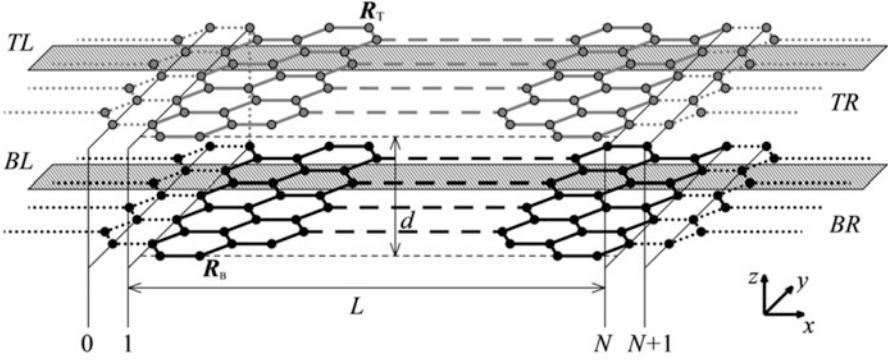
### 2.3.2.1 Simulation Technique

The simulated structure for modeling transport properties is shown in Fig. 2.8. This four-terminal structure includes two coupled graphene sheets with both bare coupling and Fock exchange interaction within, and only within, a channel of length  $L$ , through an interlayer dielectric with interlayer separation  $d$  within an effective dielectric environment  $\epsilon$ . The channel is connected to four semi-infinite, noninteracting leads. The pattern of interlayer bare coupling, as discussed above, does not have a great effect on the condensate formation but does influence the interlayer current profoundly according to Eq. (2.11). However, this pattern

**Fig. 2.6** Simulated energy band edges and Fermi level as a function of carrier imbalance between top-layer electron and bottom-layer hole concentrations for graphene bilayers separated by 1 nm through an interlayer dielectrics, with an effective dielectric constant of  $\epsilon = 3.0$  and  $V_{\text{diff}} = 0.5$  V at 300 K



**Fig. 2.7** Typical band structures for condensates in graphene bilayers with short-range disorder with a commensurate rotation of  $\theta = 2.65^\circ$  making the unit cell larger than the coherence length of the exchange interaction.  $d = 1$  nm,  $\epsilon = 2.2$ , and  $T = 0$  K. The disorder potentials are characterized by standard deviation  $\sigma_v$  and correlation length  $l$ , both extending less than one unit cell. (a)  $l \approx 0.6$  nm and  $\sigma_v \approx 0.4$  V resulting in  $E_g \approx 225$  meV, (b)  $l \approx 2.5$  nm and  $\sigma_v \approx 0.1$  V resulting in  $E_g \approx 214$  meV, (c)  $l \approx 2.5$  nm and  $\sigma_v \approx 0.4$  V resulting in  $E_g \approx 178$  meV, and (d)  $l \approx 2.5$  nm and  $\sigma_v \approx 0.5$  V resulting in  $E_g \approx 27$  meV



**Fig. 2.8** Simulated structure. Two graphene layers (infinite in the  $y$  direction) are coupled in a channel of length  $L$  and with dielectric of thickness  $d$ . The effective interlayer dielectric constant is  $\epsilon$ . Primitive cells are represented by the *longitudinal shaded stripes*, and atoms are located within slices with consecutive integer-indexed labels, starting from 1 and ending with  $N$  in the channel. There are four semi-infinite, perfectly absorbing/injecting leads connected to the graphene layers at the top left, bottom left, top right, and bottom right, labeled by TL, BL, TR, and BR, respectively

becomes at least conceptually tunable through interlayer dielectrics, although bringing about additional challenges to fabrication, a topic that will be discussed later. However, for mixed coupling patterns, either A–A hopping (coupling between the A-sublattice atoms) or, equivalently, B–B hopping is expected to dominate the contributions to interlayer current flow because of the way they align with the natural pattern of the “spontaneous” condensate [37]. In this work, we assume A–A coupling for modeling purposes. The transport is assumed along the direction of armchair chains of atoms.

In such a structure, the single-electron states  $\beta$  are now characterized by the lead of injection  $\alpha$ , the predetermined/chosen energies of injection  $E$ , the transverse wavevector  $k_y$  (also referred to as a “mode” in transport problems), and the real spin  $s$  (which can be neglected here as in the bulk for the graphene system). A thorough tutorial of quantum transport simulations can be found in Supriyo Datta’s well-known book on the subject [38]. Conceptually, an electron can be injected from any lead and then be reflected into the same lead or be transmitted to other leads. The wavefunction including the injected, reflected, and transmitted states in the leads, as well as the potentially complicated wave profile in the channel, is called the “scattering state.”

To calculate scattering states, we start with a Bloch state that describes an incoming wave in one lead, for which the wavevector  $\mathbf{k}$  is determined from the  $E$ – $\mathbf{k}$  dispersion relation in the leads and subject to the required direction of propagation being *into* the lead. In this open simulation region, the *incident* current (not counting reflection) for each of the scattering states  $\beta$ , as calculated from Eq. (2.7), is necessarily normalized consistent with the ballistic Landauer–Büttiker limit for conductance per transverse mode  $e^2/h$ . The nominal number of occupied transverse modes per unit width is then  $4k_F/\pi$  including a factor of 4 for valley and spin

degeneracy, where  $k_F$  is the Fermi wavevector,  $k_F = E_F/(\hbar v)$ ,  $E_F$  is the Fermi energy, and  $v$  is the fixed carrier velocity in graphene [34]. With the device separated into slices in the longitudinal direction, the Schrödinger's equation, Eq. (2.12), can be then generalized into a slice-by-slice form,  $\sum_j \mathbf{H}_{ij} \boldsymbol{\varphi}_j = E \boldsymbol{\varphi}_i$ , where  $\boldsymbol{\varphi}_i$  represents the wavefunction of slice  $i$  (a  $2 \times 1$  vector representing the top and bottom atoms) and  $\mathbf{H}_{ij}$  couples slices  $i$  and  $j$  via bare and any Fock exchange interaction.

The next step is to determine the boundary conditions between leads and channel, i.e., the variation of the wavefunction between slices 0 and 1 for the left leads and  $N$  and  $(N+1)$  for the right leads. Consider a left lead as the lead of injection. On the boundaries between the semi-infinite leads and the channel, beyond the region of interlayer coupling within channel, the injected and reflected waves must satisfy [31]

$$\begin{aligned} \boldsymbol{\varphi}_0 &= \boldsymbol{\varphi}_{0,i} + \boldsymbol{\varphi}_{0,r} = \boldsymbol{\varphi}_{0,i} + \mathbf{P}_{10,r} \boldsymbol{\varphi}_{1,r} = \boldsymbol{\varphi}_{0,i} + \mathbf{P}_{10,r} (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_{1,i}) \\ &= \mathbf{P}_{10,r} \boldsymbol{\varphi}_1 + (\mathbf{I} - \mathbf{P}_{10,r} \mathbf{P}_{01,i}) \boldsymbol{\varphi}_{0,i} \end{aligned} \quad (2.15)$$

Here,  $\boldsymbol{\varphi}_{k,i}$  and  $\boldsymbol{\varphi}_{k,r}$  are incident (i) and reflected (r) components of the wavefunction on slice  $k$ , respectively, such that  $\boldsymbol{\varphi}_k = \boldsymbol{\varphi}_{k,i} + \boldsymbol{\varphi}_{k,r}$ .  $\mathbf{P}_{01,i}$  and  $\mathbf{P}_{10,r}$  are  $2 \times 2$  diagonal matrices for the incident and reflected waves, respectively, the diagonal elements of which are the corresponding additional phase factors from slice 0 to 1 (incident) and 1 to 0 (reflected) determined by  $\mathbf{k}$ . If this lead is not the lead of injection and there is only an outgoing wave to consider, Eq. (2.15) reduces to simply  $\boldsymbol{\varphi}_0 = \mathbf{P}_{10,r} \boldsymbol{\varphi}_1$ . In the lead of injection, the wavevector is necessarily real. However, it is possible that the energy of injection  $E$  will fall within the band gap/forbidden region in one or more of the outgoing leads such that the corresponding wavevector of propagation  $\mathbf{k}$  will be complex. Physically meaningful solutions require that the probability density *decay* into those leads.

With Eq. (2.15), the Schrödinger's equation for transport simulation is then completely described within the channel on the left boundary as

$$(\mathbf{E}\mathbf{I} - \mathbf{H}_{10} \mathbf{P}_{10,r} - \mathbf{H}_{11}) \boldsymbol{\varphi}_1 - \mathbf{H}_{12} \boldsymbol{\varphi}_2 = \mathbf{H}_{10} (\mathbf{I} - \mathbf{P}_{10,r} \mathbf{P}_{01,i}) \boldsymbol{\varphi}_{0,i} \quad (2.16)$$

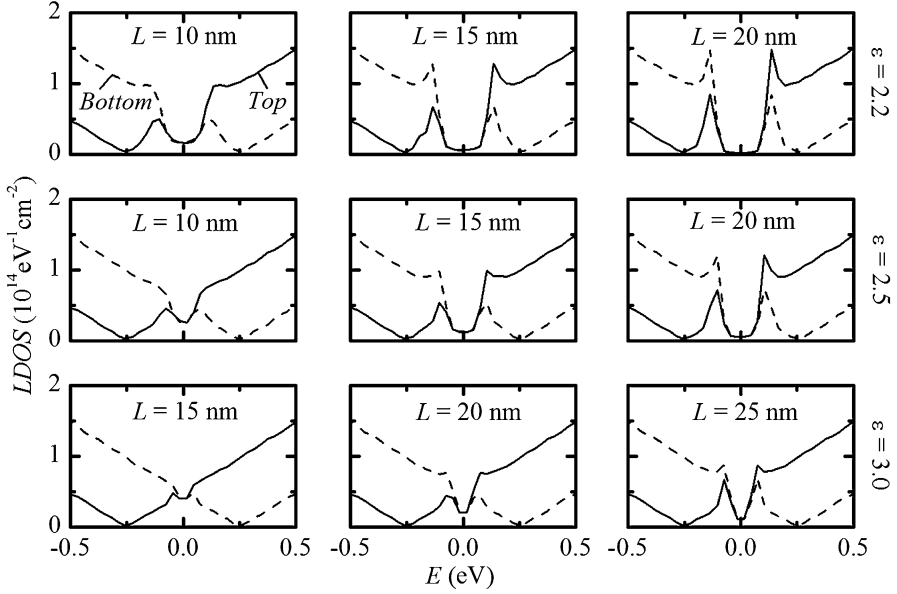
A similar equation is obtained on the right boundary (slices  $N$  and  $N+1$ ). This equation is equivalent to the nonequilibrium Green's function (NEGF) equation  $(\mathbf{E}\mathbf{I} - \mathbf{H} - \boldsymbol{\Sigma})\boldsymbol{\varphi} = \mathbf{S}$  with "self-energy"  $\boldsymbol{\Sigma}$  corresponding to boundary absorption and the "source"  $\mathbf{S}$  corresponding to the injected wavefunction. Specifically,  $\boldsymbol{\Sigma}_{1,1} = \mathbf{H}_{10} \mathbf{P}_{10,r}$ ,  $\boldsymbol{\Sigma}_{N,N} = \mathbf{H}_{N(N+1)} \mathbf{P}_{N(N+1),r}$ , and all other  $\boldsymbol{\Sigma}_{ij} = \mathbf{0}$ , and  $\mathbf{S}_1 = \mathbf{H}_{10} (\mathbf{I} - \mathbf{P}_{10,r} \mathbf{P}_{01,i}) \boldsymbol{\varphi}_{0,i}$ ,  $\mathbf{S}_N = \mathbf{H}_{N(N+1)} (\mathbf{I} - \mathbf{P}_{N(N+1),r} \mathbf{P}_{(N+1)N,i}) \boldsymbol{\varphi}_{N+1,i}$ , and all other  $\mathbf{S}_i = \mathbf{0}$  [31]. However, we do not explicitly solve for the full Green's function matrix  $(\mathbf{E}\mathbf{I} - \mathbf{H} - \boldsymbol{\Sigma})^{-1}$ , which would be required only if we were to consider "injection" from within the channel, such as when considering inelastic (and, thus, phase-breaking) phonon scattering within the simulation region. However, the scattering in a nanoscale graphene device can be expected to be very limited even at room temperature [39].

The occupation probabilities  $f_\beta$  for the states  $\beta$ , as required to calculate current flow and to obtain the exchange interaction, follow the Fermi distribution in the lead of injection  $\alpha$ ,  $f = (1 + e^{E-E_{F,\alpha}/k_B T})^{-1} = (1 + e^{E+qV_\alpha/k_B T})^{-1}$ , where  $V_\alpha$  is the applied voltage on lead  $\alpha$ . Aiming for room-temperature device operation, transport simulations are performed at a temperature of  $T = 300$  K unless otherwise noted. However, beyond the existence and strength of the condensate, temperature has little effect on simulation results.

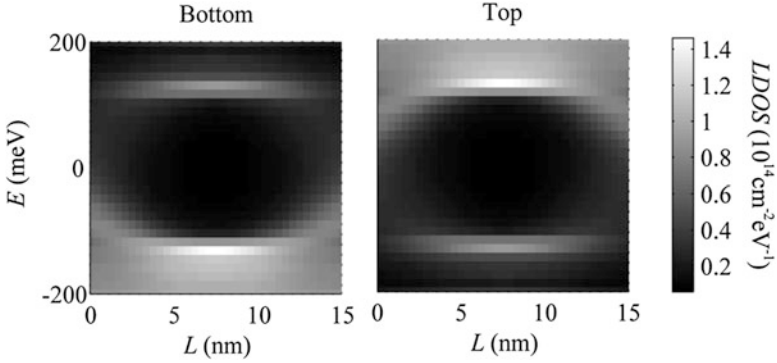
As for the bulk calculation, the carrier concentrations in top and bottom layers are conceptually created by control gates, omitted in Fig. 2.8, producing opposite and fixed electrostatic voltages on the top and bottom layers. For proposed pseudospintronic device applications, the variations in carrier concentration even with external voltages applied to the leads are negligible compared to the concentration itself and, hence, the associated self-consistent changes in electrostatic potential can be neglected. Only the Fock interaction,  $V_{\text{Fock}}(\mathbf{R}_T, \mathbf{R}_B)$ , is obtained self-consistently for now allowing the isolation of essential physics associated with the condensate formation and transport absent such perturbations.

### 2.3.2.2 Condensate Formation in the Channel

To model the condensate formation in nanoscale regions as required for device applications, the *local* density of states (LDOS), which is defined as the available states per unit (here) area per unit energy *as a function of the position*, has been calculated for different channel lengths with fixed carrier concentrations and interlayer separation. The interlayer bare coupling  $V_b$  is also very weak, on the scale of 1 meV, to guarantee the band gap/anti-crossing is primarily the result of condensation. The LDOS in the center of the channel is shown in Fig. 2.9, where the LDOS about the Fermi level approaches zero with increasing channel length. In other words, an anti-crossing band gap, although incomplete due to exponentially decaying tails of the states originating outside the condensate region, forms in the channel due to the exchange-enhanced interlayer coupling in the presence of the condensate. The size of the band gap saturates to approximately that of the bulk condensate band gap [26], implying that the condensate is nearly fully formed in the channel center. The channel length needed for the formation of the condensate also depends on the strength of the bulk condensate, as adjusted here via the effective dielectric constant; the stronger is the bulk condensate, the shorter is the channel needed to fully form the condensate. The spatial extent of the condensate within the channel also is important for device applications. In the example of Fig. 2.10, the LDOS exhibits  $\sim 5$  nm transition regions on either side of a  $\sim 5$  nm bulk-like condensate region (consistent with the coherence length of the exchange interaction in Fig. 2.5a). However, even an incompletely formed condensate will give rise to the novel transport phenomena, as will be shown.



