Spin Transfer Torque (STT)-MRAM–Based Runtime Reconfiguration FPGA Circuit

WEISHENG ZHAO, ERIC BELHAIRE, and CLAUCHE CHAPPERT
IEF, Univ. Paris-Sud and CNRS
and
PASCAL MAZOYER
STMicroelectronics

As the minimum fabrication technology of CMOS transistor shrink down to 90nm or below, the high standby power has become one of the major critical issues for the SRAM-based FPGA circuit due to the increasing leakage currents in the configuration memory. The integration of MRAM in FPGA instead of SRAM is one of the most promising solutions to overcome this issue, because its non-volatility and high write/read speed allow to power down completely the logic blocks in "idle" states in the FPGA circuit. MRAM-based FPGA promises as well as some advanced reconfiguration methods such as runtime reconfiguration and multicontext configuration. However, the conventional MRAM technology based on field-induced magnetic switching (FIMS) writing approach consumes very high power, large circuit surface and produces high disturbance between memory cells. These drawbacks prevent FIMS-MRAM’s further development in memory and logic circuit. Spin transfer torque (STT)-based MRAM is then evaluated to address these issues, some design techniques and novel computing architecture for FPGA logic circuits based on STT-MRAM technology are presented in this article. By using STMicroelectronics CMOS 90nm technology and a STT-MTJ spice model, some chip characteristic results as the programming latency and power have been calculated and simulated to demonstrate the expected performance of STT-MRAM based FPGA logic circuits.

Categories and Subject Descriptors: B.7.1
General Terms: Design, Reliability, Security, Experimentation, Performance

Additional Key Words and Phrases: MRAM, runtime reconfiguration (RTR), multicontext, low power, spin transfer torque (STT), System on Chip (SOC), FPGA, nonvolatile, architecture

The work and results reported were obtained with research funding from the European Community under the sixth Framework, Contract Number 510993: MAGLOG. The views expressed are solely those of the authors, and the other contractors and/or the European community cannot be held liable for any use that may be made of the information contained herein.

Author’s addresses: W. Zhao, E. Belhaire, and C. Chappert, IEF, Univ. Paris-Sud, UMR 8622, Orsay, F-91405, CNRS, Orsay, F-91405; email: weisheng.zhao@u-psud.fr; P. Mazoyer, STMicroelectronics, 850 Rue Jean Monnet Crolles, Grenoble 38026, France.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2009 ACM 1539-9087/2009/10-Art14 $10.00
DOI 10.1145/1596543.1596548 http://doi.acm.org/10.1145/1596543.1596548

1. INTRODUCTION

SRAM-based Field Programmable Gate Arrays (FPGAs) logic circuits have been the object of intense development in the last 20 years [Brown 1992; Chow et al. 1999]. The reconfigurability property provided by this technology compatible with standard CMOS process have made this technology quite attractive for numerous applications; however, SRAM is volatile, which means that all the functions have to be preprogrammed at each power-up, and external nonvolatile PROM memory must be integrated with the chip either in the same package or at the Printed Circuit Board (PCB) level. This leads to the loss of progress in long-running computations when a power failure occurs, increases the start-up latency, the PCB die area, and in particular, the high standby power due to the leakage currents, which has become the major issue of the current FPGA as the minimum dimensions of the CMOS transistor shrink down to 90nm or below [Kim 2003; Curd 2006]. Internal Flash memory [Actel 2007] is sometimes used to address these issues by replacing both the external memory and SRAM. However, it has some drawbacks such as slow reprogramming and sensing, limited number of writing cycles (up to $10^6$), which limit its lifetime and the reconfiguration speed in FPGA [ITRS 2007].

