Nanotechnology **7** (1996) 390–396. Printed in the UK

The bipolar spin transistor

Mark Johnson

Naval Research Laboratory, Washington, DC 20375, USA

Abstract. A novel three-terminal switching device, fabricated only from metals, has recently been demonstrated. Somewhat similar to a semiconductor transistor, the physical principles of operation are quite different. It is an active device driven by a thermodynamic force associated with the effective Zeeman energy of the spin polarized electrons in the base. 'Bipolar' has a double meaning: there are two polarities of carriers, upspin and downspin electrons; and the output can be a positive or negative voltage (or current). The three-terminal device does not have power gain, but because it shows a memory effect it is natural to use it as a storage element in a nonvolatile memory array. Power gain can be achieved in a five-terminal embodiment, allowing fanout and the linking of devices in logic operations. Switching times faster than 1 ns can be expected. Because all the layers are metals and charge carrier densities are high, fabrication of the device at submicron scales is quite possible and operation at the nanometer scale is conceivable. Thus, the possibility of achieving high packing densities is very plausible, and is not obviated by problems common to semiconductors such as heat dissipation and low carrier density.

Introduction

The bipolar spin transistor [1–3] is an active switching device fabricated using only metals. Still in an early stage of development, prototypes made on a size scale of 10–100 microns have been demonstrated. Because metals have a high density of charge carriers, it is feasible to contemplate the construction of nanoscale devices. The purpose of this article is to introduce this novel device, explaining what it is, how it works, and how it might be used in applications. Following this goal, section 1 presents a review of the phenomenology of spin polarized transport. Section 2 gives a detailed description of device operation. In section 3, the results of a prototype are presented and explained. Section 4 discusses possible applications, and section 5 reviews some of the issues involved in nanofabrication of spin transport devices.

1. Review of spin polarized transport

We begin with a brief review of conduction electron spin transport in ferromagnetic and ferromagnetic–nonmagnetic (F–N) metal systems. Figure 1(a) presents a density-of-states diagram for a simplified band picture of the 3d band of a single domain transition metal ferromagnet, like Fe, Ni, or Co. A more accurate model would include a 4s band, but realistically these two bands are hybridized, and it is appropriate to consider only the 3d band. The two spin subbands are shifted by their exchange energy, of order 1 eV, and in the idealization of figure 1(a) one of the spin subbands lies entirely below the fermi level, E_F. Since current transport involves only the conduction electrons within a thermal energy range, $E_F \pm K_B T$, with K_B Boltzmann's constant and T the temperature, it is obvious from figure 1(a) that an electric current will be carried by electrons from a single spin subband. Equivalently, this implies that the current within F is spin polarized.

Although it has been known for some time that electric current in a ferromagnet is spin polarized, it was the seminal work of Meservey *et al* [4] which showed that current tunneling out of a ferromagnetic electrode remained spin polarized as it left the ferromagnet. Aronov [5] then proposed the generalization that any transport current I_e entering a nonmagnetic metal or semiconductor from a ferromagnetic metal would be polarized. This was a nontrivial hypothesis; the prevailing opinion was that any effects of magnetic ordering in F would undergo evanescent decay in N and vanish within a few nanometers of the interface. The *spin injection* experiment [6] was the first empirical demonstration of Aronov's idea, and this is described schematically in figure 1(b). When a F–N bilayer is biased with a voltage, only one spin subband is available to carry the current across the interface into N. Conceptually, the electric current is also a current of oriented magnetic dipoles, each dipole having a magnetic moment β, where β is the Bohr magneton. Therefore, associated with I_e is a current of magnetization, $I_M = \beta I_e/e$, where I_e/e is seen to be the number current of carriers. Once in N, the spin polarized conduction electrons diffuse away from the interface and eventually lose their spin orientation by random scattering events. The characteristic spin diffusion length is defined as $\delta_s = \sqrt{DT_1}$, with D the electron diffusion constant and T_1 the spin relaxation time. Near the F–N interface, within a distance $d < \delta_s$, the steady state current flow results in a net population of spin polarized carriers. This is known as a nonequilibrium magnetization, $\tilde{M}$, and is depicted in