**Fig. 2.9** The calculated local density of states (LDOS) in the top (*solid line*) and bottom (*dashed line*) layers in the center of the channel as a function of the effective dielectric constant  $\epsilon$ . A band gap begins to form and saturates to its bulk value with increasing channel length. The other parameters are  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ ,  $V_b = 0.5 \text{ meV}$ , and  $d = 1 \text{ nm}$



**Fig. 2.10** Local density of states (LDOS) versus energy and position in the channel for  $\epsilon = 2.2$ ,  $d = 1 \text{ nm}$ ,  $L = 15 \text{ nm}$ ,  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ , and interlayer bare coupling energy  $V_b = 0.5 \text{ meV}$ . LDOS for the bottom layer is plotted on the *left* and that for the top layer on the *right*. The region of the condensate within the channel in terms of both energy and position is exhibited by the *dark center* of each figure



### 2.3.2.3 Enhanced Interlayer Tunneling

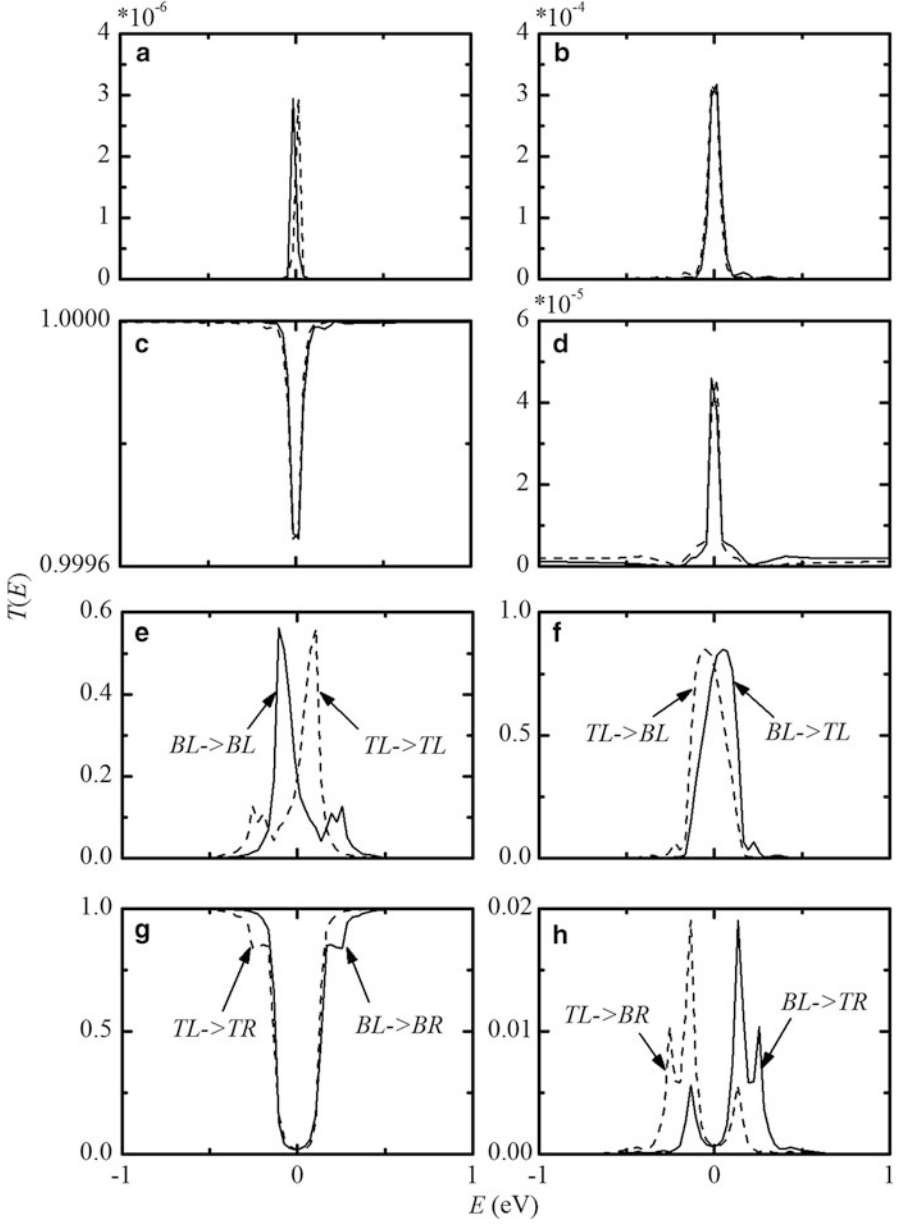
As alluded to previously and consistent with Eq. (2.11), condensate formation with the associated enhanced interlayer density matrix/pseudospin can greatly enhance interlayer current for fixed (but nonzero) bare interlayer coupling.

First, we consider the interlayer transmission probabilities in the equilibrium limit. Calculated  $k_y$ -averaged transmission coefficients among the leads for injections from two leads on the same side of the channel, BL and TL, are plotted in Fig. 2.11. The simulated individual transmission and reflection coefficients for injection from either lead add up to unity, as required for current conservation. The individual transmission probabilities from the top to bottom layer also match the inverse transmission probabilities from the bottom to top layer, as required by “detailed balance,” such that in equilibrium there is no current flow. The difference between the *average* transmission coefficients from TL to BL and from BL to TL is an artifact of the averaging process, where different numbers of states—the denominator—are available in the top and bottom leads as a function of energy except at the Fermi level. Absent the condensate, essentially 100 % of the injected carriers are transmitted to the other side of the same layer. With the condensate, about 75 % of the injected carriers are transmitted to the same side of the *other* layer, with another ~20 % reflected. In other words, the two layers would be nearly shorted together on the same end of the channel, reaching 75 % of the Landauer–Büttiker limit for ballistic current flow.

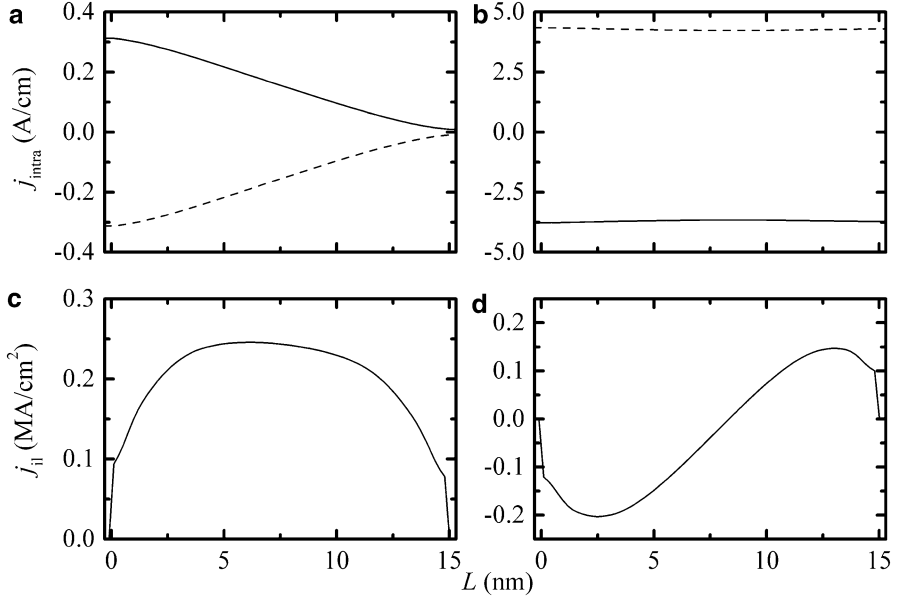
### 2.3.2.4 Detailed Current Flow Under Various Biasing Schemes

As already predicted theoretically [40] and demonstrated experimentally in cryogenic III–V systems, condensate formation can produce novel transport effects including but not limited to enhanced interlayer transport and near-perfect “Coulomb drag” in III–V double quantum wells [18–20]. In the case of Coulomb drag, the current in a biased layer produces a current of near-equal magnitude but opposite sign in the other layer. Again, the necessary interlayer electron–hole interactions—the interaction between electrons in one layer and their absences in the other layer—for these considered coherent transport effects are provided in simulation by the *many-body Fock exchange interaction within this otherwise single-particle formalism*.

The structure in Fig. 2.8 also allows for the consideration of each of these effects. To model Coulomb drag, we set  $V_{BL} = -V_{BR} = V_{al}/2$  and  $V_{TL} = V_{TR} = 0$ . To model interlayer transport, as required for the BiSFET, we set  $V_{TL} = -V_{BL} = V_{il}/2$  and  $V_{TR} = V_{BR} = 0$ . To model the BiSJT, we set  $V_{TL} = -V_{BL} = V_{ctrl}/2$  and  $V_{TR} = -V_{BR} = V_{drv}/2$ . The intra- and interlayer current flows for the BiSFET-like biasing and Coulomb drag biasing (perhaps more accurately called “current counterflow” biasing [40]) are illustrated in Fig. 2.12. These intralayer and interlayer



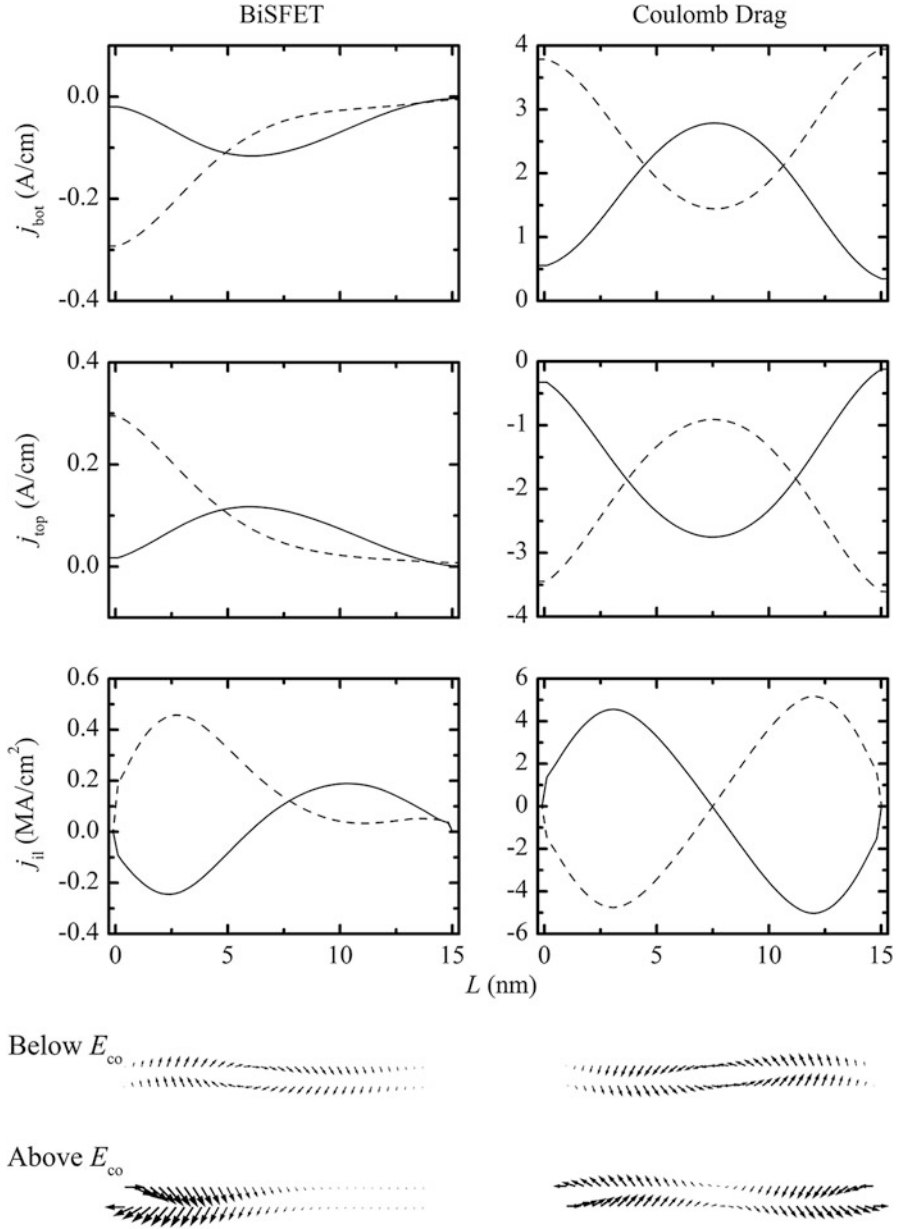
**Fig. 2.11** Averaged transmission coefficients  $T(E)$  over the propagating modes injected from leads BL and TL of the structure of Fig. 2.8. Transmission coefficients without the condensate are plotted in (a)–(d) and those with the condensate in (e)–(h). Device parameters are dielectric constant  $\epsilon=2.2$ , interlayer separation  $d=1$  nm, channel length  $L=15$  nm, interlayer bare coupling  $V_b=1$  meV, and carrier concentrations  $n=p \approx 6 \times 10^{12} \text{ cm}^{-2}$ . The *solid* and *dashed* curves in (a)–(d) correspond to the same transmission components as those in (e)–(h), respectively (but are difficult to label). Interlayer transmission on the *left side*, (b) and (f), is enhanced by more than a factor of 1,000 by condensate formation



**Fig. 2.12** Intralayer ( $j_{\text{intra}}$ ) and interlayer ( $j_{\text{il}}$ ) current distributions. (a) and (c) illustrate the BiSFET biasing condition with  $V_{\text{il}} = 2$  mV,  $V_{\text{b}} = 0.75$  meV; (b) and (d) exhibit the Coulomb drag biasing condition with  $V_{\text{al}} = 50$  mV,  $V_{\text{b}} = 0.5$  meV. Other parameters are  $\epsilon = 2.2$ ,  $d = 1$  nm,  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ , and  $L = 15$  nm. Intralayer current flowing to the right and interlayer current flowing to the bottom layer are considered positive. For the intralayer currents, those in the top layers are shown as *solid lines* and those in the bottom layer as *dashed lines*. A strong delocalization of both the intra- and interlayer current distribution is exhibited

currents as measured at the channel boundaries are entirely consistent with the above-noted prediction of novel behavior.

However, *how* the current flows is far from trivial. The injected current flows in the vicinity of and between the Fermi levels of the leads [38] and consequently into the anti-crossing band gap of the condensate. Under BiSFET biasing conditions, current flow *through* the condensate region is essentially eliminated. However, it penetrates further into the condensate region than one would expect based on very limited LDOS near the Fermi level in the channel. Moreover, interlayer current flow is actually maximized very near the center of the condensate region, not near the left edge. Under Coulomb drag biasing, the intralayer current is largely uniform across the channel despite the condensate-induced band gap. To help understand what is going on, a cutoff energy  $E_{\text{co}}$  is selected well below the Fermi levels and, thus, well below the region of current injection and also above the lower anti-crossing band-gap edge in the condensate region. As shown in Fig. 2.13, the currents above  $E_{\text{co}}$  are indeed largely restricted to the periphery of the condensate region. However, substantial currents flow below  $E_{\text{co}}$  and in a looping pattern within and between the layers. This closed current loop is just the “exciton flow”

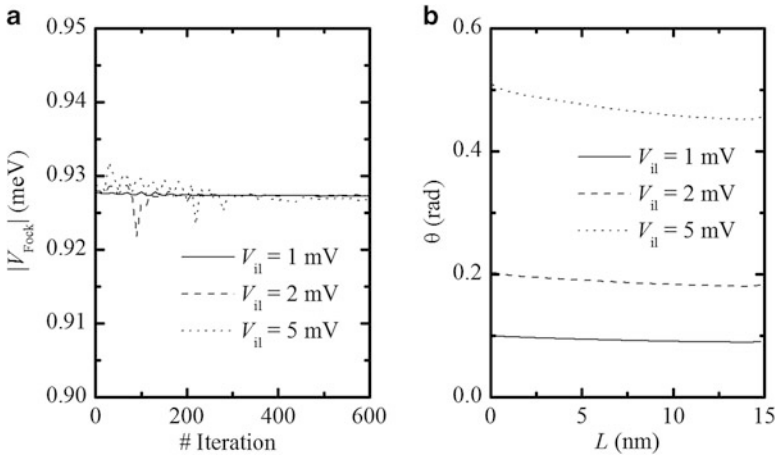


**Fig. 2.13** Bottom-layer ( $j_{\text{bot}}$ ), top-layer ( $j_{\text{top}}$ ), and interlayer ( $j_{\text{il}}$ ) distribution for BiSFET biasing condition in the *left column* and Coulomb drag in the *right column*, following the same device parameters, biasing conditions, and rules for signs of currents as those in Fig. 2.12. In the *top* six figures, current below the cutoff energy  $E_{\text{co}} = -65$  meV is plotted in *solid curves* and that below  $E_{\text{co}}$  in *dashed curves*. In the remaining figures, these distributions are visualized with *arrows* that originate at the points they represent, where the vertical (horizontal) components are proportional to the interlayer (intralayer) current densities. With different units, however, the scaling between the vertical and horizontal components of the *arrows* is arbitrary and chosen for clarity here

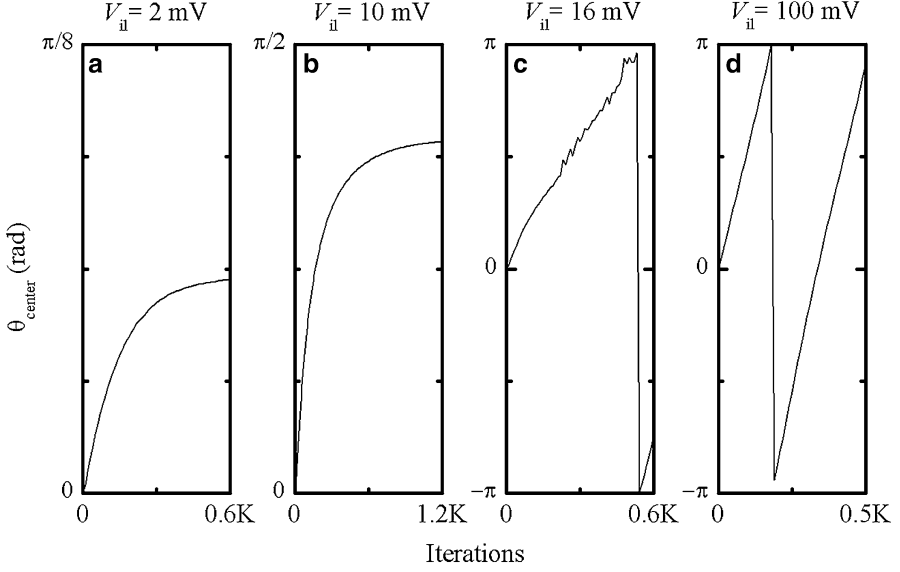
within the channel. Thus, we find that on the same side of the condensate region, an injected (extracted) electron from one layer and hole from the other are annihilated (excited), creating (annihilating) a coherent electron–hole pair/interlayer exciton within the condensate region below the anti-crossing band gap. There is, however, *no* energy loss or gain in any of this process; annihilation (creation) of the incident (extracted) electron–hole pair and, separately, the creation (annihilation) of the coherent exciton within the condensate region are both entirely elastic processes. This entire process, including required exciton conservation within the closed current loop in the condensate region, is intimately coupled to pseudospin device behavior.

### 2.3.2.5 Interlayer Critical Current and Voltage Under Various Biasing Schemes

Consistent with Eq. (2.11), the interlayer current is determined by the interlayer bare coupling energy, the pseudospin magnitude, and the sine of pseudospin phase. For small applied interlayer biases, the pseudospin amplitude, as shown in Fig. 2.14a, is independent of the applied voltage but dependent of the position in the channel. The pseudospin phase, in contrast, is largely a position-independent property but dependent of the applied voltage as shown in Fig. 2.14b. Therefore, Eq. (2.11) can be approximated as simply  $I \cong I_c \sin(\theta)$ , which reaches its maximum amplitude  $I_c$  when  $|\theta| = \pi/2$ . Thus, only a maximum current of  $I_c$  can be carried in



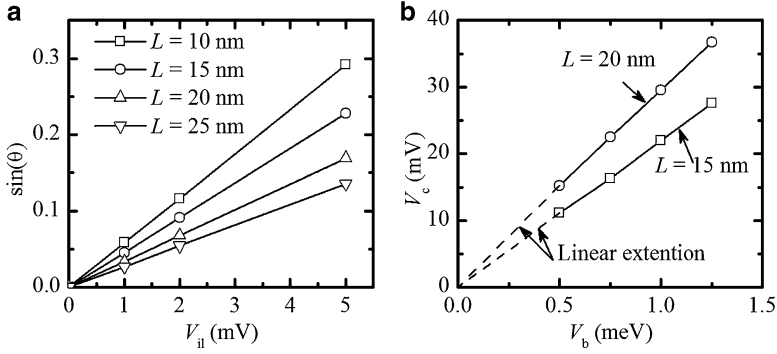
**Fig. 2.14** (a) Interlayer Fock exchange interaction strength at the center of the channel vs. iteration and (b) converged pseudospin phase vs. position in channel, both for the nearest interlayer neighbor A–A coupling. The device parameters are  $\epsilon = 2.2$ ,  $V_b = 0.5$  meV,  $d = 1$  nm,  $L = 15$  nm, and  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ . The transport problem is solved by starting from an initial equilibrium condition with grounded leads and ending with a self-consistent solution obtained iteratively with the BiSFET biasing condition



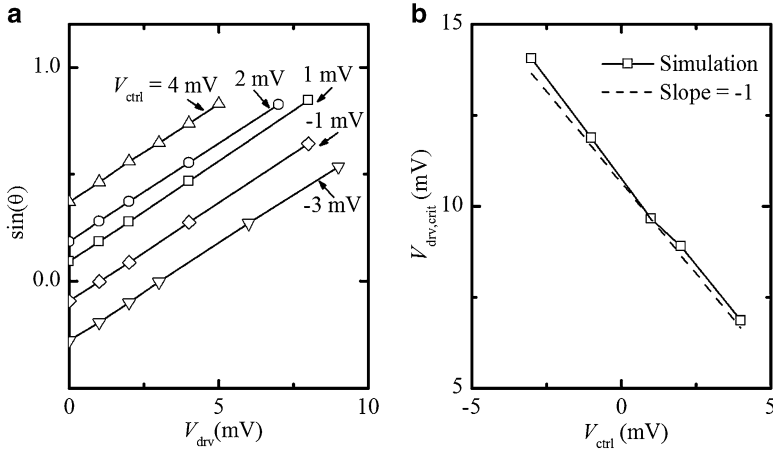
**Fig. 2.15** Evolution of the pseudospin phase with simulation iteration, starting from equilibrium conditions. Device parameters and biasing configuration are the same as in Fig. 2.14. (a)  $V_b = 0.5$  meV and  $V_{il} = 2$  mV, (b)  $V_b = 0.5$  meV and  $V_{il} = 10$  mV, (c)  $V_b = 0.5$  meV and  $V_{il} = 16$  mV, and (d)  $V_b = 0.5$  meV and  $V_{hop} = 100$  mV. The converged solutions (a) and (b) are consistent with stable DC behavior below a critical voltage; (c) and (d) are consistent with the collapse of steady-state behavior and suggestive of the expected cyclic oscillatory behavior beyond a critical voltage

this way.  $I_c$  is called the “critical current” and the associated interlayer voltage  $V_c$  is herein called the “critical voltage” (although the same term is used for different purposes elsewhere). With further increases in the interlayer voltage, the steady-state current is expected to drop drastically with the onset of current oscillation on the scale of 1 THz ( $10^{12}$  Hz) per 2 mV of interlayer voltage. Such oscillations are qualitatively analogous to those in the Josephson junction [17, 19, 41, 42]. The simulation results in Fig. 2.15 are consistent with this expectation. Below a critical voltage (Fig. 2.15a, b), our iterative calculations readily converge to provide steady-state interlayer currents for  $|\theta|$  below  $\pi/2$ . However, with higher voltages such that  $|\theta|$  reaches and exceeds  $\pi/2$  (Fig. 2.15c, d), no steady-state solution can be obtained and we observe cyclic oscillations (not just fluctuations) in  $\theta$  with numerical iteration step in our calculations, which are suggestive of such time-dependent oscillations. Actual simulation of this beyond-critical voltage behavior, however, would require time-dependent NEGF capabilities not yet available.

The  $\sin(\theta)$ – $V_{il}$  relation, or equivalently the  $I$ – $V$  relation, is a linear function of interlayer voltage  $V_{il}$  both in principle and as shown in the simulation results of Fig. 2.16a (and Fig. 2.17a below). Consequently, the critical voltage can be obtained by linear extrapolation of  $V_{il}$  to  $|\sin(\theta)| = 1$ . (In principle, one could sample values of voltage to either side of, and increasingly close to, the transition

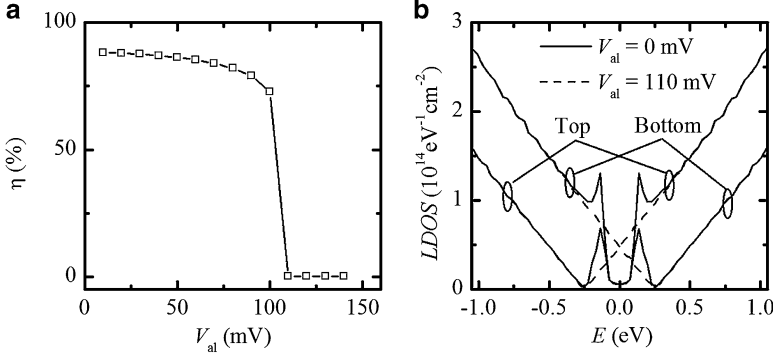


**Fig. 2.16** (a) Linear  $\sin(\theta)$ – $V_{ii}$  relation (producing a linear  $I_{ii}$ – $V_{ii}$  relation) for BiSFET biasing conditions, from which the critical voltage occurring at  $|\sin(\theta)| = 1$  can be obtained. Here  $\varepsilon = 2.2$ ,  $d = 1$  nm,  $V_b = 0.75$  meV, and  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ , with varying channel lengths. Although we have shown by direct simulation that the linear relationship extends at least very close to  $|\sin(\theta)| = 1$  (and takes increasing long to converge in the process), we have shown only a limited region here. (b) Linear dependence of the thus obtained critical voltage  $V_c$  on interlayer bare coupling energy  $V_b$  under otherwise the same conditions as (a)



**Fig. 2.17** Simulation results for a BiSJT-biased structure with  $\varepsilon = 2.2$ ,  $d = 1$  nm,  $L = 15$  nm,  $V_b = 0.5$  meV, and  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ . (a) The sine of pseudospin phase as a function of “drive-end” voltage drop from lead TR to BR, for different “control-end” voltage drops from lead TL to TR. (The slopes of the straight lines are proportional to the interlayer conductance.) (b) The effective drive-end critical voltage drop as a function of the control-end voltage drop.  $V_{ctrl} + V_{drv,crit}$  is almost perfectly conserved, with perfect conservation represented by the reference dashed line of slope =  $-1$

between convergence and oscillation in iterative numerical calculations. However, the time to converge or oscillate increases as the critical voltage is approached, as seen in Fig. 2.15, which makes this latter approach impractical outside of a few test cases in which we have got quite close.) Extracted critical voltages are presented in



**Fig. 2.18** (a) The Coulomb drag efficiency, defined as  $\eta = (|I_{TL}| + |I_{TR}|) / (|I_{BL}| + |I_{BR}|)$  under Coulomb drag biasing/current counterflow biasing/intralayer biasing applied to the bottom layer, as a function of intralayer voltage  $V_{al}$ . Here,  $\epsilon = 2.2$ ,  $d = 1$  nm,  $L = 15$  nm,  $V_b = 0.75$  meV, and  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ . Near-perfect Coulomb drag approaching 90 % is provided until the condensate, itself, collapses at relatively large intralayer voltages  $V_{al}$ , as confirmed by (b). (b) LDOS plotted in the center of the channel with the structure under low and high intralayer biasing with the same parameters as for (a), showing collapse of the anti-crossing gap, and thus the condensate, at high biases well beyond  $k_B T/q$

Fig. 2.16b. This latter figure also shows the expected linear dependence of  $V_c$  on interlayer bare coupling energy and an increase in  $V_c$  with channel length. The latter is also to be expected due to increasing size of the condensate region.

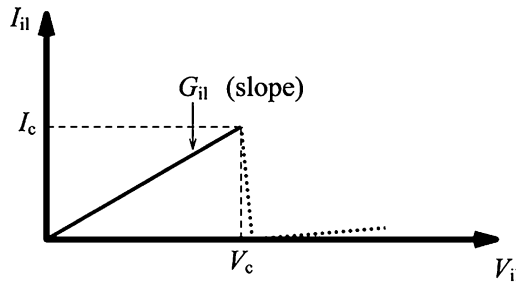
With both ends of the condensate region biased as in the BiSJT biasing condition, we find that there is a conserved total critical current,  $I_{ctrl} + I_{drv} \cong I_c$ . In this symmetric structure, that critical current conservation translates to a conserved critical voltage,  $V_{ctrl} + V_{drv} \cong V_c$ , as shown in Fig. 2.17b. Indeed, it is expected that the critical current would be conserved for any number of leads and independent of input resistance, but consideration of additional leads is beyond our current simulation capabilities.

This latter result for BiSJT biasing suggests that for equal but opposite voltage drops across the lead pairs on opposite sides of the condensate, equal but opposite currents can be maintained that, individually, are much larger than the conserved critical current for the system as a whole. Although the voltages on either end of the condensate are not evenly split between the top- and bottom-layer leads as for the BiSJT biasing condition, the current counterflow biasing condition with equal but opposite voltages applied across one layer closely approaches this limit. Indeed, we find that much larger currents and voltages can be achieved for the same bare coupling [43]. Moreover, the current flow pattern is entirely consistent with the near-perfect Coulomb drag, where the current in the unbiased drag layer closely approaches the current in biased drive layer, as shown in Fig. 2.18a. We also note that the ultimate collapse of Coulomb drag seen in Fig. 2.18a is not due to eventually exceeding the critical current and inducing oscillatory behavior but due to achieving intralayer voltages large enough to collapse the condensate, as shown in Fig. 2.18b, via a process somewhat akin to raising the temperature.



Physically, independent of the biasing condition, the critical current limit results from the requirement for a stable condensate in the channel to enhance the interlayer transmission. To maintain stability, the population of excitons within the channel needs to be conserved. However, as discussed above, electron–hole pairs are injected (extracted) consistent with external voltages and excite (annihilate) excitons in the channel. Thus, for a stable condensate, the rate of injection (extraction) of electron–hole pairs at the left lead pair must be either balanced by extraction (injection) of electron–hole pairs at the right lead pair as in the case of Coulomb drag biasing or balanced by bare-tunneling-assisted recombination (generation) within the condensate region as in the case of BiSFET biasing or balanced by a combination of the two as for BiSJT biasing. When that balance cannot be maintained, the critical current is exceeded. This required balancing also provides a physical explanation for the dependence of the (conserved) critical current on the interlayer bare coupling strength.

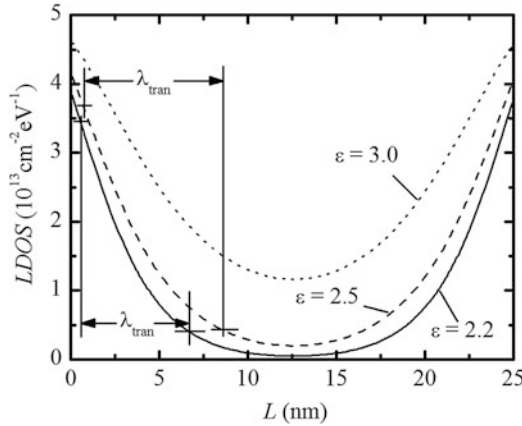
As seen in Figs. 2.16b and 2.17b, the critical voltage can be less than the thermal voltage  $k_B T/q$ , perhaps well below  $k_B T/q$ , as determined by the interlayer coupling  $V_b$ , which is controllably via interlayer dielectric engineering in principle. Subthermal voltages offer the possibility of switching at much lower voltages than CMOS, the latter limited by single-particle thermionic emission. The transition from a high conductance state below  $V_c$  to a low conductance state beyond  $V_c$  is characterized by a region of negative differential resistance (NDR), and the ON/OFF ratio with regard to DC current may be quite high. However, a factor of 10 would probably due for logic applications considered later. The sharpness of the NDR transition is not known at this point but has been observed to be quite sharp in analogous III–V systems and in Josephson junctions [20]. However, it is the transition from ON to OFF, not the region of NDR per se that is important. Such a novel  $I$ – $V$  behavior is sketched in Fig. 2.19, which is the basis for device proposals such as the Bilayer pseudoSpin Field-Effect Transistor (BiSFET) and Bilayer pseudoSpin Junction Transistor (BiSJT) [34, 44]. In some ways, it is analogous to a resonant tunneling characteristic. However, the maximum interlayer conductance is always achieved at zero interlayer voltage, and the width of the “resonance” as limited by the critical voltage can be made very small.



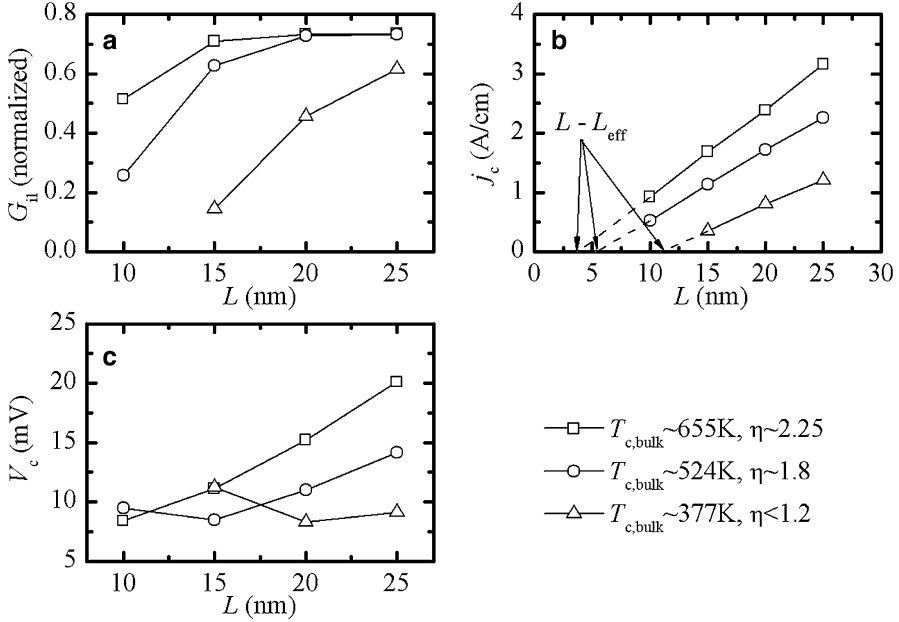
**Fig. 2.19** Schematic representation of novel  $I$ – $V$  behavior of pseudospintronic devices. The enhanced interlayer conductance  $G_{il}$  (ON-state  $G_{il}$ ) approaches  $3/4$  the Landauer–Büttiker ballistic limit. The critical voltage  $V_c$  can be lower than  $k_B T/q$  in principle. The approximate behavior beyond  $V_c$  is shown via the *dotted line*

### 2.3.2.6 Dependencies on Device Parameters

The above work on essential pseudospintronic device physics is sufficient to define device concepts. However, if such devices are ever to be practical, we must consider, beyond the ability to create a condensate, the ability to create it and control it in deeply scaled devices. Here we consider the dependencies of nanoscale condensates on device parameters. Modeling dependencies of critical current  $I_c$  and critical voltage  $V_c$  and interlayer *ON-state* conductance  $G_{il}$  on various parameters including channel length  $L$ , effective dielectric constant  $\epsilon$ , interlayer separation  $d$ , interlayer bare coupling energy  $V_b$ , and carrier concentrations  $n$  and  $p$  are all within the capability of our simulator. For weak bare coupling that is of most interest here, the bare coupling has little effect on the condensate strength [37], but it remains crucial to determining critical currents, of course. Although only semi-quantitative, such studies can identify critical design considerations. The effects of these various parameters can perhaps most easily be interpreted in terms of *bulk* condensate strength as indicated by its critical temperature,  $4k_B T_{c,bulk} \approx E_g(T=0)$ , and the ratio  $\eta = \lambda_F/\lambda_{tran}$  that characterizes the abruptness of the condensate region onset  $\lambda_{tran}$  relative to the Fermi wavelength  $\lambda_F = 2\pi/k_F$  of injected carriers.  $\lambda_{tran}$  is quantitatively defined as the length over which the LDOS at the equilibrium Fermi level decays from 90 to 10 % of its channel-edge value, as pictured in Fig. 2.20, and  $k_F$  is the magnitude of the Fermi wavevector relative to the valley center. The stronger the bulk condensate (the larger  $T_{c,bulk}$ ), the smaller the expected region required to



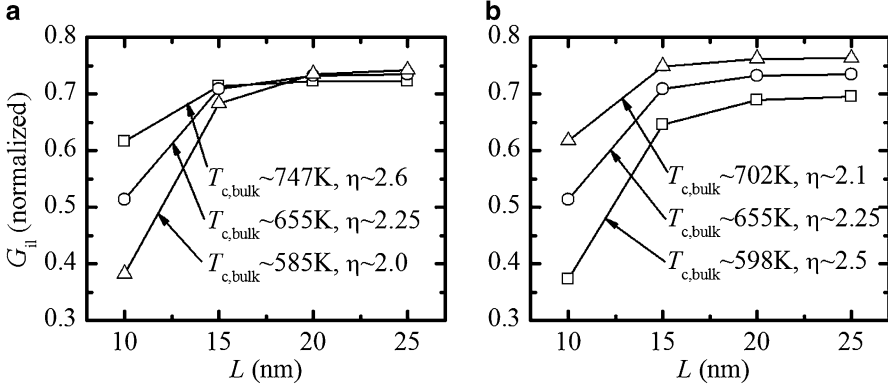
**Fig. 2.20** Estimation of the transition region length  $\lambda_{tran}$ , quantitatively defined as the length over which the LDOS at the equilibrium Fermi level decays from 90 to 10 % of its channel-edge value, in a 25 nm channel via the LDOS at  $E = 0$  eV (equilibrium Fermi level). The greater the effective dielectric constant  $\epsilon$ , the greater the bulk condensate strength, the more abruptly the condensate forms in the channel. For  $\epsilon = 3.0$ , the transition region as defined cannot be determined due to the incomplete band-gap formation. Other parameters for these simulations are  $d = 1$  nm and  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ . The weak interlayer hopping energy  $V_b$  has little effect on the condensate formation, as aforementioned



**Fig. 2.21** (a) Interlayer conductance  $G_{II}$  normalized to the Landauer–Büttiker ballistic limit, (b) interlayer critical current density  $j_c$ , and (c) critical voltage  $V_c$  with varying dielectric constant  $\epsilon$ . The dielectric constants sets are 2.2, 2.5, and 3.0. The other parameters are  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ ,  $V_b = 0.5 \text{ meV}$ , and  $d = 1 \text{ nm}$ . The corresponding  $T_c$  are 655 K, 524 K, and 377 K, respectively. For  $T_c = 377 \text{ K}$  ( $\epsilon = 3.0$ ), the band gap still does not fully form in the 25 nm channel, such that transition length  $\lambda_{tran}$  as defined cannot be determined

form the condensate. However, the more abrupt the onset of the condensate region (the larger  $\eta$ ), the more reflection is expected and the smaller the interlayer conduction.

Typical trends for  $G_{II}$ ,  $I_c$ , and  $V_c$  are shown in Fig. 2.21 [45, 46], where the devices are simulated with varying interlayer dielectric constants but fixed interlayer separation, interlayer bare coupling energy, and carrier concentrations. The interlayer conductance  $G_{II}$  in Fig. 2.21a increases and saturates to a constant value with increasing channel length, except for the smallest critical temperature (largest dielectric constant) for which the condensate is not fully formed; a longer channel length would be required to achieve saturation. The linear dependence of  $I_c$  on channel length  $L$  for a given  $T_{c,bulk}$  in Fig. 2.21b implies that the average critical current per unit channel length is constant in terms of an *effective channel length*  $L_{eff}$  that is somewhat shorter than the actual channel length, even for the case in which the condensate is not fully formed. Note that the difference between actual channel lengths and effective channel lengths for the fully formed condensates is comparable to twice (two channel ends) the exchange potential coherence radius shown in Fig. 2.5a [26, 32]. Both  $L_{eff}$  and critical current per unit effective channel



**Fig. 2.22** Interlayer conductance  $G_{II}$  normalized to Landauer–Büttiker ballistic limit with (a) varying interlayer spacing  $d$  and (b) varying carrier concentrations  $n=p$ . In (a), the interlayer spacings  $d$  are 0.8, 1.0, and 1.2 nm for  $T_{c,bulk} \sim 747$  K, 655 K, and 585 K, respectively. Other parameters are  $\epsilon = 2.2$ ,  $V_b = 0.5$  meV, and  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$ . In (b), the carrier concentrations  $n=p$  are  $\approx 3.8 \times 10^{12} \text{ cm}^{-2}$ ,  $6.0 \times 10^{12} \text{ cm}^{-2}$ , and  $8.6 \times 10^{12} \text{ cm}^{-2}$  for  $T_{c,bulk} \sim 598$  K, 655 K, and 702 K, respectively. Other parameters are  $\epsilon = 2.2$ ,  $V_b = 0.5$  meV, and  $d = 1$  nm

length  $j_c/L_{eff}$  increase with condensate strength, consistent with stronger condensates forming more abruptly and with their larger interlayer density matrices allowing larger interlayer currents per unit area as per Eq. (2.11). The critical voltage,  $V_c = I_c/G_{II}$ , decreases with increasing  $L$  initially due to the faster increase in  $G_{II}$  than in  $I_c$  and increases with increasing  $L$  subsequently because of the saturation of  $G_{II}$  and continuous increase of  $I_c$ .  $G_{II}$  is normalized to the Landauer–Büttiker ballistic limit, about 210 S/cm for carrier concentrations of  $n=p=6.0 \times 10^{12} \text{ cm}^{-2}$ , and appears largely independent of interlayer dielectric constant  $\epsilon$ .

We also note that the interlayer conductance appears to saturate to about 75 % of the Landauer–Büttiker ballistic limit for conductance within the leads in Fig. 2.21. This value, however, is not fundamental. It is a result of reflection from the channel where the more abrupt the onset of the condensate as characterized by  $\eta$ , the more reflection. In Fig. 2.22a, where condensate strength is varied via interlayer spacing, the normalized  $G_{II}$  saturates more quickly with channel length for a stronger condensate with larger  $T_{c,bulk}$ ; however, a weaker condensate provides a slightly larger value of  $G_{II}$  for larger channel lengths due to a smaller  $\eta$  (although ultimately a smaller  $I_c$ ). In Fig. 2.22b, a stronger condensate is achieved by increasing the carrier concentration, which provides a shorter Fermi wavelength and, ultimately, a decreased value of  $\eta$ . As a result, the normalized value of  $G_{II}$  increases with increasing condensate strength (and the Landauer–Büttiker conductance to which it is normalized also increases). However, the differences in saturated normalized conductance are not large, particularly considering that  $G_{II}$  are likely larger than those of practically achievable contact conductances that likely will become the primary current limiters. Indeed, even the interlayer conductance of partially formed condensates, as seen in Figs. 2.21a and 2.22, which remain a significant

fraction of the Landauer–Büttiker ballistic limit, may be comparable to or greater than lead conductances and remain orders of magnitude larger than the interlayer conductivity absent the condensate.

Collectively these results suggest that the novel interlayer  $I$ – $V$  characteristics diagramed in Fig. 2.19 for interlayer transport may well be obtainable at ultralow voltages in nanoscale devices.

### 2.3.3 *Bilayer PseudoSpin Field-Effect Transistor (BiSFET)*

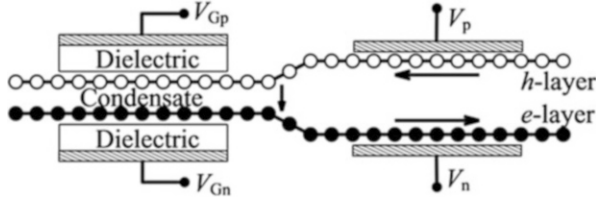
#### 2.3.3.1 BiS Junction/Diode

To realize device application of the exciton condensates, a well-localized region with exciton condensation is required. Theoretically, it can be defined by varying anything that affects the condensate strength, such as charge concentration (e.g., defined by the region of *fixed* voltage gates), dielectric environment, and interlayer spatial separation. Assuming that such a condensate can be formed in a bilayer graphene system with a pair of contacts established to the individual graphene layers, a structure much like Fig. 2.8 with floating right leads TR and BR, one already should have what can be called a “Bilayer pseudoSpin (BiS) junction” or “Bilayer pseudoSpin (BiS) diode” (BiS junction or BiS diode) with the novel  $I$ – $V$  behavior as depicted in Fig. 2.19. The BiS junction would be expected to behave much like a Josephson junction, producing a symmetric low-voltage-onset DC NDR characteristic with an AC/oscillatory current beyond. In this case, however, the oscillations would be on the multi-THz scale in this case, which would be expected to be filtered out via resistance–capacitance time-constant limits in logic circuits.

#### 2.3.3.2 BiSFET

To create transistors, however, one must have one or more additional terminals to control the  $I$ – $V$  characteristic between the original two. For the logic circuits considered in this work, two or more devices are coupled in series. Once the device with the lowest critical current reaches that current, it will turn off, dropping most of the supply voltages across it, leaving the other devices ON. The resulting resistance divider network determines the inter-device node voltage(s) and, thus, the output voltage. Under these conditions, the control input need only vary the critical current and associated critical voltage of the devices, not turn them ON and OFF directly.

The original BiSFET proposal [34] (which can be labeled as BiSFET 1 in this work), as schematically diagramed in Fig. 2.23, incorporated field-effect gating to control charge balance within the condensate, which would change the condensate strength, consistent with the examples of Fig. 2.6. Varying the condensate

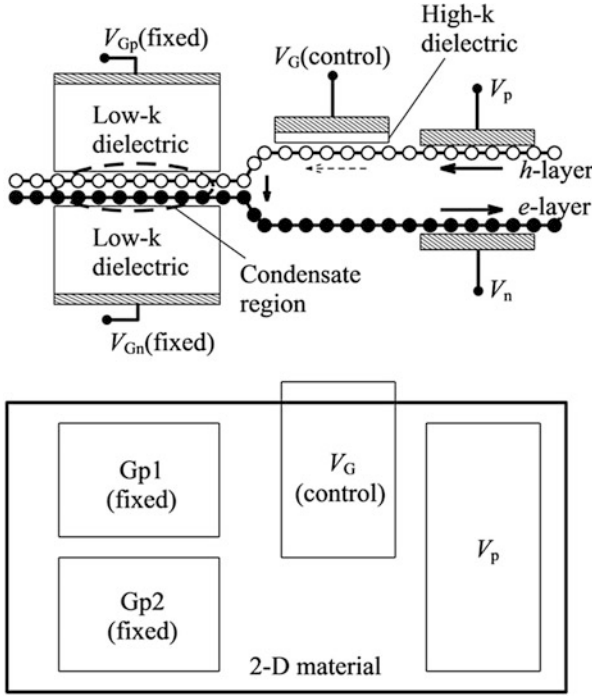


**Fig. 2.23** Initial BiSFET design (BiSFET 1), where the electrons and holes to form the condensates are created by and between control gates voltages,  $V_{Gp}$  and  $V_{Gn}$ , and the interlayer current is driven through the device via separate contacts to each layer,  $V_p$  and  $V_n$ . Nominally the condensate would be established through work-function engineering, and gate voltages would be applied only to adjust the carrier balance and, thus, the condensate strength—consistent with Fig. 2.6—and, thus, the critical current and voltage, consistent with Fig. 2.21b and c and Eq. (2.11). The direction of current flow is marked with arrows. This schematic representation is intended to show basic elements, not physical layout. It is not to scale and vertical dimensions are exaggerated as compared to in-plane dimensions

strengths, in turn, would change the critical current and voltage, consistent with Eq. (2.11) and the examples of Fig. 2.21b, c. The intent was to supply the gate voltages required to create the charged regions between the gates in Fig. 2.23 effectively via work-function engineering. This approach, in turn, would leave the gates available for also adjusting the charge balance via small voltages on the scale of room-temperature  $k_B T/q \approx 25$  mV or lower. As only small variations in the critical current should be sufficient to establish which BiSFET reaches its critical current first, even weakened input signals should be able to establish the logic state for serially connected devices as described above, providing signal restoration and/or fan-out.

However, with the improved understanding of the requirements of the electrostatic environment [25], namely, that the environmental window was much smaller than initial estimates, we realized that the gates would have to be moved further away to prevent them from partially screening the exchange interaction. As a result, work-function engineering would not be sufficient to create the required charge concentrations; actual and significant gate voltages would be required for this purpose, such that the same gates would no longer be available for application of low-voltage input signals.

To address this issue, we turned to a separate “control gate” in the design of *BiSFET 2*, as briefly illustrated in Fig. 2.24 and later discussed in this chapter and in more detail elsewhere [47]. This device is essentially two BiS diodes in parallel with gate control of the input resistance to at least one, although all incorporated into a single structure in this figure. In this way, for any given current through the device, the gate(s) can shunt more current to one or the other parallel BiS diodes/current paths, increasing the current in that path so that its critical current will be reached sooner for a given total current. However, once the critical current for that path is exceeded and that condensate region falls in a high resistance state, the current will be shunted back along the other path causing the critical current to be



**Fig. 2.24** Schematic representation of BiSFET 2, where the gate to control the critical current/voltage, labeled “ $V_G(\text{control})$ ,” is distinct from the gates to create the condensate, labeled “ $V_{Gp}$ ” and “ $V_{Gn}$ ” in the side view, which are now separated further from the 2D material layers. Two separate condensate regions are required, and the control gate would cover only approximately half the channel width. The interlayer currents are still driven by separate contacts to each 2D layer of graphene or other 2D materials. This schematic representation is intended to show basic elements, not the physical layout. It is not to scale and vertical dimensions are exaggerated as compared to in-plane dimensions

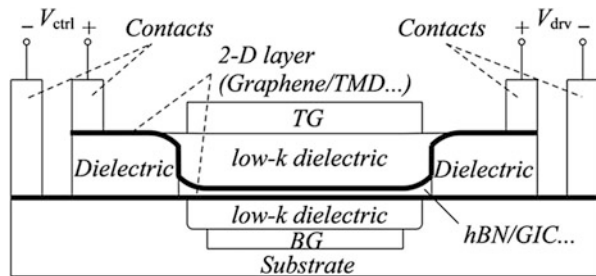
exceeded along that path as well. As a result, the BiSFET as a whole is now turned OFF. (Simulating this process is beyond our NEGF simulations capabilities but has been demonstrated in circuit simulations with compact device models consistent with our NEGF simulation to date, as to be discussed.) Again, the gate control need only establish which series-connected BiSFET has the lowest critical voltage. Thus, the gate control need only overcome unintended variations in critical current within and among devices and noise, which appears possible even in gapless graphene with  $\pm 25$  mV gates and perhaps easier in gapped TMDs. Note, however, that it is not sufficient simply to crowd the current into a small region of a larger condensate, as the critical current is expected to be a globally conserved property of each condensate region regardless of where it enters, as already discussed and consistent with the simulation results of Fig. 2.17. To minimize the screening in BiSFET 2, low-dielectric-constant dielectrics above and below the condensate, including perhaps air gaps, have been proposed, in addition to moving the gates used to create

the layer charge carriers further away from the condensate [47]. With the gates used to create the electrons and holes required for condensation moved further away, the voltages on these “condensate gates” would need to be on the scale of perhaps 10 V. However, the voltages on these gates would be fixed, and the gates would be well insulated from the circuit and act as virtual grounds with only limited capacitive coupling to the circuit.

### 2.3.4 Bilayer PseudoSpin Junction Transistor (BiSJT)

Similar in this respect to a bipolar junction transistor (BJT), the proposed BiSJT, as schematically illustrated in Fig. 2.25, is a current-controlled pseudospintronic device. It is motivated by the search for an alternative method to field-effect gating—which is subject to device-to-device variations in the electrostatic potential under identical gate voltages—to allow even lower voltage operation and perhaps simpler device design. It is inspired by the critical current conservation for multiple leads as illustrated in Fig. 2.17. As already discussed in Sect. 2.3.2 for a symmetric device, the effective critical voltage across the drive-end leads,  $V_{c,drv}$ , is controlled by a smaller voltage applied across control-end leads,  $V_{c,ctrl}$ , such that  $V_{c,drv} = V_c - V_{c,ctrl}$ , where  $V_c$  is the conserved total critical voltage. The control end of the BiSJT, therefore, plays the same role as the “control gate” in BiSFET 2, differentiating the apparent critical current of the device by an amount sufficient to overcome noise and variations in critical current among devices. However, even at room temperature, applied voltage—vs. applied electrostatic potentials—can be extremely well controlled, so that further voltage reduction becomes plausible. The additional complexity in BiSFET 2 introduced by the control gate and the need for two separate condensate regions also is bypassed, although more contacts to the 2D layers are required. In simulations and the above discussion, the “control” and “drive” ends are interchangeable, but an asymmetry between “control” and “drive” ends may be desirable. The control end alone should not be able to turn OFF the BiSJT. The use of a larger input resistance (e.g., smaller contacts) on the control end can ensure this condition. However, fan-out could serve the same purpose in some cases.

**Fig. 2.25** Schematic representation of BiSJT. Device elements are not to scale





## 2.4 Fabrication Techniques

If the theory holds, then realization of the pseudospintronic devices may be defined by fabrication technology limitations, not theoretical ones. There are four basic elements required to create a pseudospintronic device: 2D material layers (graphene, TMD, or others) separated by a dielectric tunnel barrier, a region of condensation created and localized by some means, some means of controlling the critical current, and contacts to the individual 2D material layers. (For graphene devices, one can imagine all-graphene layouts where the contacts would just be graphene leads. However, other issues arise such as doping.) Among those, the greatest technology challenge is providing the required dielectric environment necessary for the creation of the condensate and control of the single-particle/bare interlayer tunneling via the interlayer dielectric, both required in BiSFET and BiSJT proposals. Creating good contacts to graphene, especially in BiSJT design, is still a nontrivial process, and the state of the art lags considerably for TMDs. This latter challenge is not specific for pseudospintronic devices, although independently contacting the 2D layers adds to the challenge.

The dielectric requirements for creating a condensate between dielectrically separated 2D layers are fundamentally different from those required for optimal gate control in a MOSFET. Low- $k$  dielectrics are desired rather than high- $k$ . The dielectric constants of the material above and below the dielectrically separated 2D layers generally will be more important than that of the dielectric in between. It also may be the case that the high-frequency dielectric constant is more important than the low-frequency one, which is usually smaller [25]. Despite the latter positive note, air gaps in the vicinity of the condensate may still be necessary. For this purpose, one may start with sacrificial oxides which are then etched away. While certainly technologically challenging, the use of air gaps between the gate and the source and the drain has already been proposed to reduce parasitic gate-to-source and gate-to-drain capacitances in conventional MOSFETs [48], and top- and bottom-gated suspended graphene bilayers have already been reported [49]. The dielectric requirements may be at least less stringent for TMDs as previously noted or even for the graphene-analogous materials of silicene or germanene (2D layers of silicon and germanium, respectively). However, achieving sufficient carrier concentrations in TMDs will require finding pairs of TMD materials with limited *interlayer* valence to conduction band gaps, and the challenges for producing these latter materials may be greater. In contrast, the dielectric requirements for the control gate region of BiSFET 2 scheme are much the same as for “conventional” MOSFETs made in these materials. Moreover, the amount of gate control required is perhaps significantly smaller than conventional CMOS technology. The challenge will be reproducibility.

The interlayer dielectric, while also preferably low- $k$ , must also be thin and provide both limited and reproducible interlayer single-particle/bare coupling, since the critical current varies with the interlayer bare coupling. This requirement suggests naturally layered systems such as hexagonal boron nitride (hBN), TMDs

or graphene intercalants (GIC), all of which are being explored currently. The interlayer coupling also must be on the appropriate scale.

Rotational alignment of the opposing 2D layers will also be important as all of the proposed materials have their energy valleys located at the periphery of the Brillouin zone. Both simple theory and computationally intensive numerical simulations indicate that the rotational alignment of the graphene layers, and likely other materials for the same reasons, will have little impact on the spontaneous condensate [32]. However, rotational alignment will certainly affect the single-particle tunneling and, thus, the critical current through the pattern of single-particle coupling. Indeed, simply rotating graphene layers may be enough to decouple them sufficiently to produce the condensate, and no interlayer dielectric may be required. However, with the corresponding increase in interlayer capacitance, gating the required charge densities to achieve room-temperature condensation would become more problematic.

Use of graphene intercalants as the interlayer dielectric would be advantageous in this regard, as one starts with crystallographically aligned bilayers into which the intercalant is then chemically inserted. As intercalants also heavily dope the graphene layers [50], their use could lead to the ability to produce the condensate with just one gate to create the condensate, which would be used to strongly invert one layer if while reducing the carrier concentration in the other in the process.

An illustrative rough layout proposal considering such factors is illustrated in Fig. 2.26 for BiSFET 2. While daunting, having technology as the primary roadblock to be chipped away over time, as for CMOS previously, would still represent a considerable improvement over the roadblock of intractable physics of thermionic emission and source-to-drain and channel-to-gate tunneling currently standing in the way of long-term progress in CMOS.

## 2.5 Logic Applications of Pseudospintronic Devices

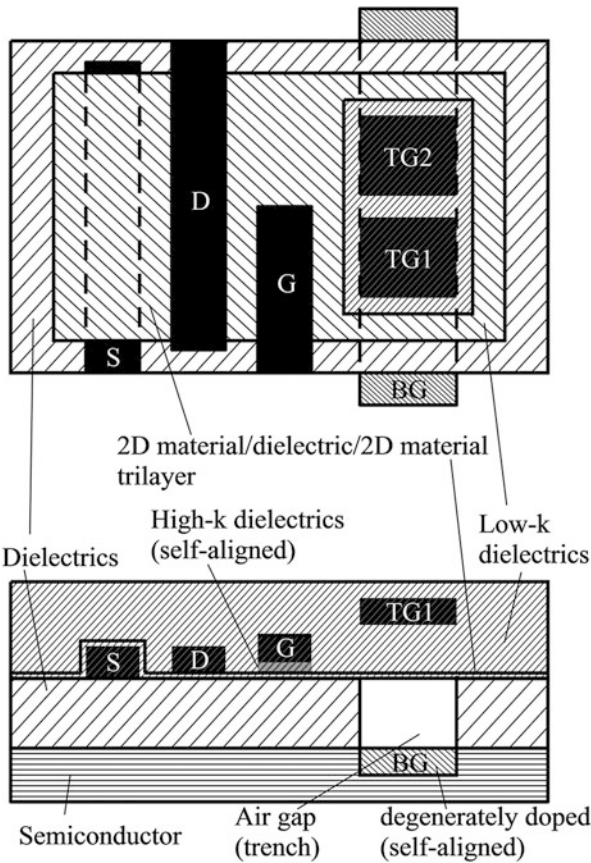
We have found that it should be possible to create a full array of logical gates using BiSFETs or BiSJT's and novel clocked power supplies. In this section, we provide examples for each, beginning with compact device models and ending with circuit-level SPICE simulations.

### 2.5.1 *BiSFET Logic*

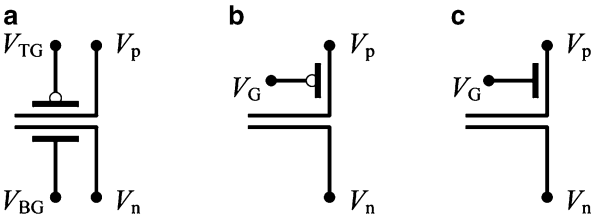
#### 2.5.1.1 BiSFET Inverter

We symbolically represent the BiSFET 1 and BiSFET 2 as shown in Fig. 2.27. “N-type” and “p-type” versions of BiSFET 2 are differentiated by the use of a negative control voltage applied to an n-type device to decrease its critical current

**Fig. 2.26** A rough layout proposal for BiSFET. 2. Vertical dimensions are somewhat exaggerated with respect to in-plane dimensions



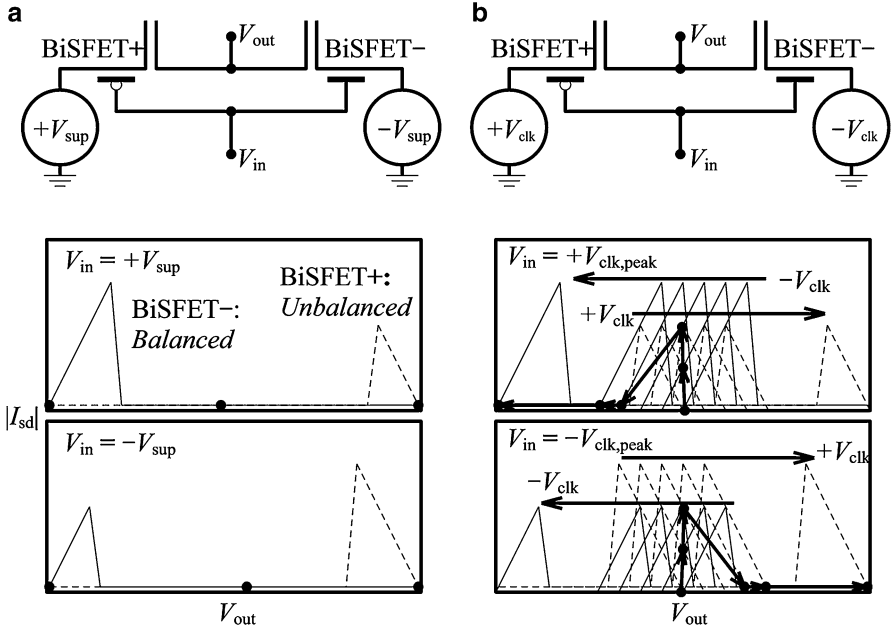
**Fig. 2.27** Symbolic representation of BiSFETs. From left to right: (a) BiSFET1 and (b) “p-type” and (c) “n-type” BiSFET 2



(which will ultimately lead to it being turned OFF) and the use of a positive control voltage applied to a “p-type” device to decrease its critical current, following the convention for CMOS devices. However, there is no need to invert the polarity of carriers in the channel for conduction, nor do “n” and “p” necessarily denote carrier type for conduction.

If BiSFETs are inserted into a conventional CMOS inverter circuit, a memory element results instead of a BiSFET inverter [51]. As schematically illustrated in Fig. 2.28a, there will be three separate operating points given fixed power supply voltages  $\pm V_{\text{sup}}$ . Two of the operating points will have one BiSFET or the other within its low conductance/OFF state beyond the NDR onset voltage  $V_c$ , while the remaining BiSFET will remain in its high conductance/ON state. The output voltage will be either high (“1”) or low (“0”). For the remaining operating point, both BiSFETs will be in their low conductance/OFF state beyond their NDR onset voltages. The output will be in a perhaps not entirely stable intermediate state and with little available current to drive a load. Moreover, altering the input voltage to produce limited changes in the NDR onset voltages will have essentially no effect on the high- or low-output states and no apparent useful effect on the intermediate state. Such static operating points are common in NDR devices. Although such a static voltage supply scheme is not applicable for a BiSFET inverter, at least this indifference to input voltage with a fixed supply voltage is compatible with memory applications as noted [51].

Output control, however, can be achieved using a dynamic voltage supply scheme. If the inputs are set *first* to differentiate the critical currents of the two BiSFETs and the power supplies are then ramped up, the desired output logic state will be obtained, as schematically illustrated in Fig. 2.28b. In this way, initially there is only one possible operating point at  $V_{\text{out}} = 0$  with both BiSFETs in their high conductance/ON state. As the balanced supply voltages are ramped, initially both BiSFETs remain in their high conductance/ON states and the current increases, while  $V_{\text{out}}$  remains approximately zero for matched BiSFETs with the same contact and ON-state interlayer resistances. However, at some point with the increasing supply voltages, the BiSFET with the lower critical current reaches its critical current and falls into its low conductance/OFF state beyond its critical voltage. With the same current funneled through both BiSFETs, the other BiSFET can never reach its critical current and, thus, remains in its high conductance/ON state. As a result,  $V_{\text{out}}$  follows the supply voltage connected to the ON device until  $\pm V_{\text{clk,peak}}$  is reached and the desired inversion operation is achieved, although delayed by the power supply ramp time. Moreover, once the clock voltage is raised and the gate output is set, the input signal can be turned off with no effect. Each gate doubles as a latch since the output = HIGH and output = LOW states are stable with the now static supply as described above. The next operation cycle will begin when  $V_{\text{clk}}$  is ramped up again. Therefore, a BiSFET inverter needs a clocked power supply and is actively switches at the leading clock edge. Moreover, if multiple stages are cascaded to create complex logic gates, a delay in the clock signal between adjacent states is necessary to allow for the time lag between input and output of a single stage. Thus, BiSFET logic requires a sequential clocking scheme [34, 51]. Finally, we note that a BiSFET follower/buffer gate (BUF) can be created simply by flipping the n-type and p-type BiSFETs. Buffers can be used to synchronize inputs and/or outputs of elements with different numbers of stages.

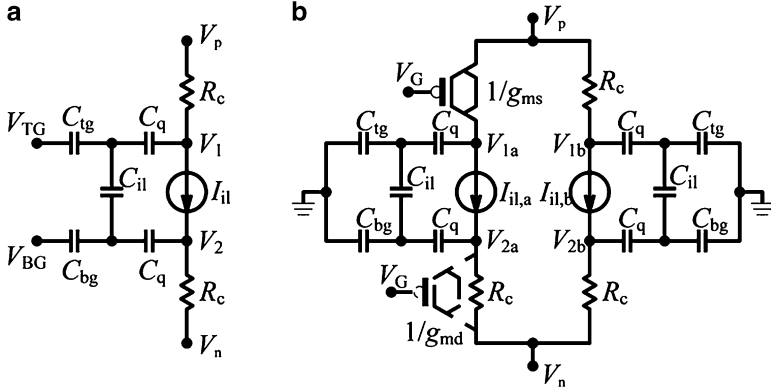


**Fig. 2.28** (a) CMOS-like inverter circuit with BiSFETs and fixed power supplies  $\pm V_{sup}$  and an associated output voltage  $V_{out}$  that is essentially independent of input voltage  $V_{in}$ . The *solid* and *dashed* curves illustrate the interlayer current magnitude vs. common source/output voltage  $V_{out}$ , for “BiSFET-” and “BiSFET+,” respectively. The intersection points of the paired BiSFET  $I$ - $V$  characteristics represent possible circuit operating points/values of  $V_{out}$ , of which there are three for each input voltage. These operating points remain largely fixed independent of  $V_{in}$ ; there is no path between operating points with changes in  $V_{in}$ . This behavior offers opportunities for memory, but not logic. (b) CMOS-like inverter circuit with the BiSFETs and power supply ramped to  $\pm V_{clk, peak}$  after the input voltage is set. The *solid* and *dashed* curves move with respect to  $V_{out}$  with increasing  $\pm V_{clk}$  and therefore so does the intersection/circuit operating point/value of  $V_{out}$  for logic application, as shown by the arrows

### 2.5.1.2 BiSFET Model

For circuit modeling purposes, simple compact device models for both BiSFET 1 and BiSFET 2 schemes were created to capture essential device elements and behavior, as shown in Fig. 2.29. These models were meant to be qualitatively accurate and conservative when in doubt in terms of desired device operation, but not to be precise as would be required for conventional devices. Fractional changes in performance are of secondary importance at this stage of development.

For BiSFET 1, the models include capacitances  $C_{tg}$  and  $C_{bg}$  between each gate and the adjacent graphene layers, capacitances  $C_{il}$  between the two graphene layers, and intralayer quantum capacitances  $C_q$  associated with moving the Fermi level with respect to the band edge for each layer. Contact resistances  $R_c$  are included. Finally, a model of the interlayer current  $I_{il}$  as a function of the interlayer voltage (Fermi level difference) and, for gating in BiSFET 1, charge balance/imbalance is provided.



**Fig. 2.29** BiSFET circuit models. (a) For BiSFET 1 and (b) for BiSFET 2

For BiSFET 2, gating has been modeled as an effective variation of the contact resistance of one (or more) leads,  $1/g_{m,s}$  and perhaps  $1/g_{m,d}$ . In circuit simulations, to date, this gating has been modeled after a graphene FET with thermal and inhomogeneous smearing of the conductance minimum. The capacitive coupling  $C_{tg}$  and  $C_{bg}$  to the gates used to create the condensate in BiSFET 2 are substantially reduced due to their much greater separation from the layers. Moreover, while these gates have fixed large voltages applied, they should act as virtual grounds in the circuit and, thus, have little effect on power consumption. However, charging and discharging the control gate, which is inherently strongly coupled, becomes a new source of power consumption.

The current flow including the transition over the NDR region has been modeled via smooth functions as [34, 51]

$$I_{sd} = G_0 V_{sd} \left[ 1 + \frac{|V_{sd}|/V_c}{\exp(1 - |V_{sd}|/V_c)} \right]^{-1} \quad (2.17)$$

where  $I_{sd}$  is the source-to-drain interlayer current,  $G_0$  is the Landauer–Büttiker ballistic conductance for injection into these highly charged regions,  $V_{sd}$  is the source-to-drain voltage drop across the condensate  $-(E_{F,s} - E_{F,d})/q$  (distinguished from the source-to-drain voltage drops across the contacts), and  $V_c$  is the critical voltage at which NDR onset occurs.

For BiSFET 1, where charge imbalance induced through  $C_{tg}$  and  $C_{tg}$  is employed to tune the critical voltage,  $V_c$  is modeled as [34]

$$V_c = V_{c,n=p} \exp\left(-10 \frac{|p-n|}{p+n}\right) \quad (2.18)$$

For the BiSFET 2 compact model, the resistance of the gated lead as a function of the control gate voltage is assumed to be limited by thermal and inhomogeneous smearing about the Dirac point as [51]

$$R^{-1} = \sqrt{\sigma_g^2 + \sigma_{\min}^2} \quad (2.19)$$

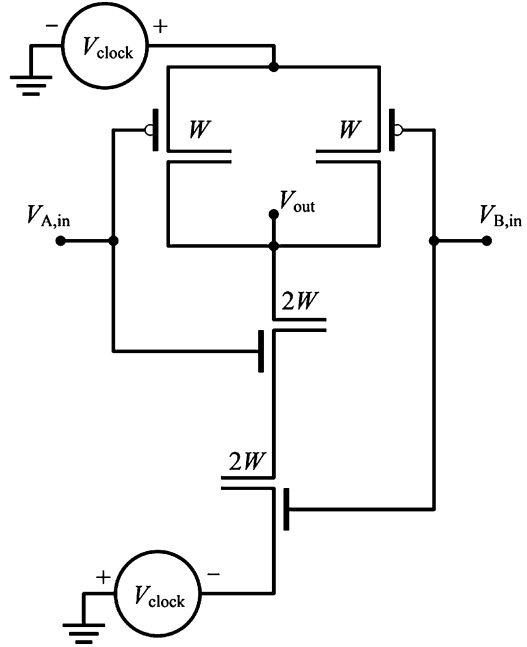
where  $\sigma_g = 8q^2 E_F W / (v_F h^2)$  is the quantum conductance due to the number of modes in a graphene layer of width  $W$  and Fermi level  $E_F$ ,  $h$  is the Planck's constant,  $q$  is the electron charge,  $v_F$  is the fixed Fermi velocity magnitude of electrons in graphene, and  $W$  is the channel width [34].  $\sigma_{\min}$  represents the minimum conductivity due to thermal and inhomogeneous smearing about Dirac point. The Fermi level in graphene under the gate is calculated using an equivalent series capacitance network of gate oxide and graphene quantum capacitance. In practice, only a  $\pm 50$  mV range ( $2V_{\text{clock}}$  in simulations to follow) is considered about the maximum resistance.

### 2.5.1.3 BiSFET Logic and Circuit Simulation

The BiSFET inverter has much the same structure as the CMOS inverter but a different clocking scheme. BiSFET NAND/NOR gates also can be similarly designed following the CMOS NAND/NOR gates. An example BiSFET 2 NAND gate is shown in Fig. 2.30, where the critical current of the “ $2W$ ” device is twice of that of the “ $W$ ” device, nominally by doubling the channel width. The maximum current allowed through the two  $2W$  BiSFETs in series will be reduced by either negative  $V_{\text{in,A}}$  or negative  $V_{\text{in,B}}$  or both, while the total current of the two  $W$  devices in parallel is approximately the same when  $V_{\text{in,A}}$  and  $V_{\text{in,B}}$  are different or increased when both  $V_{\text{in,A}}$  and  $V_{\text{in,B}}$  are negative. Therefore, the  $2W$  BiSFET(s) with the reduced critical current will be turned OFF first and the output will follow the  $+V_{\text{clk}}$  going high. On the other hand, only two positive inputs will effectively increase the maximum current allowed through the two  $2W$  BiSFETs in series, while the total current of the two  $W$  BiSFETs in parallel is reduced. Therefore, the output will follow the  $-V_{\text{clk}}$  going low after the two  $W$  BiSFETs are turned OFF. Thus, the logical operation of a NAND gate is achieved.

The NOR gate can be designed similarly with two  $W$  n-type BiSFETs in parallel and connected to  $-V_{\text{clk}}$  and two  $2W$  p-type BiSFETs in series and connected to  $+V_{\text{clk}}$ . AND and OR gates can be achieved by switching the polarity but not sizes of the n-type and p-type BiSFETs of NOR and NAND gates, respectively. In general, arbitrary logic can be achieved by a NOT gate plus only one out of AND, OR, NAND, and NOR gates. However, unlike CMOS logic, BiSFET logic does not in principle have a preferred gate selection because of the ability to make “n-type” and “p-type” devices with leads to either carrier type. In CMOS logic, NAND gates are preferred because it takes much larger area for a PMOS to carry double the current of the NMOS devices due to the lower mobility of holes.

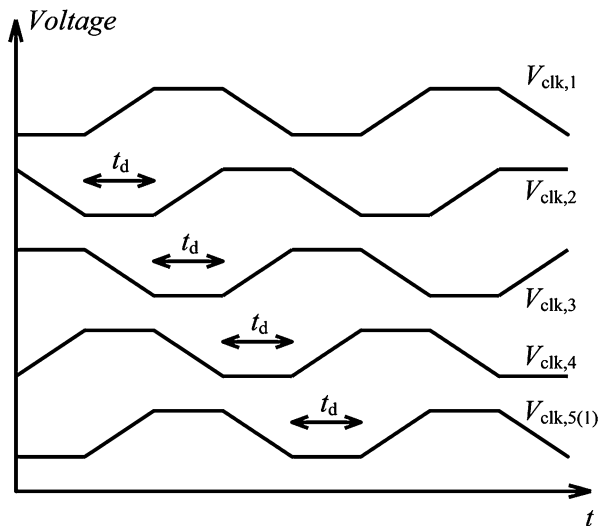
**Fig. 2.30** Schematic of BiSFET 2-based NAND circuit



To evaluate the circuit performance, Verilog-A model-based circuit-level SPICE simulations were performed using the device models described in Fig. 2.29. Except for that of the gated lead in the compact model of BiSFET 2, the contact resistances were taken to be  $320 \, \Omega \, \mu\text{m}$  at zero gate voltage, corresponding to a Fermi level of  $|E_F| = 50 \, \text{mV}$  and a carrier concentration of about  $2 \times 10^{11}/\text{cm}^2$  in the leads. The interlayer conductance is approximately  $200 \, \text{S/cm}$ , or equivalently the interlayer resistance is approximately  $50 \, \Omega \, \mu\text{m}$ . The contact resistance of the gate lead was modeled consistent with Eq. (2.19) for the simulation results shown here. Even with a large  $\sigma_{\min}$  producing an accompanying small ON/OFF ratio of 1.06, successful circuit operation has been achieved in simulation absent device-to-device variation. The interlayer capacitance was taken to be  $\approx 0.35 \, \text{fF}/\mu\text{m}$  corresponding to a  $\text{SiO}_2$ -equivalent oxide thickness (EOT) of  $1 \, \text{nm}$  and a channel length of  $10 \, \text{nm}$ , given that the injected carriers penetrate only peripherally into the condensate, i.e., only the charge in the transitional region varies with current flow. The capacitance to control gate was taken to be  $\approx 0.41 \, \text{fF}/\mu\text{m}$  corresponding to an EOT of  $0.85 \, \text{nm}$ . The capacitance to each of the gates used to create the condensate was taken to be  $0.03 \, \text{fF}/\mu\text{m}$  conservatively, corresponding to an EOT of  $10 \, \text{nm}$  and representing a low- $k$  gate dielectric environment shown in Fig. 2.26 with somewhat physically closer gates, which may not be the case. The interlayer current-voltage relation through the condensate is that of Eq. (2.17). The width  $W$  of the condensate regions, by which all of the above was scaled, was taken to be  $20 \, \text{nm}$  unless otherwise noted. The peak value of the clocked power supply voltage  $V_{\text{clk,peak}}$  was taken to be  $25 \, \text{mV}$  in all cases for illustrative purposes, approximately  $k_B T/q$  at  $T = 300 \, \text{K}$ .



**Fig. 2.31** Four-phase clocking scheme,  $V_{\text{clk},n}(t)$ , with inter-clock time delay  $t_d$  of a quarter of clock cycle



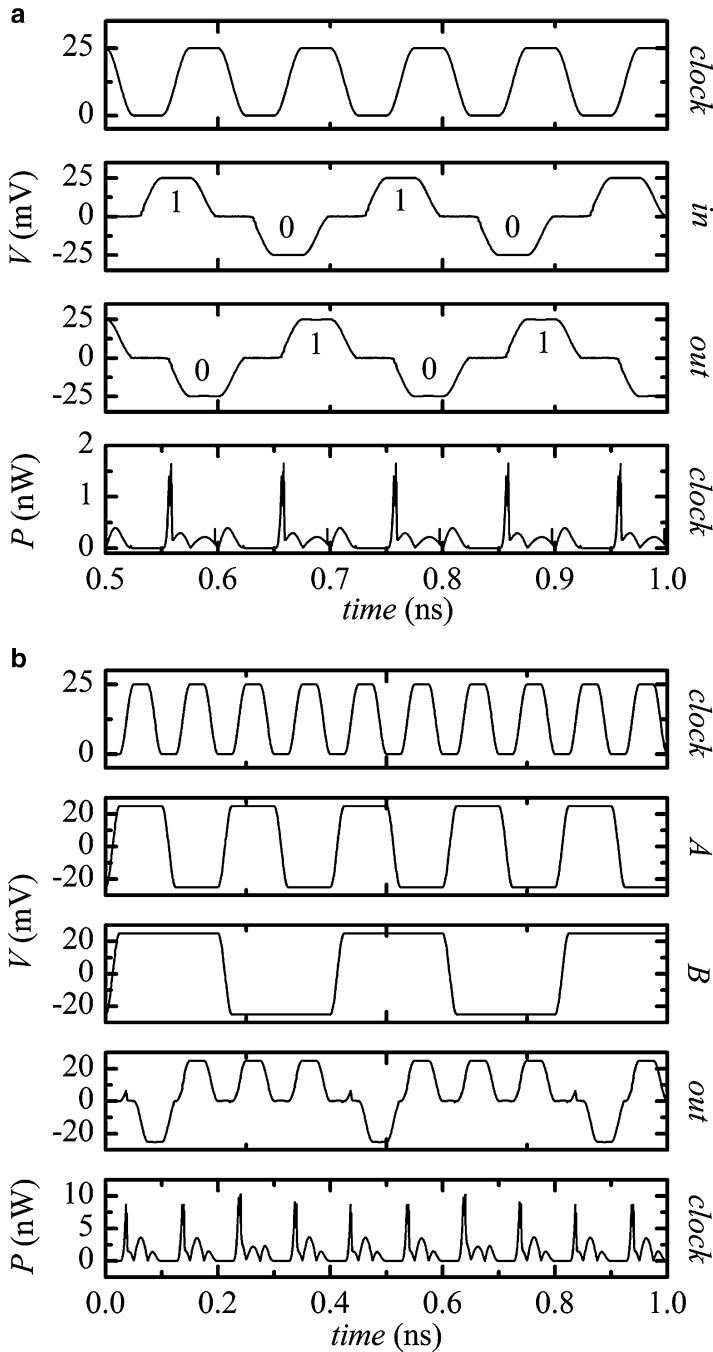
The clocks are taken to have a quarter-cycle delay between stages as illustrated in Fig. 2.31. Simulation results of a BiSFET 2-based inverter and NAND gate are provided in Fig. 2.32, where dual 10 GHz,  $\pm 25$  mV clock signals are applied. The energy per BiSFET per clock cycle is  $\sim 10$  zJ ( $1 \text{ zJ} = 10^{-21} \text{ J}$ ) in the inverter. The total energy per clock cycle is  $\sim 140$  zJ for the NAND gate (which has six times the device area of a single BiSFET). Therefore, the average energy per BiSFET per clock cycle in these simulations is on the order of tens of zeptojoules, orders of magnitude lower than end-of-the-road-map CMOS [52].

A 1-bit ripple-carry adder, with sum and carry, incorporating smaller logic elements of XOR, NAND, and buffer gates, was also simulated based on BiSFET 1 in earlier work. However, BiSFET 1 and BiSFET 2 behave essentially the same in logic circuits. The results are shown in Fig. 2.33, illustrating the ability to create complex, multistage BiSFET logic circuits.

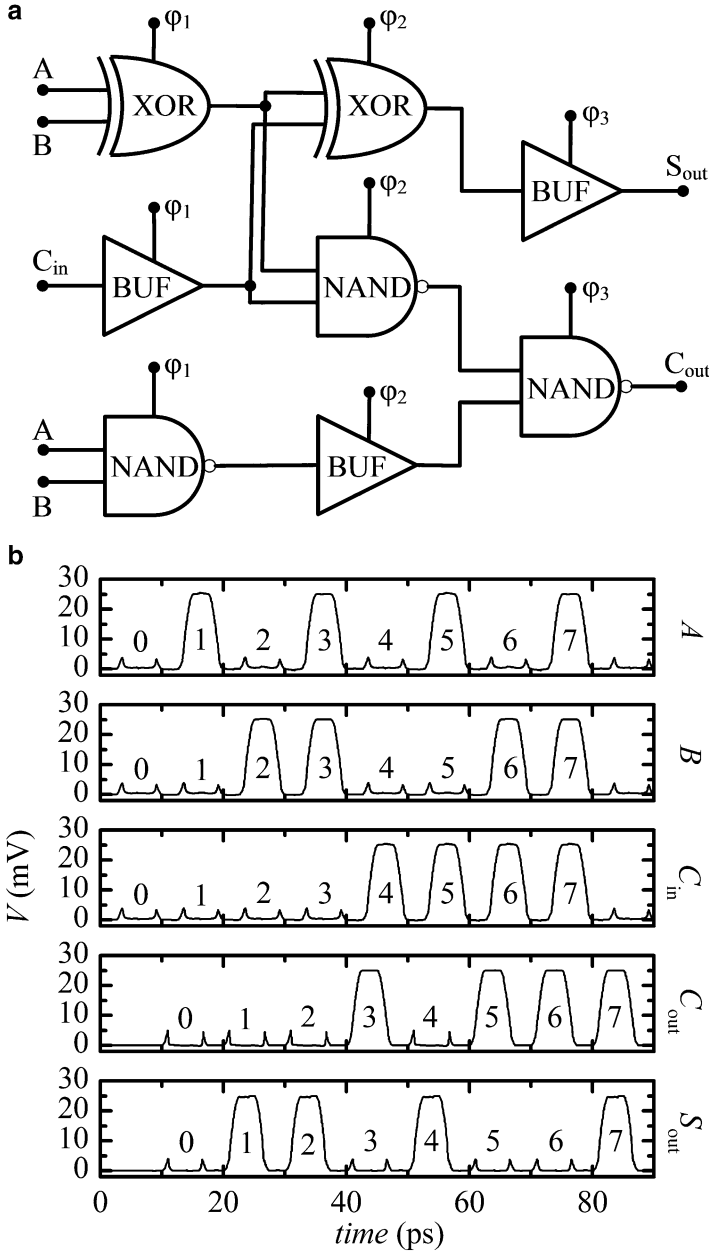
## 2.5.2 BiSJT Logic

### 2.5.2.1 BiSJT Inverter and Model

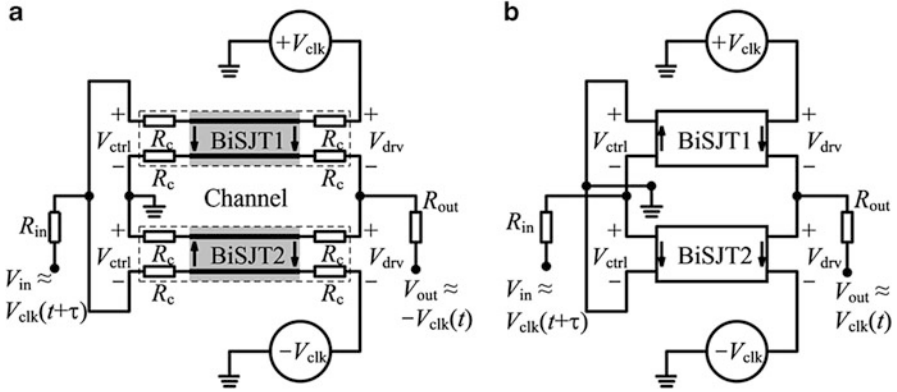
A BiSJT inverter is shown in Fig. 2.34. Like the BiSFET inverter, the clock is ramped up *after* the input signal is set, and adjacent gates are clocked sequentially with a nominally quarter-cycle delay, with each gate serving as its own latch once set until re-clocked. If a positive input voltage is provided, the effective  $I_c$  seen at the drive end of BiSJT 1 is reduced, and that of BiSJT 2 is increased (BiSJT1 and BiSJT2 are just different devices in the same circuit, not to be confused with the different varieties of BiSJTs). When the two opposite-polarity clocks are ramped



**Fig. 2.32** SPICE-level simulated response of (a) the BiSFET inverter shown in Fig. 2.28. (b) The NAND gate shown in Fig. 2.30, both of which employ BiSFET 2 with dual 10 GHz,  $\pm 25$  mV peak clocked power supplies  $\pm V_{\text{clock}}(t)$  driving a four-identical-inverter load (fan-out of four). The input signal was also taken from a preceding inverter in a similar fashion. The energy per operation/clock cycle per BiSFET is about 10 zJ in for the inverter, and the total energy consumption per operation cycle is about 140 zJ in for the NAND gate. The instantaneous power was conservatively calculated as the *magnitude* of the product of the current times voltage summed over both supplies. There actually is some power return during part of the clock cycle in simulations

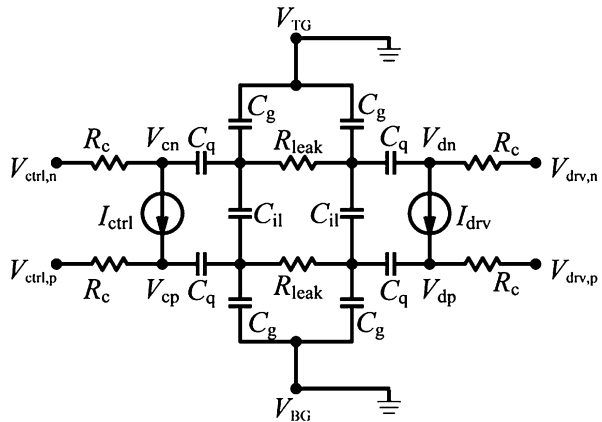


**Fig. 2.33** (a) Circuit schematic of a 1-bit full adder incorporating XOR, NAND, and buffer gates, with clocked power supplies  $\phi_1$ ,  $\phi_2$ , and  $\phi_3$ . The buffer is employed to synchronize signals. (b) Output of the 1-bit adder, including sum  $S_{out}$  and carry  $C_{out}$ . These results were obtained with BiSFET-1-based logic elements and with 100 GHz clocked power supplies separated by a 2.5 ps delay. The higher speeds were allowed because the parasitic contact resistances were not considered in these early proof-of-concept simulations



**Fig. 2.34** (a) A low-output BiSJT inverter. Sequential clocking is employed for input and output signals with a delay  $\tau = T/4$  namely. The arrows show the direction of the interlayer current flow. Contact resistances  $R_c$  are explicitly shown here, while capacitances, which affect operation speed and charging energy but not basic logic design, are implicit (but explicit in compact model of Fig. 2.35 below). The gates to create the condensate are also omitted here (and also included in Fig. 2.35). (b) A BiSJT buffer gate is achieved by switching the input and grounded nodes on the control side of the BiSJT's. Here the contact resistances are also implicit

**Fig. 2.35** The BiSJT compact model used in circuit simulations, which corresponds to the boxed “BiSJT1” or “BiSJT2” in Fig. 2.34b



up, BiSJT 1 reaches its now-smaller critical current and turns OFF for both the control and drive ends. As a result, the output is coupled to the negative clock through BiSJT 2 which remains ON having never reached its now larger critical current. Consequently, the output is pulled negative. Similarly, if a negative input is provided, a positive output is achieved.

In contrast to BiSFETs, however, inputs to both the control and drive ends of the BiSJT are coupled to the condensate. The maximum input current to the control end must be limited to below the critical current. Nevertheless, if the output state is still set when the input signal is applied, the input signal may be able to hold output state even absent the clock signal to the output side given hysteresis in these NDR devices.

Thus, in addition to the input being set before the clocks begin their ramp up, the clocks also should be fully ramped down before the next input begins its ramp up. Although not required, such was already archived in BiSFET circuits in principle, but the timing was very close and subject to slight variances in time delays. However, more care should be taken for BiSJT circuits.

Another difference from the BiSFET logic is that there is always one low-resistance path on the input end (also analogous to the BJT logic), making BiSJT less suitable for static memory applications as compared to the BiSFET. However, logic calculations remain ultralow power, as shown in the SPICE simulation results to follow. Indeed, absent concerns over electrostatic variations in the gate region, which are much more difficult to control than voltage variations, it may be possible to operate the proposed BiSJT at substantially lower voltages than the proposed BiSFET for still lower energy operation.

For the inverter element of Fig. 2.34a, the shown input side external resistance represented by  $R_{in}$  combined with the input-end contact resistances of the BiSJT (both perhaps defined by contact dimensions) must be large enough to prevent the  $I_{ctrl}$  alone from reaching the critical current of either BiSJT. Similarly, the shown load resistance represented by  $R_{out}$  combined with the output side contact resistances must be large enough to, after ramping the clocks, prevent  $I_{drv}$  of the nominally ON BiSJT from reaching its critical current. Of course, the contact resistances themselves must also be small enough to allow the critical current to be reached in one device during the clock ramping. A signal follower/buffer can be created similarly by exchanging the signal and ground terminals on the input end, as shown in Fig. 2.34b.

The compact device model used to describe the BiSJT is shown in Fig. 2.35. The charge density in the center of the condensate region, even with exciton current flow, is not a function of the control and drive voltages since the injected carriers only penetrate a short distance into the channel. Otherwise, the Fermi level within the channel lies at the center of the associated anti-crossing band gap, such that the (differential) quantum capacitance, the interlayer capacitance and the capacitive coupling to the gates used to create the condensate are all essentially zero on the scales of interest. Consequently, the  $C_{il}$  represent interlayer capacitances in the leads plus a short extension of a few nanometers (taken as 5 nm here for simulation purposes) into the channel. The  $C_q$  are quantum (density of states) capacitances for each graphene layer over the same region as  $C_{il}$ . The  $R_c$  represent the contact resistances between metal contacts and the 2D layers. The  $R_{leak}$  account for leakage currents mediated by tunneling of normal currents through the condensate-induced band-gap region, which would be quite limited for a fully formed condensate, and normal current flow around the edges of the condensate region, which could be more significant. The  $C_g$  represent (weak) capacitive coupling from the relatively remote gates of fixed voltage, which would be used to electrostatically dope the 2D layers, again subject to penetration of the charge carriers only a short distance into the channel. The near shorting of the 2D layers at each end before the total critical current is reached and the collapse thereafter are

modeled by qualitatively abrupt, but quantitatively smooth, functions  $I_{\text{ctrl/drv}}(V_{\text{cn}}, V_{\text{cp}}, V_{\text{dn}}, V_{\text{dp}})$  as

$$I_{\text{ctrl}} = G_0(V_{\text{cn}} - V_{\text{cp}})K(|V_{\text{cn}} - V_{\text{cp}} + V_{\text{dn}} - V_{\text{dp}}|) \quad (2.20)$$

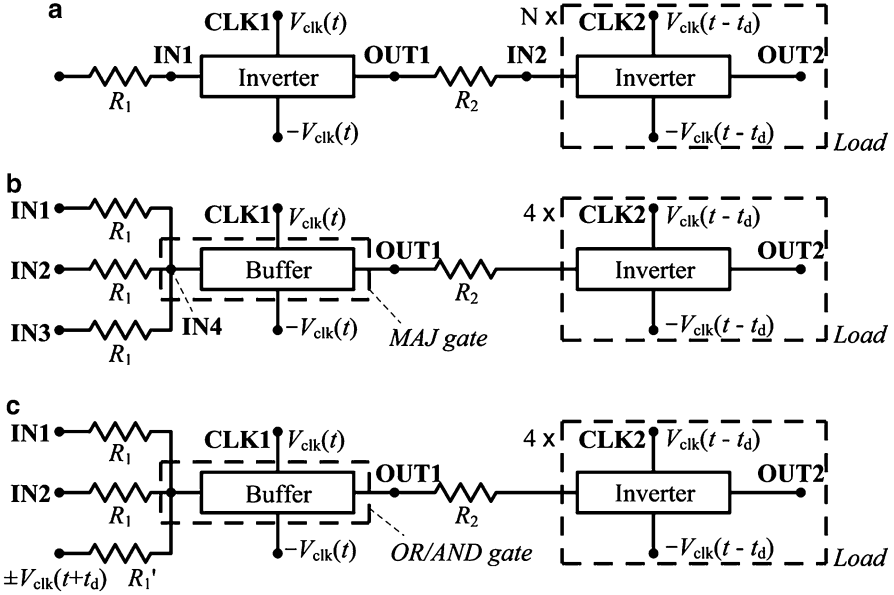
$$I_{\text{drv}} = G_0(V_{\text{dn}} - V_{\text{dp}})K(|V_{\text{cn}} - V_{\text{cp}} + V_{\text{dn}} - V_{\text{dp}}|) \quad (2.21)$$

$$K(V) = e^{-[\max(V, V_{\text{c,ttl}}) - V_{\text{c,ttl}}]^2 / V_0^2} \quad (2.22)$$

where  $G_0$  is the ON-state/enhanced interlayer conductance, approaching the Landauer–Büttiker limit as discussed previously. In these simulations where the device continues to be modeled as symmetric,  $V_{\text{c,ttl}} = I_{\text{c,ttl}}/G_0$  is the *total* critical voltage, which is a function of the adjustable bare interlayer coupling, as well as condensate region size.  $V_0$  controls the sharpness of the NDR region, which is perhaps more important to the reliability of the numerical simulations than to the physical results.

### 2.5.2.2 Complex BiSJT Logic and Circuit Simulation

As discussed previously, to the extent that device-to-device variation can be neglected and current noise is limited, the magnitude of the input signal does not matter toward determining which BiSJT or BiSJTs turn(s) off and which remain (s) on, only the sign of the input. This behavior allows for fan-out, where the more gates are driven by a single gate, the less current flows to each single second-stage gate. In addition, if combined with multiple inputs, this behavior allows for simple realization of majority gates (MAJ) and AND/OR gate by essentially averaging three inputs together at the control-end input to a buffer, as shown in Fig. 2.36 along with fan-out. The complementary NAND/NOR gates can be created similarly by averaging the inputs to an inverter. Consider the MAJ gate of Fig. 2.36b as an example. If two or three of the three *balanced* inputs are HIGH, the effective input is HIGH, even if with two possible strengths. The output is therefore HIGH. Similarly, the output is LOW if two or more of the balanced inputs are LOW. The AND and OR logic can be obtained from the MAJ logic as  $\text{AND}(A,B) = \text{MAJ}(A,B,0)$  and  $\text{OR}(A,B) = \text{MAJ}(A,B,1)$ , where the 0 or 1 reference state is provided directly from the clock of the previous stage, which is synchronized to the other input signals, if through some limiting resistances. Generally speaking, the level of the three input signals, resulting/achieved through fan-out or input resistances, must be of roughly the same scale. At least no one input signal can be larger than the other two combined. Input buffer gates could be used to balance signals if not otherwise possible. One also can imagine more than two distinct physical contacts to the condensate itself but that is, again, beyond the current capabilities of quantum transport simulation. Importantly, in addition to being qualitatively simple, this version of MAJ/AND/OR/NAND/NOR has the same device area as a simple inverter or buffer. This input arrangement does mean that the outputs of previous



**Fig. 2.36** SPICE-simulated logic circuits. (a) Inverter with variable fan-out =  $N$ , (b) majority gate with four-inverter fan-out loads, and (c) OR and AND gates with four-inverter fan-out loads. The majority gate consists simply of a buffer with three different input signals tied to its single input node. One also can imagine more than two distinct physical contacts to the condensate itself, but that is beyond this work as well as our current capabilities of quantum transport simulation. The OR and AND gates are formed by using the positive or negative clocks, respectively, of the previous stage as one of the three inputs, with a slightly increased input resistance  $R_1'$  representing the value of  $R_1$  plus that of a series ON-state BiSJT to match the current level of the other two inputs. Simulation results for these systems are provided in Fig. 2.37. An inverted majority gate and NOR and NAND gates can be obtained simply by replacing the buffer in (b) and (c), respectively, with an inverter.  $t_d = \tau/4$  where  $\tau$  is the clock cycle

logic gates are coupled together, which, beyond requiring some circuit design care, provides additional leakage paths. However, particularly with the possibility of further reduced voltage, low power can be maintained. Note that the sequential clocking scheme provides input–output state isolation essentially independent of any leakage current between the inputs and outputs.

For circuit simulations, with about 20 nm long and 45 nm wide ( $3 \times$  condensate channel length  $L$ ) leads and an interlayer effective oxide thickness (EOT) of 1.77 nm (if closer but of lower dielectric constant),  $C_{\text{il}} = 21.9$  aF for each side of the BiSJT. With  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$  over the same region,  $C_q = 86.6$  aF each. Assuming a (at least) 5 nm EOT for the gate dielectrics used to create the condensate in a conservative estimate and again an approximately 5 nm overlap under the gates into the condensate region for each lead,  $C_g = 1.5$  aF each. Consistent with simulations,  $G_0$  is taken to be 75 % of the Landauer–Büttiker ballistic limit for the graphene doping levels  $n = p \approx 6 \times 10^{12} \text{ cm}^{-2}$  or  $G_0 = 0.70$  mS [46].

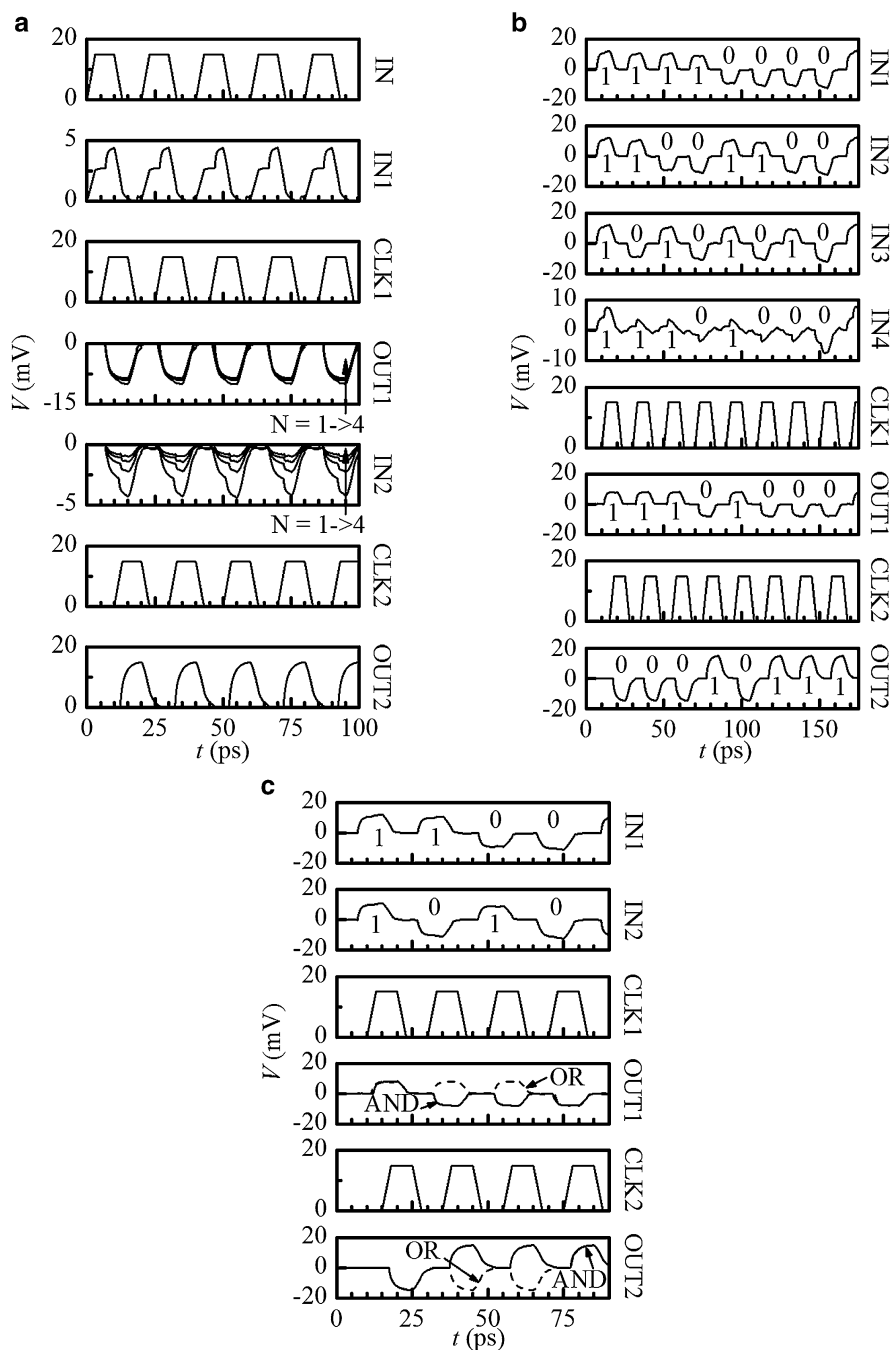
Both the  $R_c$  and  $R_{\text{leak}}$  are taken to be  $6 \text{ k}\Omega$  ( $\sim 270 \text{ }\Omega \mu\text{m}$ , approximately four times the ON-state interlayer resistance  $1/G_0$ ).  $V_{c,\text{ttl}}$  is taken to be  $1.2 \text{ mV}$ . (The abruptness factor  $V_0$  is then taken as  $0.4 \%$  of  $V_{c,\text{ttl}}$ .) For the fan-out circuit of Fig. 2.36a and the majority gate circuit of Fig. 2.36b,  $R_1 = 30 \text{ k}\Omega$  and  $R_2 = 15 \text{ k}\Omega$ . For the OR/AND gates of Fig. 2.36c,  $R_1$  and  $R_2$  remain the same as the MAJ gate, but  $R_1'$  is raised to  $43 \text{ k}\Omega$ , which is the series combination of resistance of  $R_1$  and an ON-state BiSJT in these simulations, to match the current level of the other two inputs.

Figure 2.37 includes simulation results for the BiSJT circuits of Fig. 2.36. In Fig. 2.37a, the fan-out varies from 1 to 4; in Fig. 2.37b, c, the fan-out is fixed at 4. In Fig. 2.37a, the step-like shape of the input end voltage of the BiSJTs is due to the sudden change of input resistance due to the turning off of one of the two BiSJTs. In Fig. 2.37b, despite variations in the input signal strength, the output signal OUT1 depends only on the sign of the input signal IN4, not the magnitude. The clock signals have a peak magnitude of  $15 \text{ mV}$ , a  $50 \text{ GHz}$  frequency, and a one-quarter-cycle delay between gates. IN1-3 in Fig. 2.37b and IN1-2 in Fig. 2.37c are driven by buffer gates not shown. The average energy consumption per BiSJT per clock cycle is  $\sim 60 \text{ zJ}$  for the MAJ based logic and  $\sim 30 \text{ zJ}$  for the inverter, comparable with the BiSFET logic circuits.

### 2.5.3 Comment on Energy and Noise (BiSJT)

Despite the low voltages and energies considered here, BiSFET/BiSJT logic still appears to remain in a realm of reliable operation in terms of fundamental noise and energy benchmarks. For Johnson–Nyquist (thermal) noise associated with the random motions of the carriers due to scattering [53, 54], for example, taking a  $10 \text{ k}\Omega$  resistance  $R$  and a  $10 \text{ GHz}$  clock representing the bandwidth  $\Delta f$  gives a value of voltage fluctuations of  $\Delta V = \sqrt{4k_B T R \Delta f} \cong 0.4 \text{ mV}$ , which remains small compared to the BiSFET/BiSJT critical voltage considered here for the devices as a whole. These particular input parameters are perhaps more representative of TMD systems that we have been looking at recently than of graphene systems. For graphene systems, however, the frequency would be perhaps larger, but the resistance would be much less (and only the resistance associated with the phonon system is relevant, not the Landauer–Büttiker contact resistance). We also note that cross-talk noise should scale with the overall operating voltage of the circuit. Therefore, as a percentage, it should remain constant and, thus, no more problematic for low-voltage circuits than high-voltage ones. In terms of energy, the few tens to 100-plus zJ consumed per logic gate in simulations to establish a binary logical output, although small, nevertheless remain large compared to Landauer’s limit for irreversible computation,  $k_B T \ln(2) \approx 2.9 \text{ zJ}$  at room temperature [55].





**Fig. 2.37** Simulation results for the BiSJT circuits of Fig. 2.36, (a)–(c) here for (a)–(c) of Fig. 2.36, respectively. Note that in (b) the output signal  $OUT1$  depends only on the sign of the input signal  $IN4$ , not the magnitude, showing signal restoration. In (a), the fan-out  $N$  is from 1 to 4. In (c), the dotted and solid lines are for OR and AND gates, respectively. The clock signals have a peak magnitude of 15 mV, a 50 GHz frequency, and a one-quarter-cycle delay between gates.  $IN1$ –3 in (b) and  $IN1$ –2 in (c) are driven by prior buffer gates not shown

## 2.6 Outlook and Conclusion

The goal of the pseudospintronic devices is to restart the “road map” beyond-CMOS. Of course, that is the goal of all “beyond-CMOS” work. However, with zeptojoule-scale switching energies and novel functionality, pseudospintronic devices could add a substantial amount to the road map.

Moreover, only applications to logic circuits have been considered; there should be more opportunities if pseudospintronic devices including but not limited to those discussed here can be realized. Some of these opportunities are suggested by the similarities of the  $I$ - $V$  characteristics of the proposed BiSFETs/BiSJT to those of gated resonant tunneling diodes but with lower voltages and larger ON/OFF ratios and to those of Josephson junctions but with gating and potentially room-temperature operation. The possibility of multiple contacts to the same coherent region of condensation or gate-controlled coupling and decoupling of coherent regions of condensation, even if at higher voltages, also suggests the possibility of quantum computing applications.

Theory continues to suggest that room-temperature superfluidity is possible in dielectrically separated graphene layers, although challenging. If not successful in graphene-based systems or even if successful, analogous systems of 2D materials such as TMDs [16], silicene, and germanene may prove better suited to the task of creating room-temperature condensates, even if for now these materials are more technologically challenging. Work on TMDs is underway by the authors and colleagues. In part, the experimental efforts of Ref. [20] in III-V systems were informed by theory of [40], and that theoretical effort was supported as part of the BiSFET/BiSJT effort. Qualitatively, almost all of the essential physics of interest has been seen experimentally in analogous physical systems including III-V double quantum wells and Josephson junctions, except, of course, room-temperature exciton condensation at the time of writing.

However, pseudospintronic devices are based on exotic physics in exotic material systems based on imperfect theory. Ultimately, the issue must be resolved experimentally. The challenges to observing the condensate are substantial. If we can observe the condensate in the laboratory, much more work will be required to extend those results to commercial production. Moreover, even if successful, initial efforts may not run at voltages comparable to  $k_B T/q$  or lower due to technological limitations. However, technology would be expected to improve over time, much as it did for CMOS.

The latter scenario, however, is precisely the above-stated goal of this work, to restart the “road map,” i.e., to turn making substantial long-term progress into a technological challenge rather than the physical impossibility that it is for CMOS. Moreover, room-temperature superfluidity in nanoscale devices could offer new application avenues, not just ongoing progress along existing avenues.

**Acknowledgment** This work was supported by the South West Academy of Nanoelectronics (SWAN), which in turn is supported by the Semiconductor Research Corporation (SRC) and the National Institute of Standards and Technology (NIST) through the Nanoelectronics Research Initiative (NRI).

## References

1. C.J. Davisson, The diffraction of electrons by a crystal of nickel. *Bell Syst. Tech. J.* **7**, 90 (1928)
2. J. Frenkel, On the transformation of light into heat in solids. I. *Phys. Rev.* **37**, 17 (1931)
3. J. Frenkel, On the transformation of light into heat in solids. II. *Phys. Rev.* **37**, 1276 (1931)
4. L.N. Cooper, Bound electron pairs in a degenerate Fermi gas. *Phys. Rev.* **104**, 1189 (1956)
5. S.N. Bose, Plancks Gesetz und Lichtquantenhypothese. *Z. Phys.* **26**, 178 (1924)
6. A. Einstein, Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* **1**, 3 (1925)
7. J. Bardeen, L.N. Cooper, J.R. Schrieffer, Microscopic theory of superconductivity. *Phys. Rev.* **106**, 162 (1957)
8. D.R. Tilley, J. Tilley, *Superfluidity and superconductivity* (Institute of Physics Publishing, Bristol, 1990)
9. P. Kapitza, Viscosity of liquid helium below the lambda-point. *Nature* **141**, 74 (1938)
10. J.F. Allen, A.D. Misener, Flow phenomena in liquid helium II. *Nature* **142**, 643 (1938)
11. D.D. Osheroff, R.C. Richardson, D.M. Lee, Evidence for a new phase of solid  $\text{He}^3$ . *Phys. Rev. Lett.* **28**, 885 (1972)
12. I.O. Kulik, S.I. Shevchenko, Excitonic ‘superfluidity’ in low-dimensional crystals. *Solid State Commun.* **21**, 409 (1977)
13. J.P. Eisenstein, A.H. MacDonald, Bose-Einstein condensation of excitons in bilayer electron systems. *Nature* **432**, 691 (2004)
14. H. Min, R. Bistritzer, J.-J. Su, A.H. MacDonald, Room-temperature superfluidity in graphene bilayers. *Phys. Rev. B* **78**, 121401(R) (2008)
15. D.S.L. Abergel, M. Rodrigues-Vega, E. Rossi, S. Das Sarma, Interlayer excitonic superfluidity in graphene. *Phys. Rev. B* **88**, 235402 (2013)
16. Private communication with A.H. MacDonald: Exciton condensation in bilayer  $\text{MoS}_2$  system
17. I.B. Spielman, J.P. Eisenstein, L.N. Pfeiffer, K.W. West, Resonantly enhanced tunneling in a double layer quantum Hall ferromagnet. *Phys. Rev. Lett.* **84**, 5808 (2000)
18. E. Tutuc, M. Shayegan, D.A. Huse, Counterflow measurements in strongly correlated GaAs hole bilayers: evidence for electron-hole pairing. *Phys. Rev. Lett.* **93**, 036802 (2004)
19. L. Tiemann, W. Dietsche, M. Hauser, K. von Klitzing, Critical tunneling currents in the regime of bilayer excitons. *New J. Phys.* **10**, 045018 (2008)
20. D. Nandi, A.D.K. Finck, J.P. Eisenstein, L.N. Pfeiffer, K.W. West, Exciton condensation and perfect Coulomb drag. *Nature* **488**, 481 (2012)
21. K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, M.I. Katnelson, I.V. Grigorieva, S.V. Dubonos, A.A. Firsov, Two-dimensional gas of massless Dirac Fermions in graphene. *Nature* **438**, 197 (2005)
22. Y. Zhang, Y.-W. Tan, H.L. Stormer, P. Kim, Experimental observation of the quantum Hall effect and Berry’s phase in graphene. *Nature* **438**, 201 (2005)
23. S. Reich, J. Maultzsch, C. Thomsen, P. Ordejón, Tight-binding description of graphene. *Phys. Rev. B* **66**, 035412 (2002)
24. A.K. Geim, K.S. Novoselov, The rise of graphene. *Nat. Mater.* **6**, 183 (2007)
25. I. Sodemann, D.A. Pesin, A.H. MacDonald, Interaction-enhanced coherence between two-dimensional Dirac layers. *Phys. Rev. B* **85**, 195136 (2012)

26. D. Basu, L.F. Register, D. Reddy, A.H. MacDonald, S.K. Banerjee, Tight-binding study of electron-hole pair condensation in graphene bilayers: gate control and system-parameter dependence. *Phys. Rev. B* **82**, 075409 (2010)
27. G.-B. Liu, W.-Y. Shan, Y. Yao, W. Yao, D. Xiao, Three-band tight-binding model for monolayers of group-VIB transition metal dichalcogenides. *Phys. Rev. B* **88**, 085433 (2014)
28. Feng-Cheng Wu, Fei Xue, A.H. MacDonald, Theory of spatially indirect equilibrium exciton condensates. arXiv:1506.01947. Submitted for publication (2015)
29. X. Xu, W. Yao, D. Xiao, T.F. Heinz, Spin and pseudospins in layered transition metal dichalcogenides. *Nat. Phys.* **10**, 343 (2014)
30. S. Shallcross, S. Sharma, E. Kandelaki, O.A. Pankratov, Electronic structure of turbostratic graphene. *Phys. Rev. B* **81**, 165105 (2010)
31. X. Mou, L.F. Register, S.K. Banerjee, Quantum transport simulation of Bilayer pseudoSpin Field-Effect Transistor (BiSFET) with tight-binding hartree-fock model. IEEE SISPAD 2013, 420, Glasgow, United Kingdom (2013)
32. X. Mou, D. Basu, L.F. Register, S.K. Banerjee, manuscript in preparation
33. M. Cheli, G. Fiori, G. Iannaccone, A semianalytical model of bilayer-graphene field-effect transistor. *Trans. Electron Devices IEEE* **56**, 2979 (2009)
34. S.K. Banerjee, L.F. Register, E. Tutuc, D. Reddy, A.H. MacDonald, Bilayer pseudospin field-effect transistor (BiSFET): a proposed new logic device. *Electron. Device Lett. IEEE* **30**, 158 (2009)
35. P. Avouris, Z. Chen, V. Perebeinos, Carbon-based electronics. *Nat. Nanotechnol.* **2**, 605 (2007)
36. J. Martin, N. Akerman, G. Ulbricht, T. Lohmann, J.H. Smet, K. von Klitzing, A. Yacoby, Observation of electron-hole puddles in graphene using a scan single-electron transistor. *Nat. Phys.* **4**, 144 (2008)
37. D. Basu, L.F. Register, A.H. MacDonald, S.K. Banerjee, Effect of interlayer bare tunneling on electron-hole coherence in graphene bilayers. *Phys. Rev. B* **84**, 035449 (2011)
38. S. Datta, *Electronic transport in mesoscopic systems*. Cambridge studies in semiconductor physics and microelectronic engineering (Cambridge University Press, Cambridge, 1997)
39. E.H. Hwang, S. Das Sarma, Acoustic phonon scattering limited carrier mobility in two-dimensional extrinsic graphene. *Phys. Rev. B* **77**, 115449 (2008)
40. J.-J. Su, A.H. MacDonald, How to make a bilayer exciton condensate flow. *Nat. Phys.* **4**, 799 (2008)
41. K.K. Ng, *Complete guide to semiconductor devices* (Wiley, New York, 2002), pp. 569–574
42. Y. Yu, S. Han, X. Chu, S.-I. Chu, Z. Wang, Coherent temporal oscillations of macroscopic quantum states in a Josephson Junction. *Science* **296**, 889 (2002)
43. X. Mou, L.F. Register, S.K. Banerjee, Quantum transport simulations on the Feasibility of the Bilayer pseudoSpin Field-Effect Transistor (BiSFET). IEDM 2013, 4.7.1–4.7.4, Washington DC, United States (2013)
44. X. Mou, L.F. Register, S.K. Banerjee, Bilayer pseudospin junction transistor (BiSJT): a new “beyond-CMOS” logic device. Submitted for publication (2015)
45. X. Mou, L.F. Register, S.K. Banerjee, Quantum transport simulation of exciton condensate transport physics within a bilayer graphene system. Submitted for publication (2015)
46. X. Mou, L.F. Register, S.K. Banerjee, Interplay among Bilayer PseudoSpin Field-Effect Transistor (BiSFET) performance, BiSFET scale and condensate strength. Accepted by IEEE SISPAD 2014, Yokohama, Japan (2014)
47. L.F. Register, X. Mou, D. Reddy, W. Jung, I. Sodemann, D. Pesin, A. Hassibi, A.H. MacDonald, S.K. Banerjee, Bilayer pseudo-spin field effect transistor (BiSFET): concepts and critical issues for realization. *ECS Trans.* **45**, 3 (2012)
48. K. Wu, A. Sachid, F.-L. Yang, C. Hu, Toward 44% switching energy reduction for FinFETs with vacuum gate spacer. IEEE SISPAD 2012, 253, Denver, Colorado (2012)
49. R.T. Weitz, M.T. Allen, B.E. Feldman, J. Martin, A. Yacoby, Broken-symmetry states in doubly gated suspended bilayer graphene. *Science* **330**, 812 (2010)

50. P. Jadaun, H.C.P. Movva, L.F. Register, S.K. Banerjee, Theory and synthesis of bilayer graphene intercalated with ICl and IBr for low power device applications. *J. Appl. Phys.* **114**, 063702 (2013)
51. D. Reddy, L.F. Register, E. Tutuc, S.K. Banerjee, Bilayer pseudospin field-effect transistor: applications to Boolean logic. *IEEE Trans. Electron. Devices* **57**, 755 (2010)
52. <http://www.itrs.net>
53. J.B. Johnson, Thermal agitation of electricity in conductors. *Phys. Rev.* **32**, 97 (1928)
54. H. Nyquist, Thermal agitation of electric charge in conductors. *Phys. Rev.* **32**, 110 (1928)
55. R. Landauer, Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183 (1961)

Nanoscale Materials and Devices for Electronics,  
Photonics and Solar Energy

Korkin, A.; Goodnick, S.; Nemanich, R. (Eds.)

2015, XII, 282 p. 135 illus., 65 illus. in color., Hardcover

ISBN: 978-3-319-18632-0