Modeling Spin-Based Electronic Devices


Abstract - The breathtaking increase in performance of integrated circuits is supported by the continuous miniaturization of CMOS devices. However, growing technological challenges and soaring costs are gradually bringing scaling to an end, and research on alternative technologies and computational principles becomes important. Spin attracts attention as alternative to the charge degree of freedom for computations and non-volatile memory applications. Recent progress and challenges in simulating spin-based devices are briefly reviewed. Special attention is paid to methods for enhancing the electron spin lifetime in silicon inversion layers and thin films necessary for realizing spin-based field-effect transistors. A new rapidly switching structure for spin-based non-volatile memory cell, MRAM, was optimized by means of micromagnetic simulations. Several memory cells can be assembled in logic gates capable to perform logic operations. We demonstrate that an implication gate is promising for realizing intrinsic non-volatile logic-in-memory architectures.

I. INTRODUCTION

Continuous miniaturization of CMOS devices is the driving force behind the tremendous increase in performance, speed, and density of modern integrated circuits. Numerous outstanding technological challenges have been resolved during this exciting journey. Among the most crucial technological changes recently adopted by the semiconductor industry was the introduction of a new type of multi-gate three-dimensional transistors [1]. This technology combined with strain techniques and high-k/metal gate dielectrics offers great performance and power advantages over the planar structures and allows steady scaling to 14nm. There are good indications that device miniaturization with necessary technological breakthrough will continue its pace down to the 10nm technology node. In order to sustain further scaling a breakthrough will continue its pace down to the 10nm technology node. In order to sustain further scaling a possible modification in the channel material with improved characteristics, which until recently was kept pristine, is foreseen.

The principle of MOSFET operation is fundamentally based on the charge degree of freedom of an electron: the electron charge interacts with the gate induced electrostatic field which can close the transistor by creating a potential barrier. Another intrinsic electron property, the electron spin, attracts much attention as a possible candidate for complimenting or even replacing the charge in future electron devices. The electron spin state is characterized by one of the two of its possible projections on a given axis and could be potentially used in digital information processing. In addition, it takes an amazingly small amount of energy to invert the spin orientation, which is necessary for low power applications.

Until recently silicon, the main material used by modern microelectronics, was remaining aside from the main stream of spin-related applications. Certainly, the use of silicon for spin driven devices would greatly facilitate their integration with MOSFETs on the same chip. In addition, silicon possesses several unique properties extremely attractive for spin-driven applications. As already mentioned, nuclei of the $^{28}$Si isotope are characterized by zero spin, which favors longer spin lifetime. Another source of dephasing, the spin-orbit interaction, is also weak in silicon. Because of these properties electron spin states of conduction electrons in silicon should show better stability and lower decoherence, which makes silicon a perfect candidate for spin driven device applications. However, even a demonstration of basic elements necessary for spin related applications, such as injection of spin-polarized currents in silicon, spin transport, spin manipulation, and detection, was missing until recently. Although it should be straightforward to inject spin-polarized carriers into silicon from a ferromagnetic contact, due to a fundamental conductivity mismatch problem [2] between a ferromagnetic metal contact and the semiconductor the problem was lacking a solution for a long time. A special technique [3] based on the attenuation of hot electrons with spins down anti-parallel to the magnetization of a ferromagnetic film allowed creating an imbalance between the electrons with spins up and down in silicon thus injecting spin-polarized current. Spin coherent transport through the device was studied by applying an external magnetic field causing the precession of spins during their propagation from source to drain. The detection is performed with a similar hot electron spin filter. Although the drain current is fairly small due to the carriers' attenuation in the source and drain filters as compared to the current of injected spins, the experimental set-up represents a first working spin driven device which can be envisaged to operate at room temperature. Contrary to the MOSFET, however, the described structure is a two-terminal device. Nevertheless, the first demonstration of the coherent spin transport through an undoped 350µm thick silicon wafer [4] has
II. SPIN TRANSPORT

For proper functionality of spin-based devices the necessity to transfer the spin information through the channel is essential. An example of such a device, a SpinFET, combining the advantages of spin with MOSFET compatibility was proposed more than two decades ago [6]. However, it has not been experimentally demonstrated yet. In order to realize a SpinFET, the following requirements must be fulfilled [5]. First, an efficient spin injection in the channel (and detection) must be realized. Second, because the electron spin in the channel is not a conserved quantity and thus relaxes due to spin-flip processes, the corresponding scattering mechanisms must be detected and analyzed. It is important to identify the instruments and tools compatible with modern MOSFET technology, which can enhance the spin lifetime and spin diffusion length in the silicon channel. Finally, purely electrical means of spin manipulation in the channel must be identified to control the spin and thus the current flowing to the drain. An example of such a manipulation is the gate voltage dependent effective spin-orbit interaction defining the degree of the spin precession.