0957-4484/96/040390+07$19.50

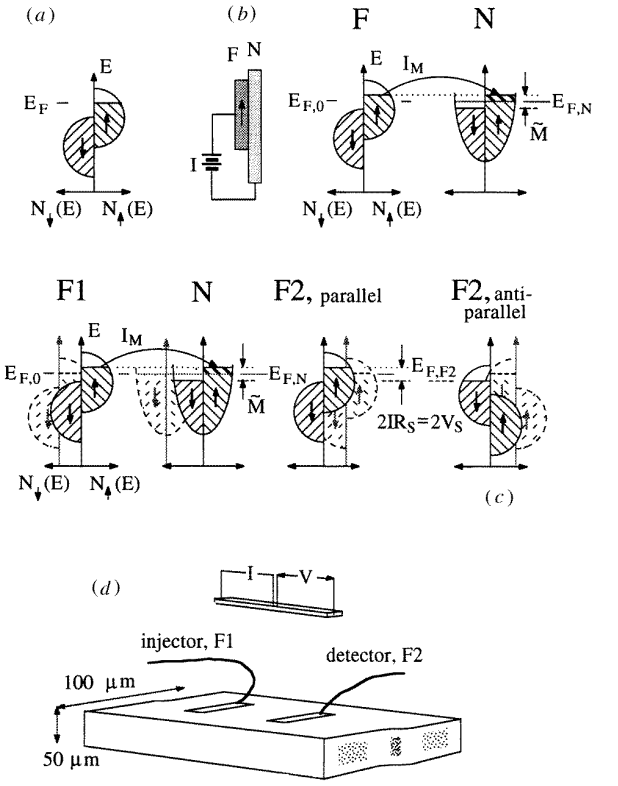

Figure 1. (*a*) Density-of-states diagram for a simple band model of a transition metal ferromagnet. (*b*) Spin injection results in a nonequilibrium magnetization in a nonmagnetic metal N, when current is driven into it from a single domain ferromagnet. The diagrams neglect any contact resistance at the F–N interface but correctly show the effect of charge-spin coupling, that the chemical potential of F rises above its equilibrium value, $E_{F,0}$, in order to align with the upspin subband chemical potential of N. (*c*) Density-of-state diagrams for an injector/detector, four-probe F–N–F structure, such as appears in (*d*). The dashed diagrams in the background show the equilibrium condition, when no current flows. The solid diagrams in the foreground refer to the dynamic condition in the presence of steady state current flow. (*d*) The geometry of the original four-probe spin injection experiment. Polyimide coats the surface of a bulk Al strip, small windows (about 10 μm wide) are etched through the polyimide, and $F1$ and $F2$ are deposited over the windows, in contact with the Al.

figure 1(*b*) as the number difference of spin-up and spin-down carriers. The occupation of the spin-down subband has been depleted because of current conservation and the stringency of charge neutrality. In other words, for every two unpolarized electrons that leave N and return to the current source (equivalently, one spin-up and one spin-down), there are two spin-up electrons, injected in the vicinity of the F–N interface, to replace them. This results, in the steady state and in the vicinity of the interface, in a net increase of spin-up electrons and a net decrease of spin-down electrons.

The converse of the spin injection process is of equal importance: the presence of spin polarized conduction electrons (equivalently, the nonequilibrium magnetization $\tilde{M}$) in N has an effect on F. Under closed circuit conditions, $\tilde{M}$ can generate electric currents, and under open circuit conditions it can alter the chemical potential of F. Silsbee [7] predicted this converse effect, called *charge-spin coupling*, and the spin injection experiment provided the first empirical proof of this as well. The concept of charge-spin coupling can be understood with the aid of the solid diagrams in figure 1(*c*). The experiment used a four-probe, injection/detection scheme with a ferromagnetic current injector, $F1$, and a ferromagnetic detector, $F2$, fabricated on bulk samples as depicted in the sketch of figure 1(*d*). The injecting electrode $F1$ pumps up a nonequilibrium magnetization $\tilde{M}$ in N, and the spins diffuse over a region of approximately δ_s on either side of the injector. When the detecting electrode $F2$ is short-circuited to ground (low impedance current detection, not depicted in figure 1(*c*)), and when the magnetization of single domain $F2$ is parallel with that of $F1$, then a current of electrons is driven through $F2$ to the ground because the up-spin subband chemical potential in N, $E_{F,N,\uparrow}$, is larger than that of $F2$, $E_{F,F2,\uparrow} = E_{F,F2}$. When the magnetization of $F2$ is antiparallel with that of $F1$, then a current of electrons is driven from $F2$ into N, because the down-spin subband chemical potential in N, $E_{F,N,\downarrow}$, is lower than that of $F2$, $E_{F,F2\downarrow} = E_{F,F2}$. The diagrams in figure 1(*c*) are drawn for the case when the detecting electrode $F2$ is a high impedance voltmeter. When the magnetization orientations of $F1$ and $F2$ are parallel, the chemical potential of the up-spin subband of $F2$ rises to align with that of N, by an amount V_s. When the magnetization orientations are antiparallel the chemical potential of $F2$ falls by an equal amount.

The results of [6] confirmed these ideas, recording positive or negative currents (or voltages) when the magnetization orientations were parallel or antiparallel, respectively. By measuring V_s as a function of distance d, the spin depth in these bulk Al samples was determined directly. It is important to note that charge-spin coupling is a bipolar effect. A positive current is pushed out through the detector (or a positive voltage is developed) when the magnetization orientations are parallel. A negative current is pulled in through the detector (or a negative voltage is developed) when the orientations are antiparallel.

2. Spin transistor, principles of operation

In the bulk samples of [6] the magnitude of the voltages V_s was small. The spin transistor [1–3] represents the application of these same ideas to a thin film geometry, where the spin coupled voltages are much larger. Figure 2(*a*) depicts a cross-sectional view of a pedagogical, three-terminal, thin film device. Nonmagnetic metal film N acts as the base, and is sandwiched between two ferromagnetic films, the emitter, $F1$, and the collector, $F2$. Each ferromagnetic film is a single domain, and the axis of magnetization of each film, $\hat{M}_e$ and $\hat{M}_c$, lies in the plane of the film. For simplicity we consider the case where $\hat{M}_e$ points down and $\hat{M}_c$ points either down or up. The device is current biased by a battery, and a third wire is attached to $F2$ and leads to a *gedanken* voltmeter, one which can read the voltage V_{F2} of $F2$ with a single input and does not need a ground. We can describe the physics

of device operation using the density of state diagrams and the microscopic model of figure 1(*c*). Here the faint diagrams in the background represent the state of the system when the switch is open, and the solid diagrams in the foreground represent the system when it is biased by the current. The following is somewhat redundant with the previous description, but is a more detailed discussion.

The band structure of the ferromagnetic emitter is such that an electric current in this metal is carried primarily by electrons of one spin subband. Thus, the electric current is a spin polarized current; it is a current of oriented magnetic dipoles and there is an associated current of magnetization, $I_M = \beta I_e/e$. We can think of the emitter as a reservoir of particles (conduction electrons) that carry charge and magnetic moment, and the band structure of the ferromagnet plus the law of detailed balance require this rule: the emitter can exchange particles with the base (the nonmagnetic layer) only if those particles are (for this arbitrary convection) spin-up electrons. Thus, the electric current carried into the base is also spin polarized. Within the base, we have steady state current flow, and current conservation requires that electrons be drawn out from the bottom of the base at the same rate that other electrons are added at the emitter–base interface. Near that interface there is a surplus of spin-up electrons. All the conduction electrons move diffusively, and each electron suffers a collision (with an impurity, a surface, or a phonon), on average, every τ sec. Each spin polarized electron has a small probability that any given collision will randomize its spin orientation, so that the spin alignment lasts for an average time $T_1 > \tau$. During this time the spin polarized electrons diffuse a distance of a spin depth, $\delta_s = \sqrt{DT_1}$. If the thickness of the base layer is much thinner than δ_s, then the spin polarized electrons cannot escape, and if we add new spin polarized electrons at a rate I_M that is faster than the rate the trapped spins are relaxing, $1/T_1$, then we will pump up a population of spin polarized electrons, $\tilde{M} = I_M T_1/\text{volume}$. $\tilde{M}$ is called a *nonequilibrium magnetization*, because it will decay back to the equilibrium state (of an equal number of spin-up and -down electrons) if we remove the driving source I_M. This is a highly ordered state, with a low entropy, and there is an energy cost to maintain this state. The spin polarized electrons generate an effective magnetic field, $\tilde{M}/\chi$, where χ is the Pauli susceptibility of the electrons, and each spin polarized electron has an interaction energy with this field, $\beta\tilde{M}/\chi$, known as an effective Zeeman energy. This is the extra energy associated with the highly ordered state. Thermodynamics demands that the system will act to lower its energy and increase its entropy when provided with an opportunity.

The ferromagnetic collector film is also a reservoir of particles. When its magnetization is parallel with that of the emitter, it can exchange particles with the base if and only if those particles are spin-up electrons. The spin polarization in the base is a *nonequilibrium* condition, having an excess energy $\beta\tilde{M}/\chi$. The spin polarization in the ferromagnetic collector is an *equilibrium* condition, and there is no extra energy associated with it. Thus, crossing the interface from the base to the collector there is a gradient of energy. A gradient of energy is a force, and this is the thermodynamic force, F_M, that pushes electrons into the collector and generates a clockwise current of electrons in the detector portion of the circuit. An ammeter measures this current; it is insensitive to magnetic moment and measures only charge flow. Conceptually, the key point is that we have generated an electric current simply by using properties of magnetization and spin. This is an active device, driven by a thermodynamic force. If we were to block the current flow, thus imposing open circuit conditions, then the reservoir of particles in the collector would adjust to align its electrochemical potential with that of spin-up electrons in the base. In other words, a high impedance voltmeter will measure a voltage V_s proportional to the (positive) Zeeman energy, $V_s = \beta\tilde{M}/e\chi$.

When the magnetization of the collector is antiparallel with that of the emitter, then the collector is a reservoir of spin-down electrons, so now it can exchange particles with the base if and only if those particles are spin-down electrons. The gradient of energy, as we cross the interface between base and collector, now has the opposite sign from the previous case because $\tilde{M}$ is negative with respect to the equilibrium magnetization in the collector. The thermodynamic force F_M has the opposite sign from the previous case, and acts to pull (spin-down) electrons into the base from the collector, generating a counterclockwise current of electrons in the detector portion of the circuit. The detector measures the same magnitude of current as before, but with the opposite sign. If we have open circuit conditions, a high impedance voltmeter will measure a voltage V_s proportional to the (negative) Zeeman energy, $V_s = \beta\tilde{M}/e\chi$, where β now has the opposite sign (for spin-down electrons).

From this simple picture, we can understand some of the attributes of the spin transistor. (i) Because it is fabricated from metals it is necessarily a low impedance device; (ii) it is biased by a current, and the bipolar output can be either a current or a voltage; (iii) the magnetization $\tilde{M}$, which is the origin of the driving force F_M, is linearly proportional to the bias current so that the output characteristics can be described in units of impedance, and the term 'spin coupled impedance' Z_s is used. In the limit $d < \delta_s$, Z_s can be expressed as

$$Z_s = \frac{V_s}{I_e} = \frac{\eta_1\eta_2}{e^2}\frac{T_1 E_F}{1.5nAd} = \eta_1\eta_2\frac{\rho\delta_s^2}{Ad} \qquad (1)$$

where η_i are phenomenological parameters that describe the efficiency of spin transport across the interface, n is the density of charge carriers, A is the area of the injecting interface, and where the second form has used an Einstein relation for the resistivity ρ of N, $\rho = 1/e^2DN(E_F)$. (iv) The prototypes have shown no nonlinearities in their I–V characteristics. This contrasts the behaviour of semiconductor p–n junctions, and implies that the basic, three-terminal embodiment of a spin transistor will not show power gain. (v) The dimensions of magnetization are dipole moments per unit volume. This implies that for a constant bias current the device characteristics (e.g. Z_s) can be improved by shrinking the device volume. Note that semiconductor devices can shrink in size until their

performance is limited by dominance of surface states. There is no analog of surface states in metals, so device dimensions can shrink so long as current densities do not exceed $\sim 10^8$ amp cm^{-2}.

It may be useful to think of the spin transistor as a device whose transimpedance g is proportional to the projection of the magnetization $\hat{M}_c$ on $\hat{M}_e$, $g \propto Z_s \cos\theta$. Thus, the transimpedance can be set, stored and reset by setting the angle θ. Because hysteresis is one of the useful properties of ferromagnetic films, the magnetizations $\hat{M}_i$ exhibit a memory effect, and we will see below that one of the natural applications of the device is that of a memory element.

3. Demonstration of a prototype device

The pedagogical device introduced in figure 2(a) is nearly the same as the prototype devices that have been fabricated. The geometry of the latter is depicted schematically in figure 2(b). Instead of grounding the third terminal (collector) to the bottom of the nonmagnetic film (the base), symmetry is used to minimize the impedance of the detector circuit by using, as a ground, a nonmagnetic film N' placed alongside the collector. Even in the presence of nonequilibrium magnetization in the base, the chemical potential of N' always aligns with the average chemical potential of the base, whereas the chemical potential of the collector will rise or fall depending on whether $\hat{M}_c$ is parallel or antiparallel with $\hat{M}_e$. This is a convenient geometry for application as a memory element, and also for demonstrating the bipolar nature of the effect.

Details of prototype fabrication are depicted in a top view in figure 2(c) and in an expanded view in figure 2(d). The bottom layer, $F1$, is a 65 nm thick permalloy film, e-beam evaporated from a $Ni_{0.8}Fe_{0.2}$ charge and deposited over a 30 nm thick Ta film used to improve adhesion to the sapphire substrate and to promote isotropic current flow in the region of the emitter. A 90 nm thick insulating film of Al_2O_3 is deposited over the permalloy, and square windows are opened through the oxide by photolithography and liftoff. The dimensions of the windows vary from 50 μm to 300 μm, with most prototypes having windows of 100 μm transverse dimension. The base is a gold film of variable thickness d, typically 100 nm or more, which is thermally deposited, and makes contact to the permalloy through the windows. A second insulating layer of Al_2O_3 is deposited with its windows in registration over those of the first layer. After outgassing the gold, a second permalloy film, 70 nm thick, is deposited over one of the windows, capped by a 20 nm thick gold film to prevent oxidation. Indium wires are pressed onto the tops of the two windows and lead to the detection circuit.

A demonstration of prototype device operation, in voltage mode and using an external magnetic field H to manipulate the orientations of the magnetizations of the emitter and collector, is presented in figure 3(a). As a function of H, the magnetization of each film follows a hysteresis loop such as one of those sketched in figure 3(b). For fields applied along a chosen axis (e.g. the $\hat{z}$-axis, in the plane of the films), the magnetization of the emitter film, for

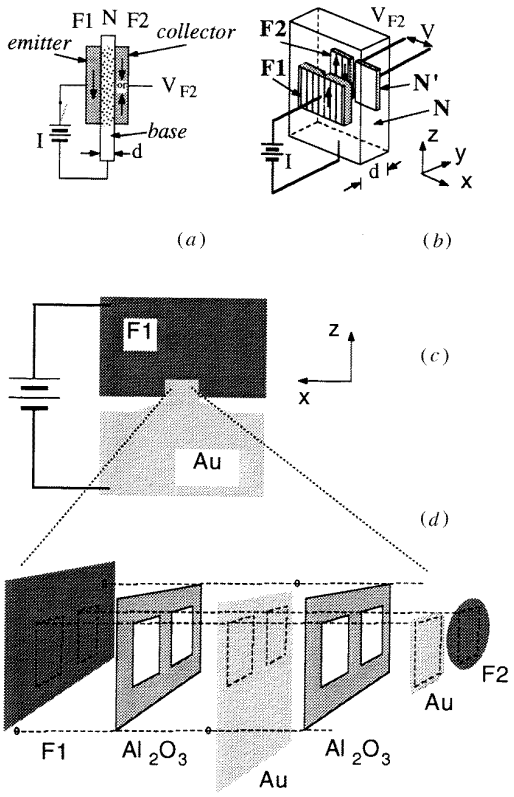

Figure 2. (a) Pedagogical model of three-terminal device. (b) The geometry used to measure V_{F2} for the data set of figure 3. N is depicted transparent. (c) Details of prototype fabrication in top view and (d) expanded view.

example, follows the dotted trace. For a large field applied along $-\hat{z}$, the magnetization is uniform (a single domain) and points along the $-\hat{z}$-direction. The magnetization maintains this state as the field is diminished to zero, and only when the field is increased along the $+\hat{z}$-direction and reaches a value known as the coercive field H_c does the magnetization reorient. When H is increased above H_c the magnetization becomes uniform again, pointing now along the $+\hat{z}$-direction. As H is reduced to zero (right to left along the top portion of the hysteresis loop in figure 3(b)), the magnetization retains its orientation along $+\hat{z}$ until H is increased along $-\hat{z}$ and the coercive value $-|H_c|$ is reached, causing the magnetization to reorient and point along $-\hat{z}$. Square hysteresis loops similar to those of figure 3(b) are characteristic of single domain films that change their orientation over a narrow field range ΔH. The coercivity of a film is determined by intrinsic material properties, substrate interactions, and shape anisotropies. Operation of the spin transistor depends on the ability to manipulate the magnetizations $\hat{M}_i$ individually, and it becomes crucial to use emitter and collector films that have different coercivities.

For the data in figure 3(a), a constant bias current is supplied and the voltage from collector to ground (refer to figure 2(b)) is recorded as the magnitude and direction (negative or positive) of an external field is changed. The solid line is a trace for which the field was swept from an initial negative value to a final positive value, and the dip cor-

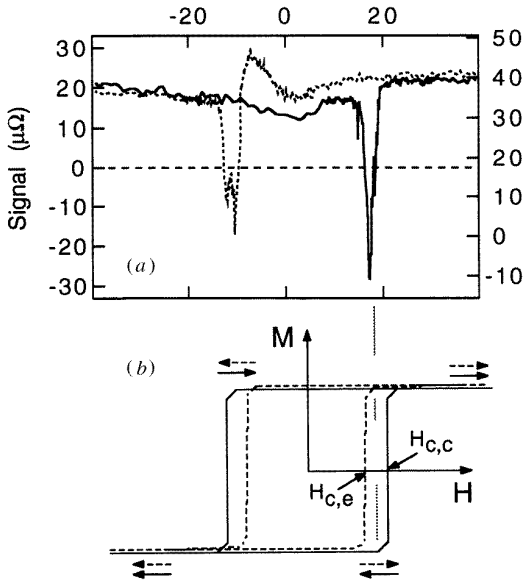

Figure 3. (*a*) Data from a sample with d = 98 nm, showing hysteretic operation. I = 0.2 mA; solid line: sweeping up in field; dotted line: sweeping down. (*b*) Hysteresis loops for the emitter (dashed) and collector (solid). Arrows represent parallel and antiparallel orientations.

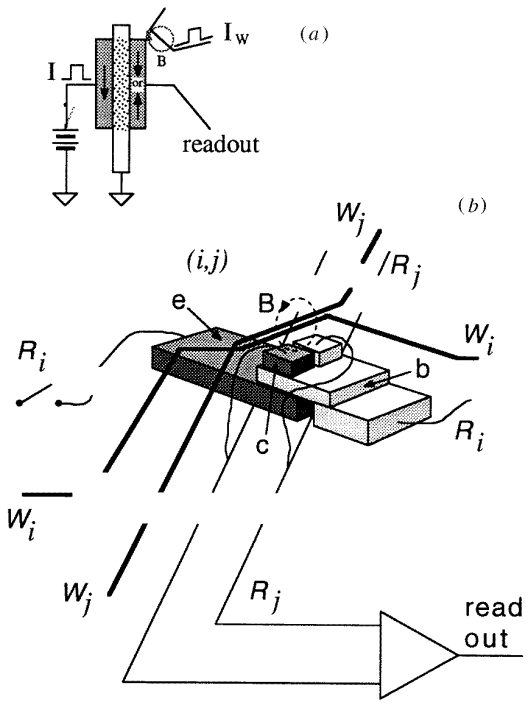

Figure 4. (*a*) Modifications necessary for device application include the addition of a write wire fabricated near the collector. The field B from the write pulse is just larger than the coercivity $H_{c,c}$. (*b*) A single element in an array of elements forming a nonvolatile random access memory (NRAM). The emitter, collector, and base are denoted by e, c and b respectively.

responds to the field value midway between $H_{c,e}$ and $H_{c,c}$ in figure 3(*b*). The recorded voltage is positive for fields $H < H_c$ because $\hat{M}_c$ and $\hat{M}_e$ are aligned parallel, along $-\hat{z}$. When H is slightly larger than $H_{c,e}$, but smaller than $H_{c,c}$, the voltage becomes negative because $\hat{M}_c$ and $\hat{M}_e$ are now aligned antiparallel. When H is increased above $H_{c,c}$ the original, positive voltage is recovered because $\hat{M}_c$ and $\hat{M}_e$ are aligned parallel again, now along $+\hat{z}$. The dotted line is for a reverse trace; the field is swept from an initial positive value to a final negative one, and the same voltages are recorded but are shifted, by hysteresis, to the opposite side of $H = 0$. In this sample, a small voltage offset has been introduced by an accidental asymmetry in the lithography. Note that the detected voltage is bipolar despite this offset (about 17 $\mu\Omega$, refer to right axis of figure 3(*a*)). The data in figure 3(*a*) were taken at cryogenic temperatures, typically in the range 4 K $< T <$ 90 K, because a superconducting magnet was employed. Recent measurements have extended the range above 150 K, and there is no physical reason to prevent room temperature operation of the device.

4. Applications

While these data demonstrate how a prototype functions, device application would incorporate a couple of simple, but important, changes. Rather than having to manipulate $\hat{M}_i$ of both ferromagnetic films, the magnetization of one would be 'frozen' (e.g. the emitter would be chosen to have a high coercivity so that its magnetization is never altered after an initial setting) and the magnetization of the other is manipulated with an applied field. The applied field would be provided by a nearby wire carrying a small current, just large enough that the magnetic field locally generated by this current is larger than $H_{c,c}$ (refer to figure 4(*a*)). Since this current sets the state of the device it can be called a write current, I_W. Thus, the device configuration of a spin transistor requires a separate write wire to control the magnetization $\hat{M}_c$, and the result is a five-terminal architecture.

Figure 4(*a*) schematically depicts a spin transistor as a single memory element. More generally, one could fabricate an array of elements (figure 4(*b*)) as a nonvolatile random access memory (NRAM). The magnetization orientation of all the emitter films is initially set by application (and subsequent removal) of a large magnetic field. The magnetization of the collector films is manipulated by sending currents through an overlayed array of write wires. Any element (i, j) is addressed by sending a pulse of write current through two write wires (W_i, W_j). By choosing the current amplitude to be $I_w/2$, only the magnetization at the collector of element (i, j) is affected. A logical '0' or '1' could correspond to parallel and antiparallel orientations of $\hat{M}_c$ and $\hat{M}_e$, respectively. Because of hysteresis, the magnetization states corresponding to '0' or '1' are stored indefinitely, even though no power is supplied. Reading is accomplished by sending a pulse of current (a read pulse) down the ith read line R_i. A plus or minus voltage pulse is generated and sent down line R_j, where it is 'read'. This kind of memory has all the advantages of DRAM: fast read, write and access times; no moving parts (integrated); and infinite read and write cycles. It has the additional advantage that the stored information is nonvolatile; it can be stored forever without drawing any power.

One can generalize these ideas and discuss other possible applications of the device. When the write current is a sequence of write pulses, i.e. a modulated input signal, and the bias current is constant, then the bipolar current through the collector circuit, the output signal, is modulated with the same pattern as the write current. With a proper choice of geometry, it is possible that the amplitude of the write current can be smaller than the amplitude of the output current, and thus the device would act as a current amplifier. The orientation of the magnetization of the collector can be changed at frequencies of order GHz, so this is a broad bandwidth application.

A trivial generalization of the device depicted in figure 4(*a*) incorporates two separate inputs in the write circuit. Choosing a collector with an appropriate coercivity permits the device to function as an 'and' or 'or' gate. For example, if the coercivity is chosen to equal the stray field from a single write pulse, then the device acts as an 'or' gate. Choosing the coercivity to be twice this value results in operation as an 'and' gate.

Since it is conceptually possible to use the spin transistor as a memory element, an amplifier, and a logic element, it becomes feasible to make a computer using only metal films, and no semiconductors at all. The desirability of this becomes apparent when one contemplates fabricating devices on very small spatial scales. Because the device characteristics of the spin transistor scale favorably with decreasing size, and because sub-micron lithographic fabrication techniques can be readily used with metal films, it is possible to contemplate a spin transistor fabricated within a cell sized 100 nm on a side (or, possibly, even smaller). The corresponding packing density is about 100 times greater than state of the art Si devices, and 10–100 times better than GaAs FETs. The power per 'gate' can be estimated, using bias currents of order 0.1 mA (for reading or writing) and a line impedance of 50 Ω, at 0.5 μW. As mentioned above, the switching times should be of order 1 ns. By comparison, Si-based devices have switching times of order 0.1–1 ns, corresponding to a power per gate of 10 mW–20 μW. GaAs FET devices are a bit faster, 0.02–0.1 ns, but dissipate more power per gate, 5–0.1 mW.

The spin transistor can be expected to switch a little slower than semiconductor devices, but should have greater packing densities and dissipate less power. Compared with Si, the factors for power and size are about the same, so that a Si chip should dissipate the same power as the same sized spin transistor chip, although the latter would contain 100 times more devices. When used as an element in a NRAM, the duty cycle would be quite low so that power dissipated would be minimized. If used as a processor, the power requirements increase. However, because the spin transistor uses metal films, the thermal conductivity is quite large, and it would be relatively easy to conduct away the dissipated power. Furthermore, the charge carrier density of semiconductors is a sensitive function of temperature, and this places restrictions on packing densities and cooling requirements. Metals suffer no such drawbacks; the carrier concentration is insensitive to temperature. Thus, venting away dissipated power is less crucial.

5. Nanofabrication issues

Looking further into the future, one can contemplate the prospects for building devices on a scale of 10 nm. Thus, consider a device fabricated from 'lines' and junctions, the active portion of which is a volume 10 nm on a side, 10^{-18} cm^3. Highly doped semiconductors have carrier densities of the order 10^{18}–10^{19} cm^{-3}, and the active volume of a semiconductor device will contain one to ten charge carriers. Carrier densities in metals, on the other hand, are of order 10^{22}–10^{23} cm^{-3}, so a spin transistor of the same dimensions will have 10^4–10^5 charge (and spin) carriers. Clearly they offer some advantage when attempting to fabricate devices on such a small scale.

Nanofabrication of devices involving ferromagnetic materials will encounter challenges presented by micromagnetics. The emitter and collector must be single domains and must have low coercivities. Although it is not yet known how microfabrication techniques affect domain structure, there is evidence that patterned edges tend to pin domain walls, increasing the coercivity and inhibiting the formation of single domains. Structures with stable domain patterns, e.g. those that incorporate flux return paths, will have to be engineered. Furthermore, these are low impedance devices with interdevice coupling achieved inductively. It is likely that hybrid devices will be needed for any high impedance interface with semiconductor electronics.

Some of these issues can be made concrete by use of an example. If the inverse volume scaling were to remain valid for five decades, then the value $Z_s = 1$ mΩ observed on a prototype with $A = (100\ \mu\text{m})^2$ would become $Z_s = 100$ Ω for a cubic device with edge length 100 nm. If linked to its neighboring device by a wire with cross-section $(30\text{ nm})^2$, then it must drive a resistance of a few Ω. Operated as a switch, one can think of switching 100 Ω into one or the other of two circuit arms, each of which has an intrinsic resistance of a few Ω. But the successful operation of this device depends on the fabrication of a stable domain in a ferromagnetic film with a dimension of 100 nm. The spin coupled impedance Z_s diminishes if the ferromagnetic films break into multiple domains, and the bias currents must be increased if the coercivity of the collector is raised by domain wall pinning.

In conclusion, while there is a long way to go from the demonstration of crude prototypes with dimensions of 100 microns to the realization of devices on a nanometer scale, two points can be emphasized: metals can supply an adequate number of charge carriers to readily permit transport, and a physical mechanism (the spin transistor mechanism) has been demonstrated to be capable of driving an active device fabricated from metals.

Acknowledgment

The author wishes to acknowledge the partial support of the Office of Naval Research, grant no N00014-95-1-0430.

References

[1] Johnson M 1993 *Phys. Rev. Lett.* **70** 2142
[2] Johnson M 1993 *Science* **260** 320
[3] Johnson M 1994 *IEEE Spectrum* **31** no 5 47
[4] Tedrow P M and Meservey R 1971 *Phys. Rev. Lett.* **26** 192
Tedrow P M and Meservey R 1973 *Phys. Rev.* B **7**
Meservey *et al* 1976 *Phys. Rev. Lett.* **37** 858
[5] Aronov A G 1976 *Pis'ma Zh. Eksp. Teor. Fiz.* **24** 37
Aronov A G 1976 *Sov. Phys. JETP Lett.* **24** 32
[6] Johnson M and Silsbee R H 1985 *Phys. Rev. Lett.* **55** 1790
Johnson M and Silsbee R H 1988 *Phys. Rev.* B **37** 5312
Johnson M and Silsbee R H 1988 *Phys. Rev.* B **37** 5326
Johnson M and Silsbee R H 1988 *J. Appl. Phys.* **63** 3934
Johnson M and Silsbee R H 1987 *Phys. Rev.* B **35** 4959
[7] Silsbee R H 1980 *Bull. Magn. Res.* **2** 284