Magnetic RAM (MRAM) has been rapidly evaluated [Redon et al. 2005; Joeng et al. 2005; Gallagher et al. 2006; Hosomi et al. 2005; Hayakawa et al. 2005] as one of the most promising Spintronics applications. It represents advantages such as nonvolatility, high write/read speed, limitless endurance, radiation-hardness, and the like. The use of MgO barrier in the Magnetic Tunnel Junction (MTJ) (see Figure 1) [Yuasa et al. 2004], basic memory cell of MRAM, improves significantly the tunnel magnetoresistance (TMR) effect (1) and its reading performance. Furthermore, the development of novel writing approach based on Spin Transfer Torque (STT) [Sun 2006] promises to greatly reduce the power and die area and improve the writing selectivity. The excellent writing/reading, power, and area performance of STT-MRAM makes it one of the best nonvolatile memory candidates [Kawahara et al. 2007].

$$TMR = \frac{R_{AP} - R_P}{R_P}$$

MRAM-based FPGA logic circuits were proposed by Black et al. [2000] and Zhao et al. [2006a]. These circuits first benefit from the nonvolatility of MRAM to store the configurations both in LUTs and Interconnects and the high writing speed of MRAM cells can also have a strong impact on the chip architecture. Intermediate data normally stored in D-flip-flops can be stored in MRAM cells, and any block can then be safely powered off, which allows the MRAM-based FPGA hard to noise or power failure. In an SRAM-based FPGA, all SRAM cells...
Fig. 1. Magnetic tunnel junction is composed mainly of four layers: an oxide barrier, such as MgO and AlxOy, a free magnetic layer and a pinned magnetic layer (typically in CoFe alloy), an antiferromagnetic layer (AF1) used to pin the magnetization of the reference via the so-called "exchange bias" phenomenon. The magnetization of the storage layer can be switched by an external magnetic field either parallel or antiparallel to that of the reference layer. When a current flows across the MTJ, a change in resistance R is observed between these two magnetic configurations (typically \( \Delta R/R \) of the order of 40% for AlxOy and 230% for MgO-based MTJ). This change of resistance is named tunnel magnetoresistance (TMR) (1).

must be initialized and kept with the configuration information during computing; besides, most of them are in "idle" state. In MRAM-based FPGA, only the part required to process data is active, the rest is able to be powered off completely. Its high reading speed allows FPGA to retrieve the data in about 200ps. This new computing architecture could greatly relax the high standby power issue in the majority of applications, particularly for the mobile solutions. Moreover, the low cell area and 3D stack structure of MRAM allows easy realization of the multicontext architecture without much additional surface for the MTJs [Wolf et al. 2001].

The applications of MRAM-based FPGA can be extended advantageously to aerospace and military fields thanks to its radiation, power failure hard property, and its limitless switching endurance. In this field, SRAM-based FPGA is difficult to implement, as data security is one of the most important characteristics. However, the high switching power consumption, large switching area, and poor cell selection performance of the conventional MRAM writing approach Field-Induced Magnetic Switching (FIMS) limits the future interest of this technology [Gallagher et al. 2006; Zhao et al. 2009a]. Another switching approach, Thermally Assisted Switching (TAS)-MRAM [Prejbeanu et al. 2004] promises to lower the reconfiguration latency and improve the writing selectivity, but it is limited to reduce the chip area and programming power due to the comparatively high switching current (~4mA) for each LUT and the heating current (~100uA to 1mA) for each bit.

In this article, we investigate how STT-MRAM can be used instead of FIMS or TAS MRAM in magnetic nonvolatile FPGA to circumvent these drawbacks. Special design techniques were used to conceive the logic components in this FPGA circuit. In the next section, we briefly introduce how the STT writing mode works. The third section presents the design of STT-look-up table (LUT)
W. Zhao et al.

Fig. 2. The Spin-MTJ state changes from parallel (P) to antiparallel (AP) if the positive direction current density $I > I_{c+}$, on the contrast, its state will return if the negative direction current density $I > I_{c-}$.

Table I. Comparison of Three MTJ Switching Technologies

<table>
<thead>
<tr>
<th>MTJ Device</th>
<th>Speed</th>
<th>Area</th>
<th>Power</th>
<th>I threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIMS-MTJ</td>
<td>High</td>
<td>Large</td>
<td>Very High</td>
<td>~10mA</td>
</tr>
<tr>
<td>TAS-MTJ</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>~4mA + ~200uA</td>
</tr>
<tr>
<td>STT-MTJ</td>
<td>High</td>
<td>Small</td>
<td>Low</td>
<td>~100uA</td>
</tr>
</tbody>
</table>

structure, STT-programmable interconnection and some advanced computing architectures. Based on the CMOS 90nm design kit [STMicroelectronics 2007] and a STT-MTJ spice model [Zhao et al. 2006b], simulations have been done to calculate and demonstrate the high chip characteristics as low power, small area, and very low reconfiguration latency of STT-LUT. At last, we discuss and conclude.

2. MRAM BASED ON SPIN TRANSFER TORQUE (STT) WRITING APPROACH

2.1 Spin Transfer Torque Mechanism

STT effect was predicted by Slonczewski [1996] and observed by a number of research group [Sun 2006; Hossameddine et al. 2007] in recent years. With this effect, the magnetization direction in the MTJ storage layer can be merely reversed by a spin-polarized current flowing through the junction. This switching mechanism occurs when the current density exceeds a critical density value (see Figure 2). This critical current density has been found lately as low as $8 \times 10^5$ A/cm$^2$ [Hayakawa 2005] in Co$_{40}$Fe$_{40}$B$_{20}$/MgO/Co$_{40}$Fe$_{40}$B$_{20}$ stack structure; as the Spin-MTJ device surface is usually small (e.g., 113nm × 75nm), the critical current is thereby less than a few hundred uA and could be easily generated by a simple minimum-sized CMOS current source. The STT writing approach resolves some major disadvantages of conventional FIMS writing mode, such as high-power dissipation, selection disturbance, and large transistors in the CMOS writing circuit.

Table I exhibits the full comparison between the Spin-MTJ and FIMS-MTJ in the terms of speed, area, power and I threshold Spin-MTJ shows the dominant advantages and will play affirmatively the leading role during the industrialization of MTJ technology.
2.2 STT-MRAM Fabrication

The stable fabrication process is always one of the major issues or obstacles for the nanodevices for mass production. As the MTJ stack has a vertical structure similar to CMOS production with low enough annealing temperatures during the process, it can take place in the backend of the CMOS process (see Figure 3). One advantage of this fabrication is that the MTJ integration does not take much die surface except for the sensing CMOS circuits and the contacts necessary to connect the MTJs with MOS transistors. The Spin-MTJ switching current is so low that it is kept below any electromigration limitation, and any thin metal layer (e.g., Metal4) is sufficient in the CMOS technology. This helps to lower the fabrication cost when compared with the conventional FIMS-MTJ and TAS-MTJ device. The total cost of STT-MRAM–based FPGA could be lower than the SRAM-based FPGA because only two or three additional masks are needed to integrate the STT-MTJ at the backend process [Kawaraha et al. 2007], and the external nonvolatile memory and configuration data transceiver in the PCB card are no longer required. One of the major issues for the Spin-MTJ fabrication is the height of oxide barrier, which should not be too low (e.g., <0.7nm) to exhibit the TMR effect and not too high (e.g., >2.5nm) to keep the low resistance value [Yuasa et al. 2004]. A well-controlled and precise deposit process for the oxide barrier is required to avoid the mismatch variation and ensure the good MTJ sensing performances.