While diffusing, the injected excess spin gradually relaxes to its equilibrium value which is zero in a non-magnetic semiconductor. The lower estimation for the spin lifetime at room temperature obtained within the three-terminal injection scheme was of the order 0.1-1ns [5]. This corresponds to the spin diffusion length - the length at which the spin relaxes - 0.2-0.5µm.

The spin lifetime is determined by the spin-flip processes. Several important spin relaxation mechanisms are identified [7]. In silicon the spin relaxation due to the hyperfine interaction of spins with the magnetic moments of the $^{29}$Si nuclei is important at low temperature. At elevated temperatures the spin relaxation due to the Elliot-Yafet mechanism becomes dominant.

The Elliot-Yafet mechanism is mediated by the intrinsic interaction between the orbital motion of an electron and its spin. The microscopic spin-orbit interaction does not commute with the spin and can therefore generate spin flips. This is the Yafet process. With the microscopic spin-orbit interaction included in the Hamiltonian the Bloch wave function with a fixed spin projection is not an eigenfunction of the Hamiltonian. Since the eigenfunction always contains a contribution with an opposite spin projection, any spin-independent scattering, for instance with phonons, generates a small but finite probability to flip the spin. This is the Elliot process.

In order to analyze the spin relaxation in silicon both Elliot and Yafet contributions must be taken into account. A good agreement between the experimentally observed and calculated spin lifetimes as a function of temperature was achieved recently confirming that the Elliot-Yafet mechanism is the dominant spin relaxation mechanism at ambient temperatures in bulk silicon [8]. The spin lifetime in undoped silicon at room temperature is about 10ns which corresponds to a spin diffusion length of 2µm. The main contribution to the spin relaxation was identified to be due to optical phonon scattering between the valleys residing at different crystallographic axes [9], [10].

In heavily doped silicon the spin lifetime is reduced due to ionized impurity scattering and is expected to be around 1ns at $N_d=10^{19}$cm$^{-3}$, in agreement with experiments [5].

A. Spin lifetime enhancement in silicon films

In gated silicon systems and MOSFETs a relatively large spin relaxation was observed experimentally [11], [12]. This indicates that extrinsic interface induced spin relaxation mechanisms become important. Because it may pose an obstacle in realizing spin driven CMOS compatible devices, a deeper understanding of fundamental spin relaxation mechanisms in silicon inversion layers, thin films and fins is urgently needed.

The theory of spin relaxation must account for the most relevant scattering mechanisms which are due to the electron-phonon interaction and surface roughness scattering. In order to evaluate the corresponding scattering matrix elements the wave functions must be known. To find the wave functions, an approach based on an effective $k\cdot p$ Hamiltonian appears to be sufficiently rigorous to capture the most important physics while still allowing to keep the computational efforts bearable. The effective $k\cdot p$ Hamiltonian must include the effective spin-orbit interaction which, apart from scattering, is the main ingredient of the Elliot-Yafet spin relaxation mechanism. In addition, a confinement potential must be included. It is also mandatory to have various other effects on the band structure, such as band non-parabolicity, warping, and splitting by external stress [13], [14].

For (001) oriented thin films one has to consider the two relevant valleys along the [001] crystallographic axis under the assumption that the spin is injected along the $Z$-axis. The spin relaxation due to surface roughness scattering and electron-phonon interaction is considered. The surface roughness at the two interfaces is assumed to be uncorrelated. At each interface it is described by its mean value and correlation length. The corresponding spin relaxation matrix elements are taken proportional to the square of the product of the subband function derivatives at the interface [15]. The spin relaxation due to the electron-phonon interaction is taken in the deformation potential approximation [16].

In unstrained films the two lower subbands are degenerate [14]. This degeneracy, originating in the $Z$-valleys’ degeneracy, produces a large mixing between the spin-up and spin-down states from the opposite valleys, resulting in hot spin relaxation spots. Figure 1 demonstrates the location of the hot spots in the momentum space. When
shear strain is applied, the hot spots are moved towards higher energy away from the subbands’ center and eventually end up in the region with no states occupied. This results in a strong reduction of the hot spots’ contribution to the spin relaxation. A strong increase of the spin lifetime with shear strain [16] is demonstrated in Figure 2. Therefore, tensile shear strain employed to enhance mobility and on-current in modern transistors can also be used to boost the spin life time in ultra-thin film/ultra-thin box silicon-on-insulator transistors.

B. Spin injection

Spin injection in silicon and other semiconductors by purely electrical means from a ferromagnetic metal electrode was not very successful until recently. The fundamental reason was identified as an impedance mismatch problem [2]. Even though there is a large spin imbalance between the majority and minority spins in a metal ferromagnet, both channels with spin up and spin down are equally populated due to the relatively small density of states in a semiconductor as compared to that for the minority spins in a ferromagnet. A solution to the problem is to introduce a potential barrier between the metal ferromagnet and the semiconductor [17]. In this case the influx of carriers from the ferromagnet into the semiconductor is reduced to such an extent that the majority spins supply just sufficient carriers to support the complete occupancy of the corresponding states in the semiconductor. The minority spin flow in the semiconductor will be a fraction of that for the majority spins defined by the spin polarization in the ferromagnet. This guarantees the existence of the spin polarized current and spin injection in the semiconductor. Room temperature spin injection into n- and p-doped silicon was first demonstrated in 2009 [3] by using a Ni80Fe20/Al2O3 tunnel contact. The authors used heavily doped silicon samples to avoid extended the depletion layer causing large tunnel barriers. It was actually the depletion layer, not the impedance mismatch problem, which for a long time prevented a successful all electrical demonstration of spin injection into silicon. It turns out that nearly all tunnel contacts fabricated so far are characterized by a contact resistance larger than the optimal one [3], and the problem of making good contacts with lower resistance-per-area characteristics still requires some attention. Recently, the tunnel contacts made of a single graphene layer [18] were shown to deliver the contact resistance close to the optimal one. Currently, a reliable injection of spin in n- and p-doped silicon has been demonstrated from a number of ferromagnetic electrodes through several dielectric tunnel barriers [3].

One of the methods to analyze spin injection into a semiconductor is a three-terminal method [3]. It uses the two contacts to introduce the electrical current into the semiconductor. One of the contacts is a tunnel contact with a ferromagnet. When current is flowing, a spin accumulation is created under this tunnel contact. The spin accumulation leads to a difference in chemical potential of spins aligned or anti-aligned to the contact magnetization direction. Since the ferromagnetic contact potential is equal to that of the spins aligned with its magnetization, a detectable voltage is generated between the ferromagnetic electrode and the third contact located at a distance larger than the spin diffusion length where the spin accumulation vanishes.
The SpinFET is a future semiconductor spintronic device with a performance superior to that achieved in the present transistor technology. SpinFETs are composed of two ferromagnetic contacts (source and drain), linked by a non-magnetic semiconductor channel region. Ferromagnetic contacts inject and detect spin-polarized electrons, in analogy to polarizer and analyzer as envisioned by Datta and Das [6]. The electrons with spin aligned to the drain magnetization direction can easily leave the channel to the drain thus contributing to the current. The total current through the device depends on the relative angle between the magnetization direction of the drain contact playing the role of an analyzer and the electron spin polarization at the end of the semiconductor channel. Additional current modulation is achieved by tuning the strength of the effective spin-orbit interaction in the semiconductor region. Because of the non-zero effective spin-orbit interaction in the channel the spin of an electron injected from the source starts to precess. The strength of the spin-orbit interaction in the channel depends on the effective electric field and can be controlled by purely electrical means applying voltage to the gate.

The strength of the spin–orbit interaction determines the minimum length of the semiconductor channel sufficient to change the orientation of spin to the opposite. The spin-orbit coupling is usually taken in the Rashba form [21]. In silicon films with interface-induced inversion symmetry broken both Rashba and Dresselhaus-like spin-orbit terms are allowed [22]. By carefully considering an atomistically irregular interface structure [23] it was shown that the dominant contribution in silicon films is of Dresselhaus type.

Stronger spin-orbit interaction leads to an increased spin relaxation. In quasi-one-dimensional electron structures, however, a suppression of this spin relaxation due to the Dyakonov-Perel’ mechanism is expected [24]. Indeed, in case of elastic scattering only back-scattering is possible. Reversal of the electron momentum results in the inversion of the effective magnetic field direction. Therefore, the precession angle does not depend on the number of scattering events along the carrier trajectory in the channel, but is a function of the channel length alone. Thus, the spin-independent elastic scattering does not result in additional spin decoherence. In the presence of an external magnetic field, however, spin-flip processes become possible, and the Elliott-Yafet spin relaxation mechanism is likely relevant [25], as already explained.

We consider square silicon fins with [100] or [110] orientation, with (001) horizontal faces. The parabolic band approach for the band structure in silicon is not sufficient to accurately obtain the subband structure in thin and narrow silicon fins. We employ the two-band k·p model proposed in [13], [14], which has been shown to be accurate up to 0.5eV above the conduction band edge in silicon [26]. The resulting Schrödinger differential equation, with the confinement potential appropriately added to the Hamiltonian, is discretized using the box integration method and solved for each value of the conserved momentum along the current direction using efficient numerical algorithms available through the Vienna Schrödinger-Poisson framework VSP [27]. The dependence of the effective mass of the ground subband in [100] fins on the film thickness t is more pronounced as compared to [110] fins. With the values of the effective masses and subband offsets obtained, we study the conductance properties for the parallel and anti-parallel configurations of the contact magnetization.

Fins with [100] orientation show a stronger dependence of the tunneling magnetoresistance on the interaction strength as compared to [110] oriented fins. Thus, [100] oriented fins are preferred for silicon SpinFETs. The reason of the stronger dependence is that the characteristic length on which the spin-orbit interaction produces the full spin precession is inversely proportional to the product of the effective mass and the strength of the spin-orbit interaction. As shown in Figure 4, the effective mass value for the [110] oriented fins is smaller compared.
to the [100] oriented fins, hence, for the same length a larger variation of the spin-orbit interaction is required to achieve the same TMR value modulation in [110] fins [28].

Fig. 5 shows the tunnel magnetoresistance (TMR) modulation as a function of the spin-orbit interaction strength at different temperatures. Thanks to the presence of the Schottky barriers between the channel and the source/drain electrodes the TMR signal is preserved at high temperatures. This opens the possibility to modulate the TMR by changing the strength of the spin-orbit interaction even at room temperature. However, because the spin-orbit interaction in silicon is weak, the channel length needed to achieve the TMR modulation is close to a micron [29]. For short channels, the only option to realize a SpinFET is to exploit the relative magnetic orientation of the source and drain ferromagnetic contacts [30, 31]. This adds a possibility to reprogram MOSFETs and to obtain a different current under the same drain and gate voltage by changing the drain magnetization orientation. It is important that, once modified, the magnetization remains the same infinitely long without any extra power applied. This property is used in emerging magnetic nonvolatile memories discussed in the next section.

IV. SPIN-BASED MEMORIES

Apart from good scalability, a new type of memory must exhibit low operating voltages, low power consumption, high operation speed, long retention time, high endurance, and a simple structure [32]. One of the most promising candidates for future universal memory among emerging technologies is spin transfer torque magnetic RAM (STT-MRAM). Currently, STT-MRAM has been demonstrated on 64Mb test chips [33]. The basic element of an STT-MRAM is a magnetic tunnel junction (MTJ), a sandwich of two magnetic layers separated by a thin non-magnetic spacer. While the magnetization of the pinned layer is fixed due to the fabrication process, the magnetization direction of the free layer can be switched between the two states parallel and anti-parallel to the fixed magnetization direction. Switching between these two states occurs due to the spin torque exerted on a free layer by spin-polarized current flowing through the MTJ. The spin-polarized current is only a fraction of the total charge current. Therefore, relatively high current densities are required to switch the magnetization direction of the free layer. The reduction of the current density required for switching and/or the increase of the switching speed without compromising the thermal stability are the most important challenges in this area [34]. Measurements [35] showed a decrease in the critical current density for a penta-layer MTJ with the two pinned magnetic layers in the anti-parallel configuration (Fig. 6, left) compared to the three-layer MTJ. Theoretical predictions showed a decrease of the switching time in penta-layer structures with an increase of the out-of-plane component of the magnetostatic field [36]. An even more pronounced decrease of the switching time (Fig. 7) and current density were recently reported in a penta-layer structure with a composite free layer (Fig. 6, right) [37]. The non-zero angle between the fixed magnetization and the magnetization in the free layer results in an enhanced spin transfer torque, when the current starts flowing. In the case of the monolithic structure, however, the torque remains marginal in the central region. As the amplitude of the end domains precession increases, the central region experiences almost no spin torque and preserves its initial orientation, thus preventing the whole layer from alternating its magnetization orientation. This is, however, not the case, when the central region is removed in the composite structure and the end domains become virtually independent. A reduction of the switching current is also
observed in MTJs with perpendicular magnetization [38].

STT-MRAM based on CoFeB-MgO MTJs with interfacial anisotropy is already close to overcome the scalability limit of charge based storage memories. However, such devices still require reducing damping and increasing thermal stability [39]. For a device with its dimension smaller than 30nm, it is necessary to use materials with higher anisotropy such as Co/Pt multilayers or a FePt system. However, radically new approaches are required to write data on such layers without sacrificing speed [40].

V. SPIN-BASED LOGIC CIRCUITS

Using spin-transfer torque to switch a magnetic tunnel junction is one of the most promising non-volatile storage technologies, which combines the advantages of CMOS compatibility, high speed, high density, unlimited endurance, and scalability. Distributing non-volatile memory elements over the CMOS logic circuit plane (so-called logic-in-memory architecture) can provide extremely low standby power consumption and instant start-up by holding the information in the MTJs and eliminating the need for refreshing pulses which are critical for CMOS-based memory elements. Furthermore, by using the MTJ technology the effective area and interconnect delay can be reduced due to easy three-dimensional integration of the MTJs on top of the CMOS layers. However, in hybrid CMOS/MTJ circuits the MTJs are used only as ancillary devices which store the computation results. Therefore, sensing amplifiers are required for reading the data at each logic stage and providing the next stage with an appropriate voltage or current signal as input. This limitation increases the device count, delay, and power consumption. Furthermore, the generalization to large-scale logic systems is problematic.

Recently, MTJ-based implication logic gates (Fig. 8a) have been proposed [41], in which the MTJs are used as the main devices for logical computations and thus intrinsically enable logic-in-memory architectures. By replacing the MTJ devices with one-transistor/one MTJ (1T/1MTJ) cells (Fig. 8b), the implication logic gates are realized in MRAM arrays to provide large-scale non-volatile magnetic circuits [42]. Due to the structural asymmetry caused by the resistor $R_G$, two MRAM arrays are required in this implementation. For performing the implication operation, a source (target) MTJ can be selected only in the MRAM array which is (not) serially connected to $R_G$. Therefore, read/write operations are required to readout the output of any logic operation from the target array and to write it in the source array as an input for the next logic steps. This asymmetry issue is addressed in [43] by using the access transistors as
Fig. 9: Coupled MRAM arrays realizing parallel MRAM-based computations [48].

voltage-controlled resistors, which enables highly flexible nonlocal computations without any need for intermediary read/write operations. Therefore, it addresses also the multiple non-volatile fan-out issue [45].

Furthermore, a MRAM-based implementation enables independent access to all input MTJs for STT writing and thus brings significant advantages related to scalability and energy consumption [46]. However, the nonzero On-resistance of the access transistors decreases the effective TMR of the 1T/1MTJ cells by about 10% [47] and this increases the average error probabilities by a factor of less than two [45].

Based on the symmetric implication, parallel MRAM arrays can be used to perform simultaneous operations on the same word lines to decrease the total number of required serial steps for implementing complex Boolean functions [48]. Parallel MRAM-based computation is performed by applying the same current pulses to arbitrary bit lines and two voltage pulses to the word lines simultaneously, by using the bit line and word line drivers shown in Fig. 9.

VI. CONCLUSION

Because of recent ground-breaking experiments silicon is now gaining momentum for electronic applications involving spin. It was demonstrated how strain could be used to boost the spin lifetime. An efficient coupling between the electrical and the magnetic degree of freedom makes STT-MRAM a viable candidate for future universal memory which is fast, non-volatile, and CMOS compatible. The use of STT-MRAM cells as implication logic gates opens the way towards intrinsic logic-in-memory architecture where the same elements are employed to store and to process the information.

ACKNOWLEDGEMENT

This work is supported by the European Research Council through the grant #247056 MOSILSPIN. The computational results have been achieved in part using the Vienna Scientific Cluster (VSC).

REFERENCES


