Bipolar spintronics: from spin injection to spin-controlled logic

Igor Žutić, Jaroslav Fabian and Steven C Erwin

1 Department of Physics, University of Buffalo, SUNY, Buffalo, NY 14260, USA
2 Institute for Theoretical Physics, University of Regensburg, 93040 Regensburg, Germany
3 Center for Computational Materials Science, Naval Research Laboratory, Washington, DC 20375, USA

Received 22 September 2006, in final form 7 November 2006
Published 6 April 2007
Online at stacks.iop.org/JPhysCM/19/165219

Abstract
An impressive success of spintronic applications has been typically realized in metal-based structures which utilize magnetoresistive effects for substantial improvements in the performance of computer hard drives and magnetic random access memories. Correspondingly, the theoretical understanding of spin-polarized transport is usually limited to a metallic regime in a linear response, which, while providing a good description for data storage and magnetic memory devices, is not sufficient for signal processing and digital logic. In contrast, much less is known about possible applications of semiconductor-based spintronics and spin-polarized transport in related structures which could utilize strong intrinsic nonlinearities in current–voltage characteristics to implement spin-based logic. Here we discuss the challenges for realizing a particular class of structures in semiconductor spintronics: our proposal for bipolar spintronic devices in which carriers of both polarities (electrons and holes) contribute to spin-charge coupling. We formulate the theoretical framework for bipolar spin-polarized transport, and describe several novel effects in two- and three-terminal structures which arise from the interplay between nonequilibrium spin and equilibrium magnetization. (Some figures in this article are in colour only in the electronic version)

1. Introduction
In contrast to well established applications based on metallic magnetic multilayers [1–14], much less is known about the prospect for utilizing semiconductors in spintronic applications. Typically, these commercial metal-based applications rely on magnetoresistive effects and employ two-terminal structures known as the spin-valves in which a nonmagnetic material is sandwiched between two ferromagnetic electrodes. The flow of carriers through a spin-valve is determined by the direction of their spin (up or down) relative to the magnetization of the
device’s electrodes, leading thus to magnetoresistance. Since magnetization in ferromagnets persists even when the power is switched off, these applications have the significant advantage of being nonvolatile. However, for advanced functions, such as signal processing and digital logic, two-terminal devices such as these are of limited use. Spin logic will also require three-terminal devices and could benefit from incorporating semiconductors with their intrinsic nonlinear current–voltage characteristics, suitable for signal amplification.

An early proposal for a semiconductor-based spin-logic device is the Datta–Das spin field effect transistor (FET) [15], depicted in figure 1. While, despite the extensive experimental efforts, there remain important challenges for its realization [14], it is helpful to illustrate a generic scheme for a spin logic device with basic elements such as spin injection and detection as well as spin transport and manipulation. The spin FET, which can be viewed as a gate-controlled (via spin–orbit coupling) spin-valve, has also spurred many related transistor schemes [14, 18–25]. However, a similar functionality has been recently realized in a very different implementation using a carbon nanotube (CNT), rather than a semiconductor [26], as the nonmagnetic material sandwiched between the ferromagnetic source and drain with tunnelling contacts. While a CNT has a negligible spin–orbit coupling, the tunability of both the magnitude and the sign of tunnelling magnetoresistance in such a CNT spin-valve was controlled by gate voltage, which changed on- or off-resonance condition [27, 28]. Another interesting feature of the Datta–Das spin FET is that it shows the importance of magnetic heterojunctions as the building block for various semiconductor spin-based devices. In this article we will review a theory for bipolar spin-polarized transport in magnetic semiconductor heterojunctions and show possible implications for spin injection and spin-controlled logic. The term *bipolar* indicates that carriers of both polarities (electrons and holes) are important. In contrast to unipolar devices, such as metallic spintronic devices [3, 30], bipolar devices exhibit large deviations from local charge neutrality and intrinsic nonlinearities in the current–voltage characteristics, which are important even at small applied bias.

---

4 This is a conventional meaning of the term bipolar, as used in the physics of semiconductors. However, the term *bipolar* has also been used to describe an analogy between the coexistence of two spin carrier populations (of spin up and spin down) in spin-polarized transport and two charge carrier populations (electrons and holes) in bipolar charge transport. See, for example [29].
These characteristics, together with the ease of manipulating the minority charge carriers, enable the design of active devices that can amplify signals—as well as providing additional degrees of control not available in charge-based electronics. Analogous to the bipolar charge transport [31, 32], which is dominated by the influence of the nonequilibrium carrier density, the nonequilibrium spin density (unequal population of ‘spin up’ and ‘spin down’ carriers) plays an important role in bipolar spintronics. A spin light emitting diode (LED), depicted in figure 2, can be viewed as a prototypical bipolar spintronic device. Similar to an ordinary LED [32], electrons and holes recombine (in a quantum well or a p–n junction) and produce electroluminescence. However, in a spin LED, as a consequence of radiative recombination of spin-polarized carriers, the emitted light is circularly polarized and could be used to trace back the degree of polarization of carrier density upon injection into a semiconductor. While spin LEDs may not directly lead to spin logic, they have been widely used as detectors for spin polarization, injected optically or electrically into a semiconductor [33–38].

Another important structure for bipolar spintronics is a semiconductor-based magnetic heterojunction and its special cases such as p–n junctions. In addition to being elements of spin FETs and spin LEDs, as we shall show, they are also the building blocks for bipolar devices which could enable a spin-controlled logic. Early experimental efforts date back to nearly 40 years ago. It was shown that a ferromagnetic p–n junction, based on the ferromagnetic semiconductor CdCr2Se4 doped with Ag acceptors and In donors, could act as a diode. Photo-voltaic diodes were also fabricated using an (Hg, Mn)Te magnetic semiconductor [39]. However, more extensive work on magnetic p–n junctions has begun since the discovery of (III, Mn)V ferromagnetic semiconductors such as (In, Mn)As [40–42], and (Ga, Mn)As [43–45], reviewed in [46–48]. Heavily doped p-(Ga, Mn)As/n-GaAs junctions were fabricated [49–53], to demonstrate tunnelling interband spin injection. Furthermore, it was shown that the current in p-CoMnGe/n-Ge magnetic heterojunction diodes can indeed be controlled by magnetic field [54].

A potentially valuable property for all-semiconductor device designs is the external control of Curie temperature ($T_C$). Carrier-mediated ferromagnetism in dilute magnetic semiconductors such as (In, Mn)As, (Ga, Mn)As, and MnGe [46, 55–58] allows for tuning the strength of the ferromagnetic interactions and, therefore, $T_C$. For example, when the number of carriers is changed, either by shining light [59, 60] or by applying a gate bias in a field effect transistor geometry [61], the material can be switched between the paramagnetic and ferromagnetic states. These experiments suggest the prospect of nonvolatile multifunctional devices with
tunable optical, electrical, and magnetic properties. Furthermore, the demonstration of optically or electrically controlled ferromagnetism provides a method for distinguishing carrier-induced semiconductor ferromagnetism from ferromagnetism that originates from metallic magnetic inclusions [62]. We predict a possibility for the reentrant ferromagnetism in such semiconductors [63]. Thermally excited carriers can enhance exchange coupling with magnetic impurities and lead to the regime in which with the increase of temperature there is an onset of ferromagnetism and enhanced magnetization.

An important challenge for potential spin logic applications is the demonstration of room temperature operations. In all-semiconductor schemes it would be desirable to have ferromagnetic materials with high $T_C$. Some of the promising developments include (Ga, Mn)As with $T_C \sim 250$ K [64] and (Zn, Cr)Te with $T_C \sim 300$ K [65]. However, there is a wide range of other materials with much higher predicted and/or reported $T_C$ [47, 66, 67] which need to be critically examined and their potential tested in the actual device structures.

An alternative route to room temperature operation is the use of hybrid structures that combine metallic ferromagnets with high $T_C$ and semiconductors. It is important to note that in such systems tailoring of interfacial properties can significantly improve magnetoresistive effects or spin injection efficiency [68–71]. For example, use of MgO (instead of Al$_2$O$_3$) as a tunnel barrier between CoFe electrodes in a magnetic tunnel junction can lead to a dramatic increase in room temperature tunnelling magnetoresistance [72, 73], confirming previous theoretical predictions [74, 75]. Furthermore, it was demonstrated that employing a CoFe/MgO tunnel injector can provide robust room temperature spin injection in semiconductors such as GaAs [76, 77] with room temperature spin polarization of injected electrons exceeding 70% [78].

We first formulate drift–diffusion equations for bipolar spin-polarized transport. Next we consider spin injection and extraction in a magnetic p–n junction as well as an interplay between equilibrium magnetization and the injected nonequilibrium spin, which leads to a strong spin–charge coupling. In the last section we review the basics of the bipolar junction transistor and our proposal for its generalization—the magnetic bipolar transistor.

2. Bipolar spin-polarized transport

2.1. Spin-polarized drift–diffusion equations

Spin-polarized bipolar transport can be thought of as a generalization of its unipolar counterpart. Specifically, a spin-polarized unipolar transport, in a metallic regime, can then be obtained as a limiting case by setting the electron–hole recombination rate to zero and considering only one type of carriers (either electrons or holes). In the absence of any spin polarization, equations which aim to describe spin-polarized bipolar transport need to recover a description of charge transport. A conventional charge transport in semiconductors is often accompanied by large deviations from local charge neutrality (for example, due to materials inhomogeneities, interfaces, and surfaces) and Poisson’s equation needs to be explicitly included. If we consider (generally inhomogeneous) doping with density of $N_a$ ionized acceptors and $N_d$ donors we can then write

$$\nabla \cdot (\epsilon \nabla \phi) = q(n - p + N_a - N_d),$$

where $n$ and $p$ (electron and hole densities) also depend on the electrostatic potential $\phi$ and permittivity $\epsilon$ can be spatially dependent. In contrast to the metallic regime, even equilibrium carrier density can have large spatial variations, which can be routinely tailored by the appropriate choice of the doping profile [$N_d(x) - N_a(x)$]. Furthermore, charge transport
in semiconductors can display strong nonlinearities, for example, exponential-like current-voltage dependence of a diode [32].

We briefly recall here a case of a unipolar spin-polarized transport in a metallic regime. The basic theoretical understanding dates back to Mott [79]. He noted that the electrical current in ferromagnets could be expressed as the sum of two independent and unequal parts for two different spin projections, implying that the current is spin polarized. We label spin-resolved quantities by $\lambda = 1$ or $\uparrow$ for spin up, $\lambda = -1$ or $\downarrow$ for spin down along the chosen quantization axis. For a free electron, spin angular momentum and magnetic moment are in opposite directions, and what precisely is denoted by ‘spin up’ varies in the literature [80]. Conventionally, in metallic systems [81], spin up refers to carriers with majority spin. This means that the spin (angular momentum) of such carriers is antiparallel to the magnetization. Some care is needed with the terminology used for semiconductors; the terms majority and minority there refer to the relative population of charge carriers (electrons or holes). Spin-resolved charge current (density) in a diffusive regime can be expressed as

$$j_\lambda = \sigma_\lambda \nabla \mu_\lambda,$$

where $\sigma_\lambda$ is conductivity and the chemical potential (sometimes also referred to as the electrochemical potential) is

$$\mu_\lambda = (q D_\lambda/\sigma_\lambda) \delta n_\lambda - \phi,$$

with $q$ the proton charge, $D_\lambda$ the diffusion coefficient, $\delta n_\lambda = n_\lambda - n_{\lambda,0}$ the change of electron density from the equilibrium value for spin $\lambda$, and $\phi$ the electric potential. We use a notation in which a general quantity $X$ is expressed as a sum of equilibrium and nonequilibrium parts, $X = X_0 + \delta X$. Here we focus on the case of a collinear magnetization. More generally, for a noncollinear magnetization, $j_\lambda$ becomes a second-rank tensor [82, 83].

In the steady state the continuity equation is

$$\nabla j_\lambda = q \left[ \frac{\delta n_\lambda}{\tau_{\lambda\lambda}} - \frac{\delta n_{-\lambda}}{\tau_{-\lambda\lambda}} \right],$$

and $\tau_{\lambda\lambda}$ is the average time for flipping a $\lambda$-spin to $\lambda'$-spin. For a degenerate conductor the Einstein relation is

$$\sigma_\lambda = q^2 N_\lambda D_\lambda,$$

where $\sigma = \sigma_\uparrow + \sigma_\downarrow$ and $N = N_\uparrow + N_\downarrow$ is the density of states. Using a detailed balance $N_\lambda/\tau_{\uparrow\downarrow} = N_{-\lambda}/\tau_{\downarrow\uparrow}$ [84] together with equations (3) and (5), the continuity equation can be expressed as [85, 86]

$$\nabla j_\lambda = q^2 \frac{N_\lambda N_{-\lambda}}{N_\uparrow + N_\downarrow} \frac{\mu_\lambda - \mu_{-\lambda}}{\tau_s},$$

where $\tau_s = \tau_{\uparrow\uparrow} \tau_{\downarrow\downarrow}/(\tau_{\uparrow\downarrow} + \tau_{\downarrow\uparrow})$ is the spin relaxation time. Equation (6) implies the conservation of charge current $j = j_\uparrow + j_\downarrow = \text{const}$, while the spin counterpart, the difference of the spin-polarized currents $j_\lambda = j_\uparrow - j_\downarrow$, is position dependent.

Following the work of Mott, a unipolar spin-polarized transport and spin injection in the metallic regime is usually described using equivalent resistor schemes with two resistors of different magnitudes, one for each spin direction, also known as the ‘two-current model’ [14, 87–89]. This approach implies a linear response with injected spin polarization and assumes that there are no interfacial spin-flip processes. However, the latter assumption, widely used since the first demonstration of spin injection in metals [90], may need to be reconsidered [85] when analysing room temperature spin injection experiments [91, 92].

Returning to the case of spin-polarized transport in semiconductors, we formulate a drift-diffusion model which will generalize the considerations of equation (2)–(6) to include both
Figure 3. Band-energy schemes for a magnetic heterojunction. In equilibrium chemical potential $\mu_0$ is constant. Conductance and valence-band edges ($E_c$ and $E_v$) are spin split in the magnetic p region, while in the nonmagnetic n region there is no spin splitting. Left and right edges of a space-charge (depletion) region are denoted by $x_L$ and $x_R$. For a sharp doping profile, at $x = w/2$, there are generally discontinuities in conduction and valence bands ($\Delta E_c$ and $\Delta E_v$) and in other quantities, such as effective mass, permittivity, and diffusion coefficient. Adapted from [97].

electrons and holes [93–95]. We recall that from equations (2) and (3) spin-resolved current has a drift part (proportional to electric field, i.e. $\propto \nabla \phi$) and a diffusive part ($\propto \nabla n_{\lambda}$), which we want to extend to also capture the effects of band bending, band offsets, various materials inhomogeneities, and the presence of two types of charge carriers. For nondegenerate doping levels (Boltzmann statistics) the spin-resolved components are

$$n_{\lambda} = \frac{N_{c,v}}{2} e^{-[E_{c,v\lambda} - \mu_{n,p\lambda}]/k_B T}, \quad p_{\lambda} = \frac{N_{c,v}}{2} e^{-[\mu_{n,p\lambda} - E_{c,v\lambda}]/k_B T}, \quad (7)$$

where subscripts c and v label quantities which pertain to the conduction and valence bands. For example, $N_{c,v} = 2(2\pi m_{c,v}^* k_B T/\hbar^2)^{3/2}$ are effective densities of states with the corresponding effective masses $m_{c,v}^*$ and $k_B$ is the Boltzmann constant. From the total electron density $n = n_\uparrow + n_\downarrow$ and the spin density $s = n_\uparrow - n_\downarrow$, we can define the spin polarization of electron density as

$$P_n = \frac{s}{n} = \frac{n_\uparrow - n_\downarrow}{n_\uparrow + n_\downarrow}. \quad (8)$$

Such a finite spin polarization does not necessarily require ferromagnetic materials or external magnetic fields at all. For example, circularly polarized light provides an effective way to generate net spin polarization in direct-bandgap semiconductors. The angular momentum of the absorbed light is transferred to the medium; this leads directly to orientation of the electron orbital momenta and, through spin–orbit interaction, to polarization of the electron spins [96]. In bulk III–V semiconductors, such as GaAs, optical orientation can lead to 50% polarization of the electrons; this can be further enhanced by using quantum structures of reduced dimensionality, or by applying strain.

We consider a general case where the spin splitting of conduction and valence bands, expressed as $2q\zeta_c$ and $2q\zeta_v$, can be spatially inhomogeneous [94]. Splitting of carrier bands (Zeeman or exchange) can arise due to doping with magnetic impurities and/or applied magnetic field. The spin-$\lambda$ conduction band edge (see figure 3)

$$E_{c\lambda} = E_{c0} - q\phi - \lambda q\zeta_c \quad (9)$$

differs from the corresponding nonmagnetic bulk value $E_{c0}$ due to electrostatic potential $\phi$ and the spin splitting $\lambda q\zeta_c$. The discontinuity of the conduction band edge is denoted by $\Delta E_c$. In the nonequilibrium state a chemical potential for $\lambda$-electrons is $\mu_{n,\lambda}$ and generally differs
from the corresponding quantity for holes. While \( \mu_{\text{el}} \) has an analogous role to electrochemical potential in equations (2) and (3), following the conventional semiconductor terminology, we refer to it here as the chemical potential, which is also known as the quasi-Fermi level. An analogous notation holds for the valence band and holes. For example, in equation (7) \( p_\lambda \) is the spin-\( \lambda \) density of holes with \( E_{v\lambda} = E_{v0} - q\phi - \lambda q\zeta_c \).

By assuming the drift–diffusion-dominated transport across a heterojunction, the spin-resolved charge current densities can be expressed as [97]

\[
\mathbf{j}_{n\lambda} = \bar{\mu}_{n\lambda} n_\lambda \nabla E_{n\lambda} + qD_{n\lambda} N_c \nabla (n_\lambda / N_c),
\]

(10)

\[
\mathbf{j}_{p\lambda} = \bar{\mu}_{p\lambda} p_\lambda \nabla E_{p\lambda} - qD_{p\lambda} N_v \nabla (p_\lambda / N_v),
\]

(11)

where \( \bar{\mu} \) and \( D \) are mobility and diffusion coefficients (we use symbol \( \bar{\mu} \) to distinguish it from chemical potential \( \mu \)). We note that ‘drift terms’ have quasi-electric fields \( \nabla E_{n\lambda,v\lambda} \) that are generally spin dependent (\( \nabla \zeta_{c,v} \neq 0 \) is referred to as a magnetic drift [94]) and different for conduction and valence bands. In contrast to homojunctions, additional ‘diffusive terms’ arise due to the spatial dependence of \( m_{c,v} \), and therefore of \( N_{c,v} \). In nondegenerate semiconductors \( \bar{\mu} \) and \( D \) are related by Einstein’s relation

\[
\bar{\mu}_{n,p\lambda} = qD_{n,p\lambda} / k_B T,
\]

(12)

which differs from the metallic (completely degenerate) case given by equation (5).

With two types of carriers the continuity equations are more complex than those in metallic systems. After including additional terms for recombination of electrons and holes as well as photoexcitation of electron–hole pairs, we can write these equations as

\[
-\frac{dn_\lambda}{dt} + \nabla \cdot \frac{j_{n\lambda}}{q} = +r_s(n_\lambda p_\lambda - n_{\lambda0} p_{\lambda0}) + \frac{n_\lambda - n_{-\lambda} - \lambda S_n}{2\tau_{mn}} - G_n,
\]

(13)

\[
+\frac{dp_\lambda}{dt} + \nabla \cdot \frac{j_{p\lambda}}{q} = -r_s(n_\lambda p_\lambda - n_{\lambda0} p_{\lambda0}) - \frac{p_\lambda - p_{-\lambda} - \lambda S_p}{2\tau_{mp}} + G_p.
\]

(14)

Generation and recombination of electrons and holes of spin \( \lambda \) can be characterized by the rate coefficient \( r_s \), the spin relaxation time for electrons and holes is denoted by \( \tau_{mn,p} \), and the photoexcitation rate \( G_{\lambda} \) represents the effects of electron–hole pair generation and optical orientation. Spin relaxation equilibrates carrier spin while preserving nonequilibrium carrier density, and for nondegenerate semiconductors \( \tilde{s}_n = n P_{n0} \), where from equation (7) an equilibrium polarization of electron density is

\[
P_{n0} = \tanh(q\zeta_c/k_B T),
\]

(15)

and an analogous expression holds for holes and \( \tilde{s}_p \).

The system of drift–diffusion equations (Poisson and continuity equations) can be self-consistently solved numerically [93, 94, 98] and under simplifying assumptions (similar to the case of charge transport) analytically [95, 97, 99]. Heterojunctions, such as the one sketched in figure 3, can be thought of as building blocks of bipolar spintronics. To obtain a self-consistent solution in such a geometry, only the boundary conditions at \( x = 0 \) and \( w \) need to be specified. On the other hand, for an analytical solution we also need to specify the matching conditions at \( x_L \) and \( x_R \), the two edges of the space charge region (or depletion region), in which there is a large deviation from the local charge neutrality, accompanied by a band bending and strong built-in electric field.

We illustrate how the matching conditions for spin and carrier density can be applied within the small-bias or low-injection approximation, widely used to obtain analytical results for charge transport [32, 100]. In this case nonequilibrium carrier densities are small compared to the density of majority carriers in the corresponding semiconductor region. For materials...
such as GaAs a small-bias approximation gives a good agreement with the full self-consistent solution up to approximately 1 V [94, 98]. To simplify our notation, we consider a model where only electrons are spin polarized ($p_\uparrow = p_\downarrow = p/2$), while it is straightforward to also include spin-polarized holes [95, 97]. Outside the depletion charge region materials parameters (such as $N_a$, $N_d$, $N_c$, $N_v$, $\bar{\mu}$, and $D$) are taken to be constant. The voltage drop is confined to the depletion region, which is highly resistive and depleted from carriers. In thermal equilibrium ($\mu_n = \mu_p = \bar{\mu}$) the built-in voltage $V_{bi}$ can be simply evaluated from equation (7) as

$$V_{bi} = \phi_0 R - \phi_0 L,$$

(16)

while the applied bias $V$ (taken to be positive for forward bias) can be expressed as

$$V = - (\delta \phi_R - \delta \phi_L),$$

(17)

implying that the total junction potential between $x = 0$ and $w$ is $V - V_{bi}$. For a heterojunction sketched in figure 3 the width of a depletion (space-charge) region is

$$x_R - x_L \propto \sqrt{V_{bi} - V},$$

(18)

where the built-in voltage is $q V_{bi} = - \Delta E_c + k_B T \ln(n_0 R N_c R / n_0 L N_c L)$. Outside the depletion region the system of drift–diffusion equations reduces usually to only diffusion equations for spin density and the density of minority carriers, while the density of majority carriers is simply given by the density of donors and acceptors [94, 95]. These diffusion equations contain spin and charge diffusion lengths

$$L = \sqrt{D \tau},$$

(19)

in which $L$ would provide a characteristic length scale for the spatial decay of nonequilibrium spin or charge by substituting for $D$ the appropriate (electron or hole) diffusion coefficient and for $\tau$ (spin or charge) the characteristic timescale. However, there are situations, due to additional effects of spin–orbit coupling or simultaneous spin polarization of electrons and holes in magnetic semiconductors, in which diffusion equations become more complicated and equation (19) needs to be generalized [97, 101]. For several decades the techniques of optical orientation have been used to directly measure the characteristic timescale for the decay of nonequilibrium electron spin [96], reaching up to 30 ns [102]. More recent optical measurements have shown at low temperatures even longer spin lifetime in GaAs (>40 ns) [103] and (>100 ns) [104, 105], which could reach ~1 ns at room temperature. Spin–orbit coupling in the valence band typically leads to much faster spin relaxation of holes than electrons (spin lifetimes are three to four orders of magnitude shorter in GaAs at room temperature [106]), further supporting our approximation of spin-unpolarized holes. The related issues of spin relaxation and spin dephasing in GaAs have been extensively reviewed in [14].

From equation (7) we rewrite electron density by separating various quantities into equilibrium and nonequilibrium parts as

$$n_\lambda = n_{\lambda,0} \exp[q(\delta \phi + \delta \mu_\lambda)/k_B T],$$

(20)

and electron carrier and spin density (for simplicity we omit subscript ‘$n$’ when writing $s = n_\uparrow - n_\downarrow$) can be expressed as [95]

$$n = e^{(\delta \phi + \delta \mu_\uparrow)/k_B T} \left[ n_{\uparrow,0} \cosh \left( \frac{q \mu_\uparrow}{k_B T} \right) + s_0 \sinh \left( \frac{q \mu_\uparrow}{k_B T} \right) \right],$$

(21)

$$s = e^{(\delta \phi + \delta \mu_\uparrow)/k_B T} \left[ n_{\uparrow,0} \sinh \left( \frac{q \mu_\uparrow}{k_B T} \right) + s_0 \cosh \left( \frac{q \mu_\uparrow}{k_B T} \right) \right],$$

(22)
where $\mu_\pm \equiv (\mu_{n\uparrow} \pm \mu_{n\downarrow})/2$, and the polarization of electron density is

$$P_n = \frac{\tanh(q\mu_-/k_BT) + P_{n0}}{1 + P_{n0}\tanh(q\mu_-/k_BT)}.$$  \hspace{1cm} (23)$$

If we assume that the spin-resolved chemical potentials are constant for $x_L \leq x \leq x_R$ (which means that the depletion region is sufficiently narrow so that the spin relaxation and carrier recombination can be neglected there) it follows, from equation (23) and $\tanh(q\mu_-/k_BT) \equiv \text{const}$, that

$$P_n^L = \frac{P_{n0}^L[1 - (P_{n0}^R)^2] + \delta P_n^R(1 - P_{n0}^L P_{n0}^R)}{1 - (P_{n0}^R)^2 + \delta P_n^R (P_{n0}^L - P_{n0}^R)}.$$  \hspace{1cm} (24)$$

where L (left) and R (right) label the edges of the space-charge (depletion) region of a p–n junction. Correspondingly, $\delta P_n^R$ represents the nonequilibrium electron polarization, evaluated at $R$, arising from a spin source. For a homogeneous equilibrium magnetization ($P_{n0}^L = P_{n0}^R$), $\delta P_n^L = \delta P_n^R$; the nonequilibrium spin polarization is the same across the depletion region. Equation (24) demonstrates that only nonequilibrium spin, already present in the bulk region, can be transferred through the depletion region at small biases [93–95].

Our assumption of constant spin-resolved chemical potentials is a generalization of a conventional models for charge transport in which both $\mu_n$ and $\mu_p$ are assumed to be constant across the depletion region [100]. From equations (17), (21), and (22) we can obtain minority carrier and spin densities at $x = x_L$

$$n_L = n_{0L} e^{qV/k_BT} \left[1 + \delta P_n^R \frac{P_{n0}^L - P_{n0}^R}{1 - (P_{n0}^R)^2} \right],$$  \hspace{1cm} (25)$$

$$s_L = s_{0L} e^{qV/k_BT} \left[1 + \delta P_n^R \frac{1 - P_{n0}^L P_{n0}^R}{P_{n0}^L - (P_{n0}^R)^2} \right],$$  \hspace{1cm} (26)$$

which in the absence of nonequilibrium spin ($\delta P_n^R = 0$) reduce to the well known Shockley relation for the minority carrier density at the depletion region [31]

$$n_L = n_{0L} e^{qV/k_BT},$$  \hspace{1cm} (27)$$

and an analogous formula holds for spin

$$s_L = s_{0L} e^{qV/k_BT}.$$  \hspace{1cm} (28)$$

2.2. Magnetic p–n junctions

Even in nonmagnetic p–n junctions the presence of nonequilibrium spin (created electrically or optically) can have interesting implications. By the term nonmagnetic we imply the limit of vanishing equilibrium magnetization or, equivalently, vanishing spin polarization since $\xi_{c,v} = 0$. We have predicted that the nonequilibrium spin polarization is bias dependent. By analogy with junction capacitance, this effect could be called spin capacitance as the amount of accumulated spin changes with applied bias. In contrast to the usual monotonic spatial decay of spin density in the nonmagnetic metal [14], away from the point of spin injection, in inhomogeneously doped semiconductors (such as p–n junctions) the spatial profile can be qualitatively different. Spin density can even increase inside the nonmagnetic region, away from the point of spin injection, which we refer to as (spatial) spin density amplification [93, 107]. Illumination of a p–n junction by circularly polarized light can lead to spin electromotive force (EMF) to generate spin-polarized currents even at no applied bias and to provide an open-circuit voltage. In addition to our proposal for a p–n-junction-based spin-polarized solar battery [98], there is a range of other structures which could be used as a
source of spin EMF [108–110] or which could display the effects of spin capacitance [111] and spatial spin amplification [108, 109, 112].

To take additional advantage of manipulating spin degrees of freedom, which could lead to strong spin–charge coupling and potential applications [14], it is useful to consider semiconductor structures with equilibrium magnetization. Such magnetization or, equivalently, band offsets (see figure 3) are negligible and the spin polarization of holes can be neglected, using electrical means. We also focus here on a diffusive regime while a magnetic diode in a ballistic carrier spin-subband splitting and the nonequilibrium spin induced, for example, by optical or electrical means. In narrow-bandgap semiconductors (in InSb |g| ≈ 500 even at room temperature). In the absence of magnetic field, equilibrium magnetization and spin splitting can be realized using ferromagnetic semiconductors. Magnetic impurities and/or an application of an inhomogeneous magnetic field could be used to obtain a desirable, spatially inhomogeneous, spin splitting. Inhomogeneous spin splitting can also occur in domain walls, discussed, for example, in [116]. By solving a system of drift–diffusion and Poisson equations, one can show that an inhomogeneous spin splitting leads to deviations from local charge neutrality [95].

We discuss several properties of magnetic p–n junctions which rely on the interplay of the carrier spin-subband splitting and the nonequilibrium spin induced, for example, by optical or electrical means. We also focus here on a diffusive regime while a magnetic diode in a ballistic regime was recently discussed in [117]. For simplicity, we look at a particular case where the band offsets (see figure 3) are negligible and the spin polarization of holes can be neglected, and both in the notation for the carrier spin splitting 2qξ and for the spin density s we can omit index n. From equations (7) and (9) we can rewrite the product of equilibrium densities as

\[ n_0 p_0 = n_i^2 \cosh(q\xi/k_BT) \]  

(29)

where \( n_i \) is the intrinsic (nonmagnetic) carrier density [100] and we notice that the density of minority carriers in the p region will depend on the spin splitting \( n_0(\xi) = n_0(\xi = 0) \cosh(q\xi/k_BT) \). In figure 4 we illustrate bipolar spin-polarized transport across a magnetic p–n junction under applied forward bias. Calculations are performed using a self-consistent solution of a system of drift–diffusion and Poisson equations. The parameters taken for \( w = 12 \mu \text{m} \) long junctions are based on GaAs-like material doped with \( N_i = 3 \times 10^{15} \text{cm}^{-3} \) acceptors to the left and \( N_d = 5 \times 10^{15} \text{cm}^{-3} \) donors to the right. Diffusion coefficients are \( D_{n}^L = 10 D_{n}^R = 103.6 \text{cm}^2 \text{s}^{-1} \), the intrinsic carrier density is \( n_i = 1.8 \times 10^{6} \text{cm}^{-3} \), the permittivity is \( \epsilon = 13.1 \), the recombination rate coefficient is \( r_f = r_i = (2/3) \times 10^{-5} \text{cm}^3 \text{s}^{-1} \), and the spin relaxation time is \( \tau_{sn} = 0.2 \text{ ns}. \) The minority diffusion lengths are [93, 98] \( L_n = 1 \mu \text{m}, L_p = 0.25 \mu \text{m}, \) and the electron spin diffusion length in the n (p-)region is \( L_{sn} = 1.4 \mu \text{m} (L_{sp} = 0.8 \mu \text{m}). \)

We first ask whether spin can be injected and extracted into/from the nonmagnetic region. At small bias (\( V < V_{bi} \approx 1.1 \text{ V} \)) there is no spin injection or extraction. As the bias increases, the injection and extraction become large and are further enhanced with \( \tau_{sn} \). The reason why there is no spin injection or extraction at small bias is that although there are exponentially more spin up than spin down electrons (recall the Boltzmann statistics) in the magnetic side, the barrier for crossing the space-charge region is exponentially larger for spin up than for spin down electrons (see figure 4). Those two exponential effects cancel out, leaving no net spin current flowing through the space-charge region. We could examine these arguments with analytical findings from equation (24), also valid for the low-bias regime. Considering spin injection in the nonmagnetic n region (\( p_{n1}^R = 0 \)) we see that in the absence of nonequilibrium spin polarization at \( x = x_R, \delta p_{n1}^R = 0 \), there is indeed no spin injection: \( p_{n1}^L = P_{n1}^L \). This is in contrast to arguments which suggest that an inefficient spin injection arises from resistance.
Figure 4. Spin injection and extraction in magnetic $p$–$n$ junctions. (a) Band-energy diagram with spin-polarized electrons (arrows) and unpolarized holes (circles) for spin injection. Spin polarization profiles for different applied biases (labelled by the numbers in volts) show that the spin polarization injected in the nonmagnetic region increases with bias and spin relaxation time $\tau_{sn}$. For $V \lesssim 0.8$ V there is only a negligible spin injection in the $p$ region while there is a sizeable equilibrium spin polarization in the magnetic $n$ region, determined by the spin splitting of $q\zeta = 0.5k_B T$. (b) Band-energy diagram for spin extraction. Electrons from the nonmagnetic $n$ region preferentially populate a lower spin level in the magnetic $p$ region, which leads to spin extraction. In contrast to spin injection, spin polarization profiles show opposite signs in the nonmagnetic and magnetic regions. Adapted from [94].

(conductivity) mismatch [118]. Here we note that even with a good conductivity match and highly polarized spin injector the spin injection can still be completely negligible. Furthermore, the efficiency of spin injection depends strongly on the applied bias rather than on the relative conductivities of the two regions.

At large bias self-consistent numerical results become indispensable, showing that spin injection/extraction is possible as a result of building up a nonequilibrium spin at the space-charge region. However, some of these trends, including our prediction for spin extraction, can already be seen analytically in the low bias regime. Using the previous assumption that $\mu_{n\uparrow,\downarrow}$ are constant for $x_L \leq x \leq x_R$ and by solving diffusion equations for $x < x_L$ and $x > x_R$ we can obtain [95]

$$\delta s_R = -\gamma_3 s_{0L} e^{qV/k_B T},$$

where

$$\gamma_3 = \left( \frac{D_L^L}{D_R^R} \right) \left( \frac{L_{1n}}{L_{2n}} \right) \frac{\tanh[(w - x_R)/L_{1n}]}{\tanh(x_L/L_{1n})},$$

implies that a contribution to the nonequilibrium spin in the $n$ region will have the opposite sign to that of the equilibrium spin in the $p$ region. Similar spin extraction was recently observed experimentally in MnAs/GaAs junctions [119], and related theoretical implications due to tunnelling from nonmagnetic semiconductors into metallic ferromagnets were considered [120]. Furthermore, it was suggested that a combination of spin injection and spin extraction could lead to completely spin polarized carriers in semiconductor nanostructures [121, 122].
We next consider a simple scheme of a magnetic p–n junction, depicted in figure 5, in which there is an external source of nonequilibrium spin, induced optically or electrically. As we discuss below, an interplay between the equilibrium spin polarization and the nonequilibrium spin source leads to the spin-voltaic effect (a spin analogue of the photo-voltaic effect) and to giant magnetoresistance.

Similar to the theory of charge transport in nonmagnetic junctions [31], the total charge current can be expressed as the sum of minority carrier currents at the depletion edges

\[ j = j_{\text{NL}} + j_{\text{PR}} \]

with

\[ j_{\text{NL}} \propto \delta n_{\text{L}}, \quad j_{\text{PR}} \propto \delta p_{\text{R}}, \]

where \( \delta n_{\text{L}} \) is given by equation (25) with \( P_{n0} = 0 \), \( \delta p_{\text{R}} = p_{0}\exp(qV/k_BT) - 1 \), and \( V \) is the applied bias (positive for forward bias). Equation (29) implies that in the regime of large spin splitting, \( q\zeta > k_BT \), the density of minority electrons changes exponentially with \( B (\propto \zeta) \) and can give rise to exponentially large magnetoresistance [94].

The interplay between the \( P_{n0} \) (recall equation (15)) in the p region, and the nonequilibrium spin source of polarization \( \delta P_n \) in the n region, at the edge of the depletion region, modifies the \( I-V \) characteristics of the diodes. To illustrate the \( I-V \) characteristics of the magnetic p–n junction, consider the small-bias limit (recall equations (16)–(28)) in the configuration of figure 5. The electron contribution to the total electric current can be expressed from equations (25) and (32) as [94, 95]

\[ j_{\text{NL}} \sim n_0(\zeta) \left[ e^{qV/k_BT} (1 + \delta P_n P_{n0}) - 1 \right]. \]

Equation (33) generalizes the Silsbee–Johnson spin-charge coupling [90, 123], originally proposed for ferromagnet/paramagnet metal interfaces, to the case of magnetic p–n junctions. The advantage of the spin-charge coupling in p–n junctions, as opposed to metals or degenerate systems, is the nonlinear voltage dependence of the nonequilibrium carrier and spin densities [94, 95], allowing for the exponential enhancement of the effect with increasing \( V \).
Equation (33) can be understood qualitatively from figure 5. In equilibrium, \( \delta P_n = 0 \) and \( V = 0 \), no current flows through the depletion region, as the electron currents from both sides of the junction balance out. The balance is disturbed either by applying bias or by selectively populating different spin states, making the flow of one spin species greater than that of the other. In the latter case, the effective barriers for crossing of electrons from the n to the p side are different for spin up and down electrons (see figure 5). Current can flow even at \( V = 0 \) when \( \delta P_n \neq 0 \). This is an example of the spin-voltaic effect, in which nonequilibrium spin causes an EMF [94, 124] and the direction of the zero-bias current is controlled by the relative sign of \( P_{n0} \) and \( \delta P_n \). We emphasize here that the spin-voltaic effect results in a build-up of electrical voltage due to the proximity of the equilibrium and nonequilibrium spin. This effect is distinct from the so-called spin Hall effect(s), which results in a build-up of a spin imbalance (different chemical potential for spin up and down), but no electric field, due to transport in a spin–orbit field [101].

A straightforward method for detecting the spin-voltaic effect follows from the symmetry properties of the different contributions to the charge current under magnetization reversal. By reversing the equilibrium spin polarization using a modest external magnetic field \( (P_{n0} \rightarrow -P_{n0}) \) a part of \( j_{nL} \), odd under such reversal, can be identified as the spin-voltaic current \( j_{sv} \) [94]. Measurements of \( j(V, P_{n0}) - j(V, -P_{n0}) = 2j_{sv}(V, P_{n0}) \) would then provide (1) cancellation of contributions to the charge current that are not related to the injected spin and (2) a choice of \( V \) to facilitate a sufficiently large \( j_{sv} \) for accurate detection. Unlike the spin LEDs, this approach does not rely on direct bandgap material and injected spin could be detected even in silicon [97]. Magnetic semiconductors approximately lattice matched with Si could be used for spin injection and detection ((Ga, Mn)As has already been grown on Si [125]). For example, the Mn-doped chalcopyrite ZnGeP2 (mismatch < 2%) [126] has been reported to be ferromagnetic at room temperature. Another Mn-doped chalcopyrite, ZnSiP2, was recently predicted [67] to be ferromagnetic, as well as highly spin polarized and closely lattice matched with Si (mismatch < 1%). Mn doping of the chalcopyrite alloy ZnGe1-xSi_xP2 would likely lead to an exact lattice match, since the lattice constant of Si is between those of closely matched ZnSiP2 and ZnGeP2.

Additionally, from \( j_{sv}(V) \) one could also determine a spin relaxation time by all-electrical means [99]. A particular assumption of a magnetic homojunction is not essential. One could also generalize this analysis to heterojunctions, which would include (see figure 3) band offsets and spin splitting in both conduction and valence bands.

Several experimental efforts have recently observed the spin-voltaic effect in semiconductor heterojunctions. One of the approaches used a p–n (In, Ga)As/(Al, Ga)As heterojunction [127] in the applied magnetic field. Circularly polarized light was used to inject nonequilibrium spin in (Al, Ga)As while an applied magnetic field created equilibrium spin splitting in (In, Ga)As (a particular Al composition in the (Al, Ga)As region can produce nearly zero \( g \)-factor). An interesting implication of this device is that it operates as a spin photodiode [127]. By converting circular polarization directly into an electrical signal it is a counterpart of a spin LED, which converts electrical signal into emission of circularly polarized light. In another approach both the spin injection and the detection were realized electrically. Iron was used as a spin injector into n-doped GaAs, while the spin splitting in p-doped (Ga, Mn)As enabled spin detection [128].

We will revisit the implications of spin-voltaic effect in three-terminal structures, discussed in the section on magnetic bipolar transistors.

In addition to the spin-voltaic effect, the spin-charge coupling in magnetic p–n junctions can produce a giant-magnetoresistance- (GMR-) like effect, which follows from equation (33) [94]. The current depends strongly on the relative orientation of the
nonequilibrium spin and the equilibrium magnetization. Figure 6 plots $j$, which also includes the contribution from holes, as a function of $2q\zeta/k_B T$ for both the unpolarized, $\delta P_n = 0$, and fully polarized, $\delta P_n = 1$, n regions. In the first case $j$ is a symmetric function of $\zeta$, increasing exponentially with increasing $\zeta$ due to the increase in the equilibrium minority carrier density $n_0(\zeta)$. In unipolar systems, where transport is due to the majority carriers, such a modulation of the current is not likely, as the majority carrier density is fixed by the density of dopants. A realization of exponential magnetoresistance was recently demonstrated in a very different materials system of manganite–titanate heterojunctions [129], in which an applied magnetic field affected the width of a depletion layer. If $\delta P_n \neq 0$, the current will depend on the sign of $P_{n0} \cdot \delta P_n$. For parallel nonequilibrium (in the n region) and equilibrium spins (in the p region), most electrons cross the depletion region through the lower barrier (see figure 5), increasing the current. In the opposite case of antiparallel relative orientation, electrons experience a larger barrier and the current is inhibited. This is demonstrated in figure 6 by the strong asymmetry in $j$. The corresponding GMR ratio, the difference between $j$ for parallel and antiparallel orientations, can also be calculated analytically from equation (33) as $2|\delta P_n P_{n0}|/(1 - |\delta P_n P_{n0}|)$ [95].

3. Spin transistors

Thus far we have mostly considered two-terminal spintronic devices in which we were concerned with spin injection and spin-voltaic phenomena. However, the greatest strength of the semiconductor spintronics should lie in the possibility to fabricate three-terminal structures that would allow current gain. Two goals can be set: first, to extend the functionalities of the existing transistors by adding spin control, and, second, to improve the performance of the current technology in terms of speed, power consumption, or sensitivity. Whether or not these goals will be reached depends much on the progress in fabrication and materials development,
as well as on our understanding of the physics of spin-charge coupling in semiconductor heterojunctions.

We will briefly discuss a few proposed and existing spin transistor designs (some of them are further reviewed in [14]), before we analyse in more detail our proposal for magnetic bipolar transistors. The canonical example of a spin transistor is that of Datta and Das [15], depicted in figure 1. The gating does not involve charge build-up, so the transistor could be faster and less power consuming than conventional field effect transistors. In addition, the magnetic configurations of the electrodes can be useful for storing information. However, despite many experimental efforts, the Datta–Das transistor has not been realized. There are inherent difficulties in its semiconductor-based design, the most important being the spin injection to the quasi-one-dimensional conduction channel (which, as noted in section 1, could be avoided by using carbon nanotubes). An interesting alternative to the spin field-effect transistor, a spin MOSFET, has been a more conventional structure employing ferromagnetic source and drain and using the spin-valve effect to control the current [24, 25]. Since the proposed structure includes silicon substrate it could be potentially useful for silicon spintronics. An important prerequisite for a spin MOSFET would be a demonstration of efficient spin injection in Si [97] and there are recent efforts to fabricate suitable ferromagnet/Si contacts [130, 131]. Another effort to incorporate silicon in spintronics devices relies on a spin diffusion transistor with a silicon base [132]. The emitter and collector contacts are ferromagnetic metal–insulator–semiconductor junctions and early encouraging results show both magnetoresistive effects and current gain greater than unity.

There have been other transistor designs using metallic layers in the structure. They go under the name of hot electron spin transistors or spin-valve transistors [14, 133]. For example, one realization (which is also called the magnetic tunnelling transistor) uses a combination of a ferromagnetic tunnel junction and a Schottky barrier collector [134–139]. The tunnel junction plays the role of the emitter–base junction, supplying hot spin-polarized electrons into the magnetic base. The base–collector junction is a Schottky barrier. The hot electrons from the base can overcome the barrier only if their energy is higher than the barrier height. Since the hot electrons lose their energy depending on the spin, the magnetic junction can effectively control the collector current: for parallel magnetizations of the junction the current is large, while for antiparallel it is small, since the spin up, say, hot electrons in a spin down base equilibrate more efficiently than in a spin up base. This is the physics behind the high magnetocurrent ratios (reaching thousands of percent) in these transistors. The disadvantage of these hybrid metal/semiconductor transistors is the absence of gain (the word transistor here points to the three-terminal geometry rather than to the ability to amplify currents). Nevertheless the hybrid designs have been successful in achieving the large magnetocurrent ratios and spin injection into semiconductors. A direct connection of such structures with bipolar transport, discussed in this paper, was recently realized in the spin-valve structures which contain a nonmagnetic p–n junction, which can have useful effects as the energy barrier [140, 141].

Motivated by the potential ease of the integration of magnetic semiconductors with conventional devices, we have proposed what we call magnetic bipolar transistors (MBTs), in which one or more regions (emitter, base, and collector) are formed by a magnetic or ferromagnetic semiconductor [142]. We have shown that such structures can exhibit giant magnetooamplification and a significant control of electrical properties by magnetic field, as well as spin injection all the way from the emitter to the collector. The magnetic bipolar transistor was later discussed in terms of spin currents in [143] and in terms of magnetooamplification in [144]. A bipolar transistor-like scheme has been recently presented in [145]: the semiconductor spin-diffusive channel is topped with three ferromagnetic electrodes, enabling the amplification of magnetoresistance.
4. Magnetic bipolar transistors

4.1. Connection to bipolar junction transistors

We next describe the operation principles of the magnetic bipolar transistor and their conventional (nonmagnetic) counterparts. Consider an npn transistor, as in figure 7(a). The transistor consists of three regions: emitter, base, and collector. There are two p–n junctions in series: base–emitter and base–collector. Depending on the polarity of the bias across the junctions the transistor exhibits different functionalities. The current gain in bipolar junction transistors appear only in the active forward and active reverse regimes. In both regimes one junction is forward, the other reverse biased. The only distinction between the two regimes comes from the asymmetry of the actual device. In the active forward bias the base–emitter junction is forward biased, while the base–collector junction is reverse biased. In the active reverse regime the bias polarities as switched. Typically, the emitter is more heavily doped than the collector, which results in a much greater current gain for the active forward regime due to the increased electron injection efficiency from the emitter (see below).

There are two more possibilities for the transistor operation. In the saturation mode both of the junctions are forward biased. The transistor exhibits no gain, but this configuration is used in logic operations to represent the on state. Similarly, the configuration in which both junctions are reverse biased, also called cut-off, is used to represent the off state.

We will see that those configurations have a much richer structure in magnetic bipolar transistors, shown schematically in figure 7(b). In fact, there exists an additional configuration,
Table 1. Operational modes of bipolar junction transistors (BJTs) and magnetic bipolar transistors (MBTs). Forward (F) and reverse (R) bias means positive and negative voltage, respectively. Symbols MA and GMA stand for magnetoamplification and giant magnetoamplification, while ON and OFF are modes of small and large resistance, respectively; SPSW stands for spin switch. Adapted from [146].

| Mode         | $V_{bc}$  | $V_{bc}$ | BJT  | MBT            |
|--------------|-----------|----------|------|----------------|
| Forward active | F         | R        | Amplification | MA, GMA |
| Reverse active | R         | F        | Amplification | MA, GMA |
| Saturation   | F         | F        | ON    | ON, GMA, SPSW |
| Cut-off      | R         | R        | OFF   | OFF            |
| Spin voltaic | 0         | 0        | OFF   | SPSW           |

in which both junctions are unbiased but spin-charge coupled (through the spin-voltaic effect), which can be used as a spin switch. All the possible functionalities are summarized in table 1.

We have seen in section 2.2 on magnetic diodes that the spin-charge coupling (or, more specifically, the spin-voltaic effect) across magnetic p–n junctions can either intensify or inhibit carrier injection. The electron current in a magnetic junction was given by equation (33). This spin-charge coupling induces significant changes in the operation of the magnetic bipolar transistor in terms of what we have named magnetoamplification and giant magnetoamplification (GMA) [142, 147–149].

Let us first see what is the mechanism behind the current gain in conventional bipolar transistors. The amplification factor $\beta$ is customarily written as [150]

$$\beta = \frac{1}{\alpha_T' + \gamma'}.$$  \hspace{1cm} (34)

Here

$$\alpha_T' = \frac{1 - \alpha_T}{\alpha_T},$$  \hspace{1cm} (35)

$$\gamma' = \frac{1 - \gamma}{\gamma},$$  \hspace{1cm} (36)

and $\alpha_T$ and $\gamma$ are the base transport factor and the emitter injection efficiency, respectively. We can then call $\alpha_T'$ and $\gamma'$ the base transport inefficiency, which is due to the electron–hole recombination in the base, and emitter injection inefficiency, respectively.

The base transport inefficiency is given by the expression

$$\alpha_T' = \cosh \left( \frac{w_b}{L_{ab}} \right) - 1,$$  \hspace{1cm} (37)

which is valid for both conventional and magnetic transistors. Here $w_b$ is the width of the base and $L_{ab}$ is the electron diffusion length. Typically, the width of the base, $w_b$, is much smaller than the electron diffusion length in the base, $L_{ab}$, in which case $\alpha_T' \approx (w_b/L_{ab})^2/2$. Typically $L_{ab}$ is about a micron, so the base transport factor does not play a significant role in modern bipolar transistors with a narrow base. Another possibility for how to reduce the base factor is to employ graded semiconductors with built-in electric fields that help the diffusing electrons from the emitter to reach the collector faster to inhibit electron–hole recombination.

Unlike the base inefficiency, the emitter inefficiency depends rather strongly on the spin properties of the magnetic transistor. Let us write

$$\gamma' = \gamma' \eta,$$  \hspace{1cm} (38)

17
\[ \gamma_0' \] is the emitter inefficiency for conventional spin-unpolarized transistors, while the spin-charge coupling is contained in the factor \( \eta \). We can then write, for the practical limit of \( \alpha_0' \ll \gamma' \), that
\[
\beta \approx \eta / \gamma_0' \approx \eta \beta_0.
\]
(39)

We will be concerned only with \( \eta \); \( \beta_0 \) is the conventional transistor current gain. For us it is just a numerical factor of the order of a hundred.

4.2. Magnetooamplification

In the absence of an external spin injection, all the spin properties of magnetic bipolar transistors are determined by the equilibrium magnetization of the transistor regions. Similarly to magnetic diodes, this equilibrium magnetization influences electrical properties as well, due to the influence on the equilibrium minority carrier density. Consider a magnetic base, for example, assuming its spin polarization to be \( P_{0b} \). The equilibrium electron density there is
\[
n_{0b} = n_i^2 \frac{1}{N_{ab} \sqrt{1 - P_{0b}^2}},
\]
(40)
where \( n_i \) is the intrinsic density of the underlying semiconductor material and \( N_{ab} \) is the acceptor doping. Since the current across p–n junctions, in our case the emitter–base junction, is linearly dependent on \( n_{0b} \), the electron emitter current and the emitter efficiency will be linearly proportional to \( n_{0b} \) as well. The spin-charge factor \( \eta \) then stands out as
\[
\eta = \frac{1}{\sqrt{1 - P_{0b}^2}}.
\]
(41)

This finally gives for the gain of the transistor
\[
\beta = \frac{\beta_0}{\sqrt{1 - P_{0b}^2}}.
\]
(42)
It turns out that the gain can also be controlled by the emitter spin polarization [144], but is unaffected by the possible equilibrium collector spin [148]. The control of the current gain by the equilibrium spin polarization has been termed magnetooamplification [95].

4.3. Giant magnetooamplification

Giant magnetooamplification is a direct consequence of the spin-charge coupling across a p–n junction. Consider a magnetic bipolar transistor with a magnetic base having equilibrium spin polarization \( P_{0b} \). The emitter and the collector are nonmagnetic. Suppose we can excite nonequilibrium spin in the emitter, giving it a nonequilibrium spin polarization \( \delta P_e \). As a result of the proximity of the equilibrium and nonequilibrium spin there appears an EMF across the junction and the modification of the electron injection efficiency. This spin-charge coupling is reflected in the \( \eta \):
\[
\eta = \frac{1 + \delta P_e P_{0b}}{\sqrt{1 - P_{0b}^2}}.
\]
(43)
The current gain factor becomes
\[
\beta = \beta_0 \frac{1 + \delta P_e P_{0b}}{\sqrt{1 - P_{0b}^2}}.
\]
(44)
The spin dependence of the current gain due to the spin-charge coupling has been termed giant magnetoamplification [148], due to the potentially giant relative difference of amplification for parallel \( \delta P_e P_{0b} > 0 \) and antiparallel \( \delta P_e P_{0b} < 0 \) configurations. The corresponding giant magnetoamplification coefficient is

\[
\text{GMA} = \frac{\beta_p - \beta_{ap}}{\beta_{ap}},
\]

where the subscripts p and ap represent the parallel and antiparallel configurations, respectively. For the above case of the magnetic base we obtain (recall also section 2.2 and equation (33))

\[
\text{GMA} = \frac{2|\delta P_e P_{0b}|}{1 - |\delta P_e P_{0b}|}.
\]

If the equilibrium and nonequilibrium spin polarizations were about 50%, the corresponding GMA factor would be about 67%.

4.4. Spin injection

We have seen in the previous section that spin injection from a magnetic n region to a nonmagnetic p region of a p–n junction is not possible at small biases due to balancing thermodynamics of the equilibrium spin polarization and the spin-dependent thermal activation. Only nonequilibrium spin can be injected; that is, the (nonequilibrium) spin has to first accumulate in the magnetic region.

Although the magnetic bipolar transistor comprises two magnetic p–n junctions in series, spin injection is possible even at low biases in the forward active regime. The reason for the possibility of spin injection is that minority electrons injected from the emitter to the magnetic base accumulate in the base. There is thus nonequilibrium electron density with a gradient sufficient to drive the electrons by diffusion to the collector. The electron spins equilibrate in the base to the equilibrium spin polarization. We thus have nonequilibrium spin density in the base, with the equilibrium spin polarization [93]. It is this nonequilibrium spin density that drives the spin injection to the collector: as the spin-polarized electrons move towards the base–collector depletion layer, they are swept by the built-in electric field of the layer to the collector. In effect, we have a minority electron spin pumping, similar to what happens in spin-polarized p–n junctions or solar cells [93]. The resulting spin polarization in the collector, in the limit of a narrow base (the width smaller than the spin diffusion length), is [148]

\[
\delta P_c \approx P_{0b} \frac{L_{sc} \epsilon_0 v_s / k_B T}{w_b N_{dc}},
\]

where \( L_{sc} \) is the spin diffusion length in the collector and \( N_{dc} \) is the collector donor density. For a spin polarization of \( P_{0b} \approx 0.5 \) one can achieve \( \delta P_c \) as large as 0.1, mainly due to the large ratio of the base width and the spin diffusion length in the collector.

Much less surprising is the spin injection possibility for a source spin from the emitter to the collector. Suppose we induce a spin polarization \( \delta P_e \) in the emitter. In the active forward regime the spin is injected to the collector through the following sequence of steps: first, the source spin diffuses towards the base–emitter junction. If the spin diffusion length is larger than the length the spin travels from the injection point to the base–emitter depletion layer, the spin will attenuate only a little. Then the spin is transferred to the base through the depletion region. If we now take a nonmagnetic base, the spin polarization in the base will be roughly \( \delta P_e \). The nonequilibrium spin polarizations on the two sides of a p–n junction are the same, as follows from the generalized Shockley theory of spin-polarized p–n junctions [95]. As a result of the nonequilibrium spin polarization, there appears nonequilibrium spin density in the base.
This density will be injected into the collector as a result of the minority electron spin pumping, similarly to the case of the no source spin case above. We have derived the following formula for the injected spin polarization in the collector:

\[
\delta P_c \approx \delta P_e \frac{L_{sc} n_{be} \sigma V_{be}}{w_b N_{dc}}.
\]

Comparing with equation (47), we see that the role of the source spin polarization in the emitter is similar to the role of the equilibrium spin in the base. Both can be efficiently injected through to the collector.

In summary, magnetic bipolar transistors offer new functionalities to conventional semiconductor electronics. The most exciting is the possibility of large magnetoamplification effects. In all other aspects the magnetic bipolar transistors will have a similar performance to their conventional counterparts, since they are based on the same physical principles governing electronic transport.

5. Conclusions

We have reviewed here several phenomena associated with bipolar spin-polarized transport in semiconductors. Our findings for two-terminal structures, such as magnetic heterojunctions, can also be applied to more complicated multi-terminal geometries. We show that the interplay of the magnetic region with equilibrium spin polarization and injected nonequilibrium spin leads to the spin-voltaic effect in a heterojunction. This theoretical prediction, a spin analogue of the photo-voltaic effect, was also recently confirmed experimentally. The direction of the charge current, which can flow even at no applied bias, can be switched by reversal of the equilibrium magnetization or by reversal of the polarization of the injected spin. In three-terminal magnetic bipolar transistors the spin-voltaic effect implies that one could effectively control the gain or current amplification in such devices. We predict the possibility for giant magnetoamplification, which could be viewed as a generalization of the spin-valve effect to semiconductor structures with strong intrinsic nonlinearities suitable for spin-controlled logic.

Acknowledgments

We thank S Das Sarma, M Fuhrer, B T Jonker, H Munekata, S S P Parkin, A Petrou, A Petukhov, and E I Rashba for useful discussions. This work was supported by the US ONR, NSF-ECCS CAREER, DARPA, NRC-NRL, and SFB 689. This work was performed in part at SUNY’s Buffalo Center for Computational Research and used the resources of the Center for Computational Sciences and the Center for Nanophase Materials Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the US Department of Energy under contract No DE-AC05-00OR22725.

References

[1] Maekawa S (ed) 2006 Concepts in Spin Electronics (Oxford: Oxford University Press)
[2] Maekawa S and Shinjo T (ed) 2002 Spin Dependent Transport in Magnetic Nanostructures (New York: Taylor and Francis)
[3] Parkin S S P, Jiang X, Kaiser C, Panchula A, Roche K and Samant M 2003 Proc. IEEE 91 661
[4] Prinz G 1998 Science 282 1660
[5] Ansermet J-P 1998 J. Phys.: Condens. Matter 10 6027
[6] Hartman U (ed) 2000 Magnetic Multilayers and Giant Magnetoresistance (Berlin: Springer)
[7] Hirota E, Sakakima H and Inomata K 2002 Giant Magnetoresistance Devices (Berlin: Springer)
[8] Gregg J et al 1997 J. Magn. Magn. Mater. 175 1
[9] Johnson M 1994 IEEE Spectr. 31 47
[10] Moodera J S and Mathon G 1999 J. Magn. Magn. Mater. 200 248
[11] Parkin S S P et al 1999 J. Appl. Phys. 85 5828
[12] Tehrani S, Engel B, Slaughter J M, Chen E, DeHerrera M, Durlam M, Naji P, Whig R, Janesky J and Calder J 2000 IEEE Trans. Magn. 36 2752
[13] Johnson M 2001 J. Supercond. 14 273
[14] Žutić I, Fabian J and Das Sarma S 2004 Rev. Mod. Phys. 76 323
[15] Datta S and Das B 1990 Appl. Phys. Lett. 56 665
[16] Bychkov Y A and Rashba E I 1984 J. Phys. C: Solid State Phys. 17 6039
[17] Winkler R 2003 Spin–Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems (Berlin: Springer)
[18] Wang B, Wang J and Guo H 2003 Phys. Rev. B 67 092408
[19] Schieblmann J, Egues J C and Loss D 2003 Phys. Rev. Lett. 90 146801
[20] Matsuyma T, Hu C-M, Grundler D, Meier G and Merkt U 2002 Phys. Rev. B 65 155322
[21] Mirales F and Kirchenow G 2001 Phys. Rev. B 64 024426
[22] Cui C, McGuire J F and Shams L J 2002 Appl. Phys. Lett. 81 4781
[23] Nikonorov D E and Bourianoff G 2005 IEEE Trans. Nanotechnol. 4 206
[24] Sugahara S and Tanaka M 2004 Appl. Phys. Lett. 84 2307
[25] Sugahara S and Tanaka M 2005 J. Appl. Phys. 97 10D505
[26] Schäpers T, Nitta J, Heersche H B and Takayanagi T 2001 Phys. Rev. B 64 125314
[27] Sahoo S, Kontos T, Furer J, Cheong H D, Kioseoglou G and Petrou A 2000 Phys. Rev. B 62 8180
[28] Young D K, Johnston-Halperin E, Awschalom D D, Ohno Y and Ohno H 2000 Appl. Phys. Lett. 80 1598
[29] Hanbicki A, van ’t Erve O M J, Magnon R, Kioseoglou G, Li C H, Jonker B T, Itskos G, Mallory R, Yasar M and Petrou A 2003 Appl. Phys. Lett. 82 4092
[30] Johnson M 1993 Science 260 320
[31] Matsuyma T, Takahashi S and Imamura H 2002 Spin Dependent Transport in Magnetic Nanostructures ed S M Macka and T Shino (New York: Taylor and Francis) pp 143–236
[32] Shockley W 1950 Electrons and Holes in Semiconductors (Princeton, NJ: van Nostrand-Reinhold) 84
[33] Sze S M 1981 Physics of Semiconductor Devices (Princeton, NJ: van Nostrand-Reinhold)
[34] Fiederling R, Kleim M, Reuscher G, Ossau W, Schmidt G, Waag A and Molenkamp L W 1999 Nature 402 787
[35] Jonker B T, Park Y D, Bennett B R, Cheong H D, Kioseoglou G and Petrou A 2000 Phys. Rev. B 62 8180
[36] Van Roy W, Van Dorpe P, De Boeck J, and Loss D 2003 Phys. Rev. Lett. 90 156603
[37] Van Roy W, Van Dorpe P, De Boeck J and Borghs G 2006 Mater. Sci. Eng. B 126 155
[38] Jančík E and Karčzewski G 1988 Acta Phys. Pol. A 73 439
[39] Munekata H, Ohno H, von Molnár S, Segmüller A, Chang L L and Esaki L 1989 Phys. Rev. Lett. 63 1849
[40] Ohno H, Munekata H, Penney T, van Molnár S and Chang L L 1992 Phys. Rev. Lett. 68 2664
[41] Munekata H, Ohno H, Ruj R R, Gambino R J and Chang L L 1991 J. Cryst. Growth 111 1011
[42] Ohno H, Shen A, Matsukura F, Oiwa A, End A, Katsumoto S and Iye Y 1996 Appl. Phys. Lett. 69 363
[43] Van Esch A, Van Bockstal L, Van Bogaerts R, Herlach F and Borghs G 1999 Phys. Rev. B 59 13103
[44] Hayashi T, Tanaka M, Nishina T, Shimada H, Tsuchiya T and Ota K 1997 J. Cryst. Growth 175/176 1063
[45] Ohno H 1998 Science 281 951
[46] Dietl T 2002 Semicond. Sci. Technol. 17 377
[47] Jungwirth T, Sinova J, Masek J and MacDonald A H 2006 Preprint cond-mat/0603080
[48] Ohno H, Arata I, Matsukura F, Ohtani K, Wang S and Ohno H 2000 Appl. Surf. Sci. 159/160 308
[49] Kohda M, Ohno Y, Takamura K, Matsukura F and Ohno H 2001 Japan. J. Appl. Phys. 40 L1274
[50] Johnston-Halperin E, Lofgreen D, Kawakami R K, Young D K, Coldren L, Gossard A C and Awschalom D D 2002 Phys. Rev. B 65 041306
[51] Arata I, Ohno Y, Matsukura F and Ohno H 2001 Physica E 10 288
[52] Van Dorpe P, Liu Z, Roy W V, Motsnyi V F, Sawicki M, Borghs G and De Boeck J 2004 Appl. Phys. Lett. 84 3495
[53] Tsai F, Ma L and He L 2003 Appl. Phys. Lett. 83 954
[54] Dietl T 2003 Nat. Mater. 2 646
[55] Samarth N, Chiu S H, Ka K C, Potashnik S J and Schiffer P 2003 Solid State Commun. 127 173
[56] Park Y D, Hanbicki A T, Erwin S C, Hellberg C S, Sullivan J M, Mattson J E, Ambrose T F, Wilson A, Spanos G and Jonker B T 2002 Science 295 651