3. STT-MRAM–BASED NONVOLATILE LOOK-UP TABLE (STT-LUT)

3.1 MTJ Write and Read Circuit in STT-LUT

As mentioned earlier, the state switching of STT-MTJ requires a dual directional current source, thereby four NMOS (MN0 through MN3) powered by Vdda and Gnda are used to generate the current passing through the STT-MTJs (see Figure 4). Each time only two of these NMOS are active, a control circuit composed by two NOR logic gates is addressed to define the activation by the control signal “EN” and the direction of current by “input,” respectively. Figure 4 shows an example of 2 inputs STT-LUT with four STT-MTJs, the reconfiguration of this LUT can be operated bit by bit in a series to economize the chip area. In this case, only one current source is required, but each STT-MTJ should be associated with one additional NMOS transistor, which gives it
Fig. 4. Two inputs STT-LUT programming schematics; where there are four NMOS controlled by C0-C3, respectively, to select the STT-MTJs to be programmed, two of NMOS MN0-MN3 can be active each time to generate the bidirectional current.

the corresponding addresses (e.g., C0-C3). Benefiting from the high switching speed of STT effect, lower than 1ns [Devolder et al. 2005], the reconfiguration in series will not slow the speed and it takes only some hundred nanoseconds for the reconfiguration of complex STT-LUT with more than five inputs.

Numerous sense amplifiers (SAs) were proposed to read the state of MTJ by detecting its resistance difference [Black et al. 2000; Durlam et al. 2002]. An SRAM-based sense amplifier (MN4-6 and MP0-1; see Figure 5) is used here, as it is capable to sense a pair of MTJs in different resistance and demonstrates very high reading speed lower than 200ps [Zhao et al. 2006a]. The sensing begins by briefly turning on the switch MN6 with “SEN” to place the amplifier in a metastable state. The MTJ (MTJR: MTJ Reference, respectively) modulates the source of the transistor NMOS MN4 (MN5, respectively) which forms an inverter with the PMOS transistor MP0 (MP1, respectively). The pull-down strength of the inverter is then modulated by the MTJ resistance value, and when MN6 is turned off, the amplifier reaches one of its two stable states depending on the sign of the difference of resistance between MTJ and MTJR. The output of the amplifier then restores a digital level “Out” whose value depends on the bit stored in the MTJ. The high speed of this sense amplifier allows the output to be retrieved rapidly from the MTJs in case of power failure or soft error due to the radiation. The sensing power maybe very low et ignorable because its low frequency and sensing current as low as some uA (2).

\[
P_{\text{dynamics}} = f_{\text{input}} \times \int_{0}^{T} V_{dd} \times I_d (t) \, dt
\]  

(2)
The resistance of MTJR should be in the middle of the P resistance and AP resistance of MTJ to get the largest margin and improve the sensing stability; that means the TMR ratio will be halved in this sensing mechanism. However, the bias voltage dependence is observed for the TMR ratio and is reduced as bias voltage increases [Yuasa et al. 2004]. In our design, the TMR is set to 275% and the maximum bias voltage may be up to 100mv. In this case, the real TMR during sensing is about 250%, and the resistance difference between the P state of MTJ and MTJR should be 125% (3). The higher TMR ratio has
Fig. 6. With the same magnetic process, the different resistance of MTJ depends only on its surface.

Fig. 7. The full schematic of one nonvolatile computing bit includes the switching and data-sensing circuit.

been observed more than 500% [Lee et al. 2007], which will further improve the sensing performance of the sense amplifier.

\[ R_{MTJR} = R_{MTJ} \times \left( 1 + \frac{TMR(V)}{2} \right) \]  

(3)

As the MTJs and MTJR should be implemented with the same magnetic process to save the fabrication cost; we cannot change the height of oxide barrier (tox) to determine the resistance difference between the MTJ and MTJR. Therefore, we design the MTJ and MTJR in different dimensions (4) [Brinkman et al. 1970]. Figure 6 exhibits the dimension of MTJ0 and MTJR, the TMR ratio between them is 125%. As MTJR is always in a parallel state, no write current source is required for this reference resistance. Noted that MN7 and MN8 are necessary to balance the resistance with the context selecting NMOS transistors, they should be always active during data sensing (see Figure 5).

\[ R_{MTJ} = \frac{tox}{223.76 \times \varphi^{1/2} \times \text{surface}} \times \exp(1.025 \times tox \times \varphi^{1/2}) \]  

(4)

Combining the write and read circuit, the full schematic of one nonvolatile computing bit can be designed (see Figure 7). During the reconfiguration, C0 is active but the A, B selecting signals and the “EN_bar” control signal should be inactive. In this case, the current source will write the new data as the state of MTJ. During the computing, C0 is inactive but A, B and R/W are active. The current passing through the MTJ will be referenced with the current passing through the MTJR and sense amplifier outputs the result in logic value.

It is important to mention that the power source of the sensing and switching current source Vdd_{Logic}, Vdda should be different to optimize area performance. In Equation (5), \( \mu_n \) is the electron surface mobility, \( C_{ox} \) is the gate...
Table II. Transistor Number Comparison of Four LUT

<table>
<thead>
<tr>
<th>Configuration Memory</th>
<th>Two Input</th>
<th>Three Input</th>
<th>Four Input</th>
<th>Five Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRAM</td>
<td>34</td>
<td>66</td>
<td>132</td>
<td>262</td>
</tr>
<tr>
<td>FIMS-MRAM*</td>
<td>1,046</td>
<td>1,874</td>
<td>3,383</td>
<td>6,848</td>
</tr>
<tr>
<td>TAS-MRAM*</td>
<td>122</td>
<td>154</td>
<td>271</td>
<td>352</td>
</tr>
<tr>
<td>STT-MRAM</td>
<td>35</td>
<td>48</td>
<td>75</td>
<td>126</td>
</tr>
</tbody>
</table>

*Large transistor is considered as some minimum transistors in parallel.

Fig. 8. (a) all the logic blocks are busy (black), (b) almost half of them are in the standby state (white), (c) most of the logic blocks are in the standby state.

oxide capacitance, and \( V_{Nth} \) is the threshold voltage of NMOS transistor, they depend only on the CMOS process. Thereby improving the voltage supply of the NMOS transistor allows it to reduce its width keeping the same drain current; however, the dynamic power will be increased proportionally as a trade-off (2). We should thus choose the Vdda to reach a compromise between the power and area performance. In our design based on 90nm technology, Vdd_Logic and Vdda are set to 1.2V and 5V, respectively. The die area of the switching circuit can be thus economized up to about 80%.

\[
I_D = \frac{k_n}{2} (2(V_{GS} - V_{Nth}) \times V_{DS} - V_{DS}^2)
\]

\[
k_n = \mu n \cdot C_{ox} \cdot \frac{W_n}{L}
\]

3.2 Area, Power, and Speed Performance of STT-LUT

As all the bits of STT-LUT share the same current source and sense amplifier, its number of transistors can be reduced greatly (see Figures 4 and 5). Table II demonstrates that STT-MRAM–based 5-inputs LUT can economize about 50% the number of transistors than the SRAM-based LUT; we compare also with the proposals of the FIMS-MRAM–based LUT [Zhao et al. 2006a] and the TAS-MRAM–based LUT [Zhao et al. 2007; 2009b]. Unlike the use of some very large transistors in the current source for the FIMS and TAS switching approaches, all the NMOS transistors in the current source for STT switching mode could be at the minimum width (e.g., 012um for CMOS 90nm technology).

The total power of this STT-LUT is very low. First, zero standby consumption benefiting from the non-volatility and high data-sensing speed, the logic blocks in standby state can be powered down completely (see Figure 8). For
low-power CMOS 90nm technology, the leakage power for each bit SRAM is down to about 9.8pW. If we assume that the four SRAM cells in the 4-input LUT are in the “idle” state for only 1 second, the total power will be dominated by the standby power as high as 313.6pJ. This decreasing of standby power is very important for LUT applications, as it operates with stored data and there are always some logic blocks in the “idle” state to wait the active command for most of the applications. Second, the low switching current (∼200uA) significantly reduces the dynamic reconfiguration power. Based on the STMicroelectronics CMOS 90nm low power design kit [STM 2007] and a STT-MTJ macromodel (see Figure 9), the energy simulated and calculated for the reconfiguration of a 5-input STT-LUT is as low as 35.2pJ. Table III exhibits the estimated dynamic power for the reconfiguration of LUT based on different configuration memories.

Although the dynamic power of STT-LUT is still much higher than conventional LUT due to the write current, it is comparatively ignored to the decreasing of standby power. Multi-Vt and multi-Vdd SRAM have been proposed to store the configurations in FPGA circuit because the higher Vt or lower Vdd of MOS transistor could significantly reduce leakage currents, thereby economizing the standby power. However, these technologies bring much higher extra

Table III. Reconfiguration Power Comparison of Four LUT

<table>
<thead>
<tr>
<th>Configuration Memory</th>
<th>Two Input</th>
<th>Three Input</th>
<th>Four Input</th>
<th>Five Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRAM</td>
<td>0.44 pJ</td>
<td>0.88 pJ</td>
<td>1.76 pJ</td>
<td>3.52 pJ</td>
</tr>
<tr>
<td>FIMS-MRAM</td>
<td>1,000 pJ</td>
<td>1,800 pJ</td>
<td>3,400 pJ</td>
<td>6,600 pJ</td>
</tr>
</tbody>
</table>

Fig. 9. DC simulation of STT-MTJ model, Ic+ is about 133.2uA, and Ic− is about 217.6uA.
Table IV. Reconfiguration Latency Comparison of Four LUT Configuration Memory

<table>
<thead>
<tr>
<th>Configuration Memory</th>
<th>Two Input</th>
<th>Three Input</th>
<th>Four Input</th>
<th>Five Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRAM</td>
<td>some ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIMS-MRAM</td>
<td>∼2ns</td>
<td>∼4ns</td>
<td>∼8ns</td>
<td>∼16ns</td>
</tr>
<tr>
<td>TAS-MRAM</td>
<td>∼25ns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STT-MRAM</td>
<td>∼4ns</td>
<td>∼8ns</td>
<td>∼16ns</td>
<td>∼32ns</td>
</tr>
</tbody>
</table>

Fig. 10. NMOS MNB should be added to isolate the computing logic value “Out” from “Qm” during reconfiguration. Qm will change during the data write and become indeterminate as “SEN” equal to “1.”

cost due to the additional masks, and it can not overcome the other drawbacks of conventional SRAM-based FPGA [Kaptanoglu Sinan Atel Corporation 2007].

The data write in this series determines that the reconfiguration latency of STT-LUT increases linearly with the number of logic bits; it is much lower than conventional LUT, which requires to be updated with the configuration data from the external memory. Benefiting from the parallel reconfiguration architecture, the TAS-LUT demonstrates the same delay no matter how many inputs. For the LUT having more than five inputs, the reconfiguration of TAS-LUT is a little more rapid than STT-LUT. The comparison in term of area, power, and speed shows that STT-LUT is able to completely overcome the drawbacks of FIMS-LUT and TAS-LUT, and it promises to replace the conventional LUT for all the applications (see Table IV).

3.3 STT-LUT Runtime Reconfiguration (RTR) Architecture and Multicontext Configuration

The high write speed and nonvolatility of STT-MTJ allows the LUT to be reconfigured at runtime. In this case, an additional NMOS transistor MNB is required to avoid the disturbance from data programming and protect the data “Out” from “Qm” for the runtime computing function (see Figure 10). Before the data write in the STT-MTJs, MNB should be closed; it should be opened to finish the reconfiguration and to enable the LUT operate with the new function.

Figure 11 exhibits an example of run-time reconfiguration for five-input STT-LUT, the reconfiguration flow begins at the closing of MNB by “R/W”; in this case, the computing data is isolated from the write/read circuit. In the following,
Fig. 11. Spice simulation of the reconfiguration for five inputs STT-LUT, about 40ns is required to reconfigure the configuration of STT-LUT, but the reconfiguration delay from one function to the other one is lower than 100ps thanks to the runtime reconfiguration (RTR).

MNA is closed and the current source is enabled by “EN” (see Figures 4 and 5). Because there are 32 bits in this LUT and the data write operates in series, the LUT reