Chapter 2
Fundamentals of NVM Physics and Computing

Abstract The bistable states are the foundation of all memory devices to store data. For conventional memory devices, the bistable states are represented by voltage levels and the transition is described by the charging and discharging of the capacitors. The transition dynamics is critical in order to obtain important figures of merit such as device operation speed and energy. Therefore, it is of great importance to quantitatively understand the physical mechanism and transition dynamics of the emerging nonvolatile devices, whose states are represented by nonelectrical variables. For the magnetoresistive random-access memory family, including toggled MRAM, STT-MRAM, and racetrack memory, the magnetization dynamics is the fundamental physics behind, while for the resistive random-access memory category, including memristor and CBRAM, the ion migration effect is the shared physics. In this chapter, both the magnetization dynamics and ion migration dynamics are introduced.

Keywords Magnetization • Ion migration • Logic-in-memory architecture • In-memory computing

2.1 Nonvolatile Memory Physics

2.1.1 Magnetization

A large portion of the emerging nonvolatile memory devices are magnetization based. The MRAM usually is formed by one insulator in the middle, sandwiched by two ferromagnetic layers, namely fixed layer that is strongly magnetized and free layer that can be easily changed. Differed by the approaches for writing, there are several phases of MRAM technology. The first-generation MRAM needs external magnetic field to switch the free layer magnetization. The second-generation STT-RAM is introduced and the free layer magnetization can be altered by polarized current, which brought significant advantages such as easy

H. Yu and Y. Wang, *Design Exploration of Emerging Nano-scale Non-volatile Memory*,
DOI 10.1007/978-1-4939-0551-5_2,

integration with current CMOS technology and high density, high reliability, etc. Recently, the third-generation domain-wall racetrack is introduced, with a series of magnetization domains in one ferromagnetic thin-film nanowire and additional shift ability. The shift is also current-induced operation. In this section, the magnetization dynamics under external field and spin current in nanosecond regime will be introduced.

2.1.1.1 Basic Magnetization Process

As an intrinsic property, electrons spin about its axis and produce magnetic field like current-carrying wire loop. From macrospin point of view, the relation between magnetization M and angular momentum associated with electron spin S can be expressed as

$$M = -\gamma S, \tag{2.1}$$

where $\gamma = 2.21 \times 10^5\ \mathrm{mA^{-1}\ s^{-1}}$ is the gyromagnetic ratio. A uniform magnetic field exerts no net force on a current loop, but it does exert a net torque, and the torque T, on the current-carrying loop under applied magnetic field H, can be expressed as

$$T = M \times H. \tag{2.2}$$

By definition, the time derivative of angular momentum is called torque. The relation between the angular momentum L and torque T reads

$$\frac{dL}{dt} = T. \tag{2.3}$$

The quantum form of Eq. (2.3) still remains valid, and then we have

$$\frac{dS}{dt} = T. \tag{2.4}$$

By combining Eqs. (2.1), (2.2), and (2.4), we can obtain the motion equation of magnetization under applied magnetic field:

$$\frac{dM}{dt} = -\gamma M \times H. \tag{2.5}$$

The processional motion described by Eq. (2.5) indicates that the magnitude of magnetization will not change and also the angle between H and M will not change, which is depicted in Fig. 2.1. This is based on that no energy loss is assumed during this process.

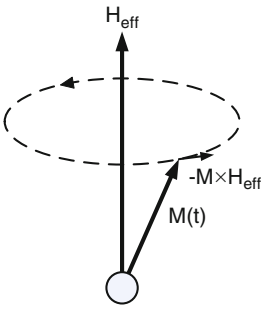

Fig. 2.1 The magnetization precession

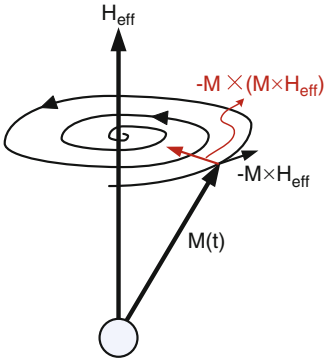

Fig. 2.2 The magnetization precession with damping

2.1.1.2 Magnetization Damping

In real systems, however, energy is dissipated through various means and the magnetization motion is damped until an equilibrium is reached. Energy dissipation can occur through the spin–spin, spin–photon, and spin–electron interactions through which the spin energy is transferred. The approach followed by Landau and Lifshitz is to introduce dissipation in a phenomenological way. In fact, they introduce an additional torque term that pushes magnetization in the direction of the effective field. Landau–Lifshitz equation in the Gilbert form, or LLG equation, then reads

$$\frac{dM}{dt} = -\gamma M \times H + \frac{\alpha}{M_s} M \times \frac{dM}{dt}. \tag{2.6}$$

The magnetization dynamics described by Eq. (2.6) is sketched in Fig. 2.2.

2.1.1.3 Spin-Transfer Torque

In 1996, Berger [3] and Slonczewski [14] predicted, which later has been confirmed experimentally [7, 16, 20], that electrons that carry enough angular momentum are able to cause magnetization precession and switching by spin-transfer torque effect. When a current passes through a ferromagnetic layer, its electron spins are polarized

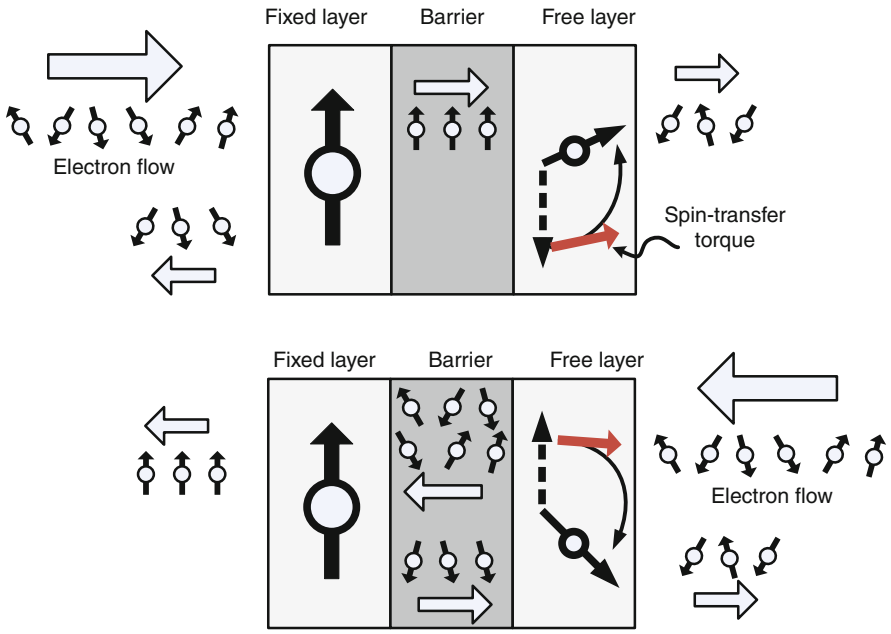

Fig. 2.3 The spin-transfer torque effect

along the magnetization direction and the current becomes spin polarized and hence carries angular momentum. And the spin-polarized current, when flows through a second ferromagnetic layer, exerts a spin torque on the local magnetic moment of the magnetic layer and causes the magnetization precession and switching when the current is large enough (Fig. 2.3).

Therefore, the dynamics of the free layer magnetization can be determined by the LLG equation in conjunction with an additional term for spin-transfer torque:

$$\frac{dM}{dt} = -\gamma M \times H + \frac{\alpha}{M_s} M \times \frac{dM}{dt} - \frac{a_J}{M_s}[M \times (M \times P)], \tag{2.7}$$

where P is the magnetization of fixed layer, M_s is the saturation magnetization, a_J is a factor related to the interfacial interaction between magnetic moment and spin-polarized electrons, which is proportional to the current density, and the sign of a_J depends on the direction of current. When applied properly, the current is able to cancel the damping and switch the magnetization of free layer by spin-transfer torque.

For the current-induced magnetization precession, there exists a threshold current density J_{c0}, and by applying current larger than J_{c0} the magnetization can be switched back and forth:

$$J_{c0} = \frac{2e\alpha M_s t_F (H_K + H_{\text{ext}} + 2\pi M_s)}{\hbar \eta}, \tag{2.8}$$

where e is the electron charge, α the damping constant, M_s the saturation magnetization, t_F the thickness of free layer, $\hbar$ the Planck constant, η the spin-transfer efficiency, H_K the anisotropy field, and H_{ext} the external applied magnetic field.

There are three modes for the current-driven magnetization switching: thermal activation associated with switching time longer than 10 ns, processional switching associated with switching time less than a few nanoseconds, and dynamic reversal as a compound process of both. The above three modes reveal the switching time and current density relationship.

For the fast processional switching, the switching time is reversely proportional to the applied current:

$$\tau_p \propto \frac{1}{(J - J_{c0})} \ln\left(\frac{\pi}{2\theta_0}\right), \tag{2.9}$$

where θ_0 is the initial magnetization angle deviated from the easy axis. At finite temperature, θ_0 is determined by thermal distribution. For the fast processional switching in the regime of nanosecond, it usually takes a current density that is several times greater than J_{c0}.

In the slow thermal-activated switch regime, the switching current is dependent on the current pulse width and thermal stability factor $\Delta = K_u V / k_B T$ of the free layer. Interestingly, the current density can be smaller than the critical density and therefore is useful for current reduction. In this case, the standard thermal agitation will be assisted by spin current, which introduces extra energy to reach enough energy for the magnetization switching. The relation reads

$$J(\tau) = J_{c0}\left[1 - \frac{K_u V}{k_B T} \ln\left(\frac{\tau}{\tau_0}\right)\right], \tag{2.10}$$

where $\tau_0 \propto 1$ ns is the inverse of the attempt frequency and $K_u V$ the anisotropy energy.

As the fast processional switching requires large current density which reduces the robustness and causes undesired switching and the slow thermal activation process takes too long time, the most interesting switching mode is the dynamic reversal at intermediate current pulses. Although the dynamic switching mode corresponds to the operating speed of practical STT-RAM, the explicit formula is hard to be derived due to its complicated process. Therefore, the dynamic reversal is usually studied as a combination of processional and thermally activated switching.

2.1.1.4 Magnetization Dynamics

Magnetic domains are formed by the competition between the various energy terms involved in a magnetic object. The energy of a magnetic structure is the sum of the exchange energy, the anisotropy energy, the Zeeman energy, and the demagnetization energy. The magnetic system seeks to minimize its overall

free energy. Since the magnitude of the magnetization cannot change the way to minimize the energy is to vary the direction of the magnetization. The exchange energy seeks to align the spins with each other, the anisotropy energy seeks to align the spins with an axis determined by the crystal structure, and the Zeeman energy aligns the spins with an external field. When the magnetostatic dipole–dipole interaction is also taken into account, known as the demagnetization energy, a nonuniform magnetization will generally be found as the lowest compromise of overall energy. Short-range exchange energy will prevail a configuration with the spins aligned, and long-range dipole–dipole interaction will however prevail a magnetic state with minimal net magnetization. In the macrospin model of magnetization dynamics study, the short-range exchange energy can be ignored. The energy associated with anisotropy field can be written as

$$\epsilon = K(1 - m_x^2), \tag{2.11}$$

where K is the anisotropy constant and m_x the normalized magnetization in x-direction, defined as the in-plane easy axis.

For the Zeeman energy by external applied field, we have

$$\epsilon = -\mu_0 M H_{\text{ext}} \tag{2.12}$$

in which μ_0 is called the vacuum permeability.

Demagnetization field represents the work necessary to assemble magnetic poles in a given geometric configuration. Since the thickness is so small compared to the in-plane dimensions, the dominant term is approximately the demagnetizing field of a uniformly magnetized thin film with infinite lateral dimensions, namely $H_D = [0, 0, -M_S m_z]$. The associated energy density is

$$\epsilon = -\frac{1}{2}\mu_0 M H_D = -\frac{1}{2}\mu_0 M_s^2 H_D^2. \tag{2.13}$$

Combining all three terms together, we have the overall energy:

$$\epsilon = K(1 - m_x^2) + \frac{1}{2}\mu_0 M_s^2 m_z^2 - \mu_0 M \cdot H_{\text{ext}}. \tag{2.14}$$

When pulled out of equilibrium, the magnetization is subject to an effective field. Therefore, we are able to calculate the effective field H in the LLG Eq. (2.7):

$$H_{\text{eff}} = -\frac{1}{\mu_0 M_s}\frac{\delta \epsilon}{\delta m} = [H_x^{\text{ext}} + H_K m_x, 0, -M_s m_z], \tag{2.15}$$

where $H_K = 2K/(\mu_0 M_s)$ is the anisotropy field.

In dimensionless form, we have

$$\omega = \frac{1}{2}Q(1 - m_x^2) + \frac{1}{2}m_z^2 - m \cdot h^{\text{ext}} \tag{2.16}$$

and

$$h_{\text{eff}} = -\frac{\delta\omega}{\delta m} = [h_x^{\text{ext}} + Qm_x, 0, -m_z] \tag{2.17}$$

with $Q = 2K/(mu_0M_s^2) = H_K/M_s$. Since the magnitude of m does not change, which suggests that it can be rewritten in spherical coordinates,

$$\frac{d\theta}{d\tau} = h_\phi - \alpha \sin\theta \frac{d\phi}{d\tau} \tag{2.18}$$

$$\sin\theta \frac{d\phi}{d\tau} = -h_\theta + \alpha \frac{d\theta}{d\tau} \tag{2.19}$$

by multiplying α to Eq. (2.18), adding Eq. (2.19), multiplying α to Eq. (2.19), and subtracting Eq. (2.19), we could obtain a set of first-order differential equations:

$$(1+\alpha^2)\frac{d\theta}{d\tau} = h_\phi - \alpha h_\theta \tag{2.20}$$

$$(1+\alpha^2)\sin\theta \frac{d\phi}{d\tau} = -h_\theta + \alpha h_\phi. \tag{2.21}$$

The stable processional states can be obtained by numerical integration of Eqs. (2.20) and (2.21), as first demonstrated by Sun [17]. With approximation, the current threshold for the establishment of a stable magnetization trajectory may be simply derived from standard perturbation theory. Clearly, in the absence of current and under the action of any applied field $h_x^{\text{ext}} > 0$, the stable magnetization direction satisfies $m_x = 1$, or, $\theta = \pi/2$, $\phi = 0$. As the damping constant α is generally small, so that in the investigated trajectory zone, the spin-transfer torque roughly balances the damping. Therefore χ is of the same order of magnitude as α, and both can be treated as small parameters. To investigate the stable processional states, we focus on trajectories in which the magnetization is close to its equilibrium states. This suggests the replacement $\theta = \pi/2 + \xi$, so that ξ and ϕ can be treated as small. Taking $1 + \alpha^2 \cong 1$ leads to the following linearized equations of magnetization motion:

$$\frac{d\xi}{d\tau} = h_\phi - \alpha h_\theta \tag{2.22}$$

$$\frac{d\phi}{d\tau} = -h_\theta + \alpha h_\phi \tag{2.23}$$

with

$$h_\theta = -(1 + Q + h_x^{\text{ext}})\xi - \chi\theta \tag{2.24}$$

$$h_\phi = +\chi\xi - (Q + h_x^{\text{ext}})\phi. \tag{2.25}$$

Let $u = Q + h_x^{\text{ext}}$ and λ be the first derivative operator $d/d\tau$; the first-order differential equation then reads

$$\lambda^2 + [\alpha + 2(\alpha u - \chi)]\lambda + u(1+u) = 0 \tag{2.26}$$

solution precession

$$\theta = e^{-\frac{t}{t_0}} \cos(\omega t + \Phi_0), \tag{2.27}$$

where

$$t_0 = \frac{1}{\gamma_0 M_s(\chi_{\text{crit}} - \chi)} \tag{2.28}$$

$$\omega = \gamma_0 M_s \sqrt{u(1+u) - (\chi_{\text{crit}} - \chi)^2} \tag{2.29}$$

with

$$\chi_{\text{crit}} = \alpha(\frac{1}{2} + Q + h_x^{\text{ext}}) \cong \frac{\alpha}{2} \tag{2.30}$$

if $Q, h_x^{\text{ext}} \ll 1$.

2.1.1.5 Domain-Wall Propagation

Recalling spin-transfer torque effect that when a current is passed through a ferromagnetic material, electrons will polarize, that is, the spin of the conduction electron will align with the spin of the local electrons carrying the magnetic moment of the material. When the conduction electrons subsequently enter a region of opposite magnetization they will eventually become polarized again, thereby transferring their spin momentum to the local magnetic moment, as required by the law of conservation of momentum. Therefore, when many electrons are traversing a domain wall (DW), magnetization from one side of the DW will be transferred to the other side. Effectively the electrons are able to push the DW in the direction of the electron flow i.

The influence of current on DW dynamics is often treated by including two spin torque terms in the LLG equation, Eq. (2.6). When the current, with current density J, is flowing in one direction, the x-direction, the LLG equation including the spin torque terms can be written as

$$\frac{\partial M}{\partial t} = -\gamma M \times H + \frac{\alpha}{M_s} M \times \frac{\partial M}{\partial t} - \eta J \frac{\partial M}{\partial x} + \beta \eta J M \times \frac{\partial M}{\partial x}, \tag{2.31}$$

where last two terms are added to the regular LLG equation to describe the effect of current on the magnetization dynamics. The first of these terms expresses the

adiabatic spin-transfer torque as exerted by a current on magnetic DWs with η the strength of the effect. The second STT term in the equation describes the nonadiabatic current-induced effect which relative strength is parameterized by β. The strength of the adiabatic spin torque, η, is widely agreed on [2, 10, 18, 19] and given by

$$\eta = \frac{g\mu_B P}{2eM_s}, \tag{2.32}$$

where g is the Land factor, μ_B the Bohr magneton, e the electron charge, M_s the saturation magnetization, and P the electron polarization, all of which values are very well known except for the electron polarization. Estimates for P range from $P = 0.4$ to $P = 0.7$ [1].

2.1.2 Ion Migration Dynamics

Unlike the nonvolatile memories from MRAM family that they share same physical dynamics as magnetization precession, the physics behind other emerging memories are more or less different from each other. From a mechanistic point of view, therefore, it is hard to conclude one universal physical model that is suitable for all. Nevertheless, they still share many similarities that make it possible for all to have similar phenomenological model, especially for various resistive random-access memory (ReRAM) technologies. For example, the ion migration model proposed by Mott–Gurney [13] is used to describe the kinetics of both CBRAM and memristor devices according to [15, 21], though their types of ion and reaction equations are different from each other. In this section, we will review the ion migration physics that is able to describe dynamics of thin-film ReRAM devices.

The ion migration is involved in many ReRAM switching dynamics. For example, for CBRAM device, its metal electrode will dissolve and the metal ions will transport under applied electric field, and for memristor, oxygen ion and oxygen vacancy migrate according to applied electric field. For thick films, the ion transport follows linear velocity and electric field relationship due to its low electric field [5]; however, for very thin film, which is the case for most nanoscale ReRAM devices, electric field may be high and linear drift law will be invalidate. For example, when a thin-film ReRAM device with $\sim$10 nm thickness is operated at a few volts, its electric field may exceed 1 MV/cm. Such a high electric field can significantly reduce the activation barrier for migration inside the solid, and hence it is natural to suspect there might be strongly nonlinear ionic transport. In fact, such nonlinear ionic transport under strong electric field has been experimentally confirmed [4]. Figure 2.4 shows the linear ionic transport for small electric field and strongly nonlinear transport while electric field is large.

The hyperbolic sine function in the nonlinear drift model predicts the existence of threshold voltage for thin-film ReRAM devices. As the drift velocity has positive

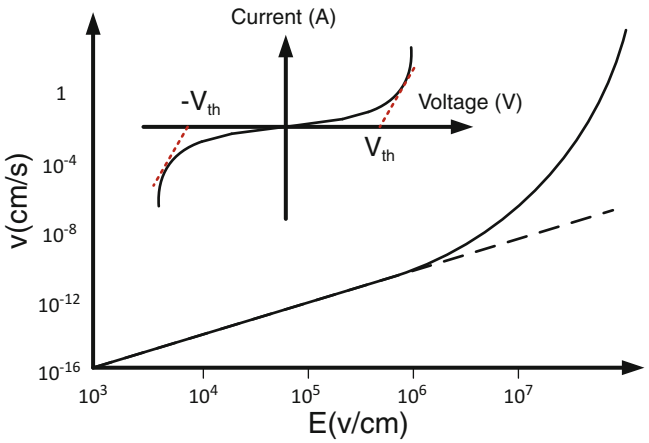

Fig. 2.4 The nonlinear ion drift velocity under strong electric field with introduced bidirectional thresholds for thin-film ReRAM devices

correlation with current density, the nonlinear drift therefore brings nonlinear I-V curve, as shown in Fig. 2.4. Similar to diode, the nonlinear I-V curve also brings a threshold voltage, below which the device can be considered turned off. Different from diode, the hyperbolic sine function indicates that the thin-film ReRAM devices will behave like a bidirectional diode with one positive voltage threshold as well as a mirrored negative voltage threshold.

According to [5], such relationship can be described by hyperbolic sine function,

$$J = 4avn_0 e^{-\frac{U_a}{kT}} \sinh\left(\frac{qaE}{kT}\right), \tag{2.33}$$

where N_i is the density of the metal ions in the solid electrolytes, f is the attempt-to-escape frequency, E_a is the activation energy, E is the electric field, kT is the thermal energy, and a is the effective hopping distance. The illustration of energy profile with definitions of variables is shown in Fig. 2.5.

A simplified derivation of above continuum transport equation that takes account of both concentration gradients and high electrostatic potential gradients will be reviewed in the following. When dealing with equilibrium phenomena of ion migration, one must take consideration of electrochemical potential as they carry a net charge, instead of only normal chemical potential. The electrochemical potential of a species can be defined by

$$\overline{\mu} = \mu_0 + kT \ln(n) + q\phi, \tag{2.34}$$

where μ_0 is the standard chemical potential of the species, q is the charge, ϕ is the local electrostatic potential, and n is the species concentration which can be approximated by

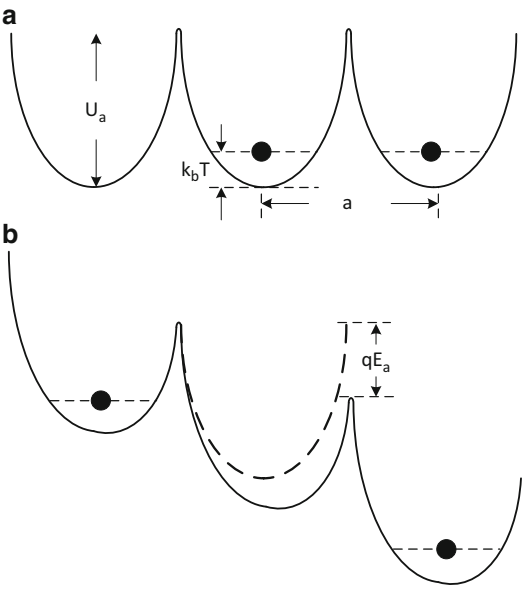

Fig. 2.5 Illustration of vacancy diffusion (**a**) without and (**b**) with an applied electric field with definitions of physical parameters

$$n \propto \exp\left(\frac{-q\phi}{kT}\right). \tag{2.35}$$

It is clear that any valid continuum equation must satisfy the condition that for zero electrochemical potential gradient, the flux density of the corresponding species must be zero. Thus

$$J = 0 \text{ for } \frac{\partial \overline{\mu}}{\partial x} = 0, \tag{2.36}$$

where x is a position variable and J the flux density of the species in the x-direction.

At low field strengths, any general continuum transport equation must reduce to the normal linear transport phenomenological equations. Thus, the following equation can be written as

$$J = -D\frac{\partial n}{\partial x} + nvE\frac{q}{|\,q\,|} \text{ for small } E, \tag{2.37}$$

where D is the diffusion coefficient, v the mobility of the migrating species, and E the electric field strength in the x-direction. By substituting the Einstein relation $v/D = |q|/kT$, Eq. (2.37) can be rewritten in the alternative form:

$$J = \frac{-Dn}{kT}\frac{\partial \overline{\mu}}{\partial x} \text{ for small } E. \tag{2.38}$$

Apparently, Eq. (2.38) satisfies the boundary condition (2.36). And in the absence of a concentration gradient, and for high field strengths, we shall for present purposes regard the flux density as given by equation

$$J \propto \sinh\left(\frac{\mu^* E}{kT}\right) \text{ for } \frac{\partial n}{\partial x} = 0, \tag{2.39}$$

where μ^* is a phenomenological coefficient with the dimensions of an electric dipole. In fact, Eq. (2.39) is constructed more from an experimental phenomenological approach than it is from theoretical mechanistic approach. Experimental justification for the hyperbolic sine function dependence on the field, and for its limiting form an exponential dependence on field, can be found in [13]. Then we have

$$J = \frac{-Dn}{kT} \cdot \frac{\partial \overline{\mu}}{\partial x} \cdot \frac{\sinh(\mu^* E/kT)}{\mu^* E/kT}. \tag{2.40}$$

It is then obvious that Eq. (2.40) satisfies boundary condition (2.36) over the entire range of all the variables. In addition, when field strength is sufficiently low (i.e., $\mu^* E/kT \ll 1$), the sinh function could be replaced by its argument, which reduces conditions Eq. (2.40) to Eq. (2.38). Again for zero concentration gradient, in the case of $\partial\overline{\mu}/\partial x = -qE$, Eq. (2.40) reduces to Eq. (2.39). Equation (2.40) is therefore consistent with the asymptotic equations (2.38) and (2.39) and furthermore satisfies condition (2.36). It is clear that Eq. (2.40) cannot be regarded as unique in satisfying these conditions.

By substituting Eq. (2.34), $D = 4a^2 v \exp(-U_a/kT)$ and $\mu^* = qa$, Eq. (2.40) can be rewritten as

$$J = 4avn_0 e^{-\frac{U_a}{kT}}\left[\sinh\left(\frac{qaE}{kT}\right) - \frac{\partial \ln(n)}{\partial x} a \frac{\sinh(\mu^* E/kT)}{\mu^* E/kT}\right]. \tag{2.41}$$

Practically for thin films it can be approximated that the concentration of mobile ions at the interface towards which the ions are moving can have no direct influence on the transport rate; the latter terms can be omitted, and thus Eq. (2.41) becomes Eq. (2.33).

2.2 Nonvolatile In-Memory Computing

The memory technology can affect the computing performance from two aspects. The first aspect is from memory technology itself. For example, memory access latency, access energy as well as memory density are important figures of merit of memory that tell how fast and how efficiently data can be stored and retrieved. Besides the memory technology itself, the second aspect is the way how memory is integrated with logic. This will greatly affect how fast and how efficiently the retrieved data can be processed by logic units. In this part, the memory and logic integration issues will be discussed.

2.2.1 Memory-Logic-Integration Architecture

Current memory and logic integration has hit the memory wall. That is to say, the memory is the bottleneck of the whole system which is not able to provide data at the rate that processor requires. In this case, the processor is not fully operating. Such hardware resource waste is especially severe for data-intensive applications. This is because current memory and logic components are separated. The data required by logic components will be read out from memory and write back the results to memory through I/O after execution is done. In other words, the linking bridge between logic and memory is the limited I/Os.

The memory-logic throughput can be determined by two factors: I/O pin numbers and how fast the I/O can be operated. Speed-wise, the current I/O can be operated ranging from 100 MB per second per pin for flash memory to 10,000 MB per second per pin for GDDR memory. Although higher I/O operating frequency is desirable, it has fundamental limits such as signal propagation physics (Maxwell's equation) and signal integrity issues (cross talk, loss, and reflection). On the other hand, the I/O number depends on the packaging technology, number of ball bumps, for example. Wider I/O requires higher power and cost budget, and current maximum is about 512 bits. Higher number of pins requires packaging and interconnect breakthroughs. As a rule of thumb, a new generation of packaging technology comes at every six years: 1994 lead technology TSOP, 2000 FBGA (0.8 mm), 2006 PoP/MCM (0.4 mm) and 2012 die stack (40–50 μm). As the pitch of interconnects is getting smaller, it is promising to have more I/O for memory chip.

2.2.2 Logic-in-Memory Architecture

Instead of improving memory bandwidth, it is also possible to reduce the required data communication traffic between memory and processor. The basic idea behind this is that, instead of feeding processor large volume of raw data, it is beneficial to preprocess the data and provide processor only intermediate result. The key to lower communication traffic is the operand reduction. For example, to perform a sum of ten number, instead of transmitting ten number to processor, in-memory architecture is able to obtain the sum by in-memory logic and transmit only one result, thus reducing traffic by 90%. To perform in-memory logic, it is necessary to implement logic inside memory so that preprocessing logic can be done. Such architecture is called logic-in-memory architecture [6, 8, 9, 11, 12]. Figure 2.6 shows logic-in-memory architecture at memory cell level. The example illustrated here is an in-memory full adder with both *sum* logic and *carry* logic.

The basic circuitry, including the access transistor, word-line, and bit-lines, is to ensure memory access. The data is stored in nonvolatile memory devices which have either low or high resistance. Redundant data is required for each bit of data for logic purpose. Combinational logic circuit is added inside which the nonvolatile devices are equivalent to transistors: considered turned on if at low-resistance state

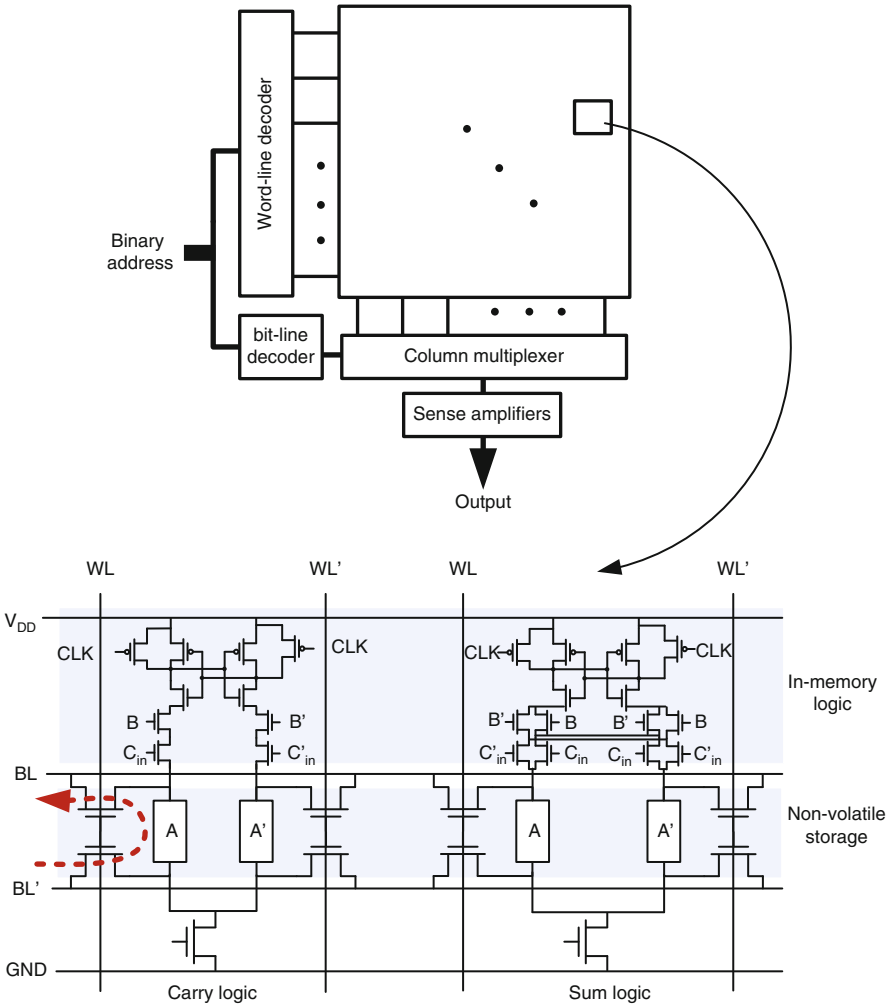

Fig. 2.6 In-memory computing architecture at memory cell level

or turned off if at high-resistance state. In such architecture, the desired result can be obtained immediately without reading operands as if the results are already stored in data array and it can just be "read out." This is very useful for some specific applications as this architecture is able to preprocess data without loading data to processor with extremely short latency.

As the logic is inserted to one cell or a few cells, it is limited to small size; thus, it cannot be made complex. Usually only simple logic is suitable for such architecture; otherwise, the overhead would be overwhelming. However, though simple logic in such architecture is able to share the workload of processor, its effect to reduce communication traffic is not obvious due to limited operand reduction. In addition,

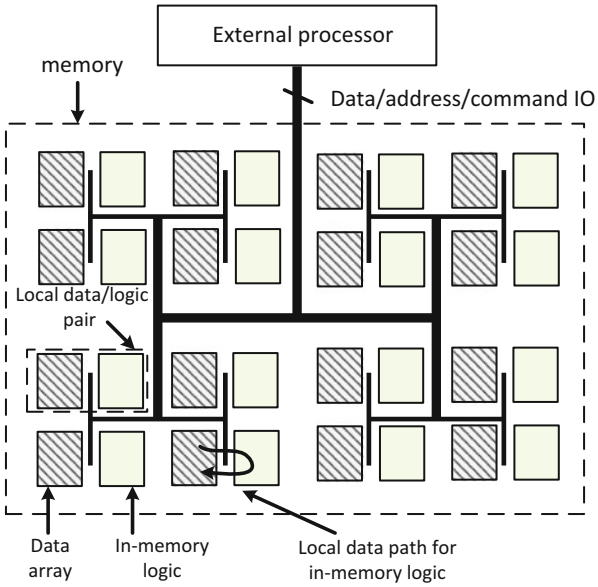

Fig. 2.7 In-memory computing architecture at memory block level

similar to the operation of memory, for the whole data array, only a few logic can be active concurrently at one time. This leads most logic circuit to be idle at most of the time, which not only is a waste of computational resources but also incurs leakage power for CMOS logic. Another disadvantage is that the data needs to be stored in a very strict manner, determined by how the in-memory is organized.

An alternative in-memory architecture at block level which is more suitable for traffic reduction is illustrated in Fig. 2.7. A memory data is usually organized in H-tree fashion, and the data block can be the data array or a number of data arrays that belong to same "H-tree" branch. Instead of inserting in-memory logic at memory cell level inside the data array, the architecture in Fig. 2.7 pairs each block of data with in-memory logic (accelerators). Different from the cell-level in-memory architecture, the accelerators can be made with higher complexity, and the number of accelerators for each data block can also be customized. The data flow of the block-level in-memory architecture is to read out data from data block to in-memory logic, which performs particular functionality and then writes back the result. The data also needs to be stored in assigned blocks, but it is much more flexible than that of cell-level in-memory architecture. The block-level in-memory architecture is very effective to reduce communication traffic between memory and processor. This is because high operand reduction can be achieved due to higher accelerator complexity. For example, for face recognition in image processing application, instead of transmitting a whole image to obtain a Bool result, the result can be directly gained through in-memory logic. In other words, the block-level in-memory architecture is suitable for big data-driven applications where traffic reduction is more important than latency reduction.

References

1. Beach G, Tsoi M, Erskine J (2008) Current-induced domain wall motion. J Magn Magn Mater 320(7):1272–1281
2. Berger L (1978) Low-field magnetoresistance and domain drag in ferromagnets. J Appl Phys 49(3):2156–2161
3. Berger L (1996) Emission of spin waves by a magnetic multilayer traversed by a current. Phys Rev B 54(13):9353
4. Cabrera N, Mott N (1949) Theory of the oxidation of metals. Rep Progress Phys 12(1):163
5. Dignam M (1968) Ion transport in solids under conditions which include large electric fields. J Phys Chem Solids 29(2):249–260
6. Hanyu T, Teranishi K, Kameyama M (1998) Multiple-valued logic-in-memory vlsi based on a floating-gate-mos pass-transistor network. In: Solid-state circuits conference, 1998. Digest of technical papers, 1998 IEEE International. IEEE, New York, pp 194–195
7. Katine J, Albert F, Buhrman R, Myers E, Ralph D (2000) Current-driven magnetization reversal and spin-wave excitations in co/cu/co pillars. Phys Rev Lett 84(14):3149
8. Kautz WH (1969) Cellular logic-in-memory arrays. IEEE Trans Comp 100(8):719–727
9. Kimura H, Hanyu T, Kameyama M, Fujimori Y, Nakamura T, Takasu H (2004) Complementary ferroelectric-capacitor logic for low-power logic-in-memory vlsi. Solid State Circ IEEE J 39(6):919–926
10. Li Z, Zhang S (2004) Domain-wall dynamics and spin-wave excitations with spin-transfer torques. Phys Rev Lett 92(20):207–203
11. Matsunaga S, Hayakawa J, Ikeda S, Miura K, Hasegawa H, Endoh T, Ohno H, Hanyu T (2008) Fabrication of a nonvolatile full adder based on logic-in-memory architecture using magnetic tunnel junctions. Appl Phys Expr 1(9):1301
12. Matsunaga S, Hayakawa J, Ikeda S, Miura K, Endoh T, Ohno H, Hanyu T (2009) Mtj-based nonvolatile logic-in-memory circuit, future prospects and issues. In: Proceedings of the conference on design, automation and test in Europe. European Design and Automation Association, Leuven, pp 433–435
13. Mott NF, Gurney RW (1964) Electronic processes in ionic crystals. Dover, New York
14. Slonczewski JC (1996) Current-driven excitation of magnetic multilayers. J Magn Magn Mater 159(1):L1–L7
15. Strukov DB, Williams RS (2009) Exponential ionic drift: fast switching and low volatility ofáthin-film memristors. Appl Phys A 94(3):515–519
16. Sun J (1999) Current-driven magnetic switching in manganite trilayer junctions. J Magn Magn Mater 202(1):157–162
17. Sun J (2000) Spin-current interaction with a monodomain magnetic body: a model study. Phys Rev B 62(1):570
18. Tatara G, Kohno H (2004) Theory of current-driven domain wall motion: spin transfer versus momentum transfer. Phys Rev Lett 92(8):086–601
19. Thiaville A, Nakatani Y, Miltat J, Suzuki Y (2005) Micromagnetic understanding of current-driven domain wall motion in patterned nanowires. EPL (Europhys Lett) 69(6):990
20. Tsoi M, Jansen A, Bass J, Chiang WC, Seck M, Tsoi V, Wyder P (1998) Excitation of a magnetic multilayer by an electric current. Phys Rev Lett 80(19):4281
21. Yu S, Wong HS (2011) Compact modeling of conducting-bridge random-access memory (cbram). Electron Dev IEEE Trans 58(5):1352–1360

http://www.springer.com/978-1-4939-0550-8

Design Exploration of Emerging Nano-scale Non-volatile Memory

Yu, H.; Wang, Y.

2014, X, 192 p. 377 illus., Hardcover
ISBN: 978-1-4939-0550-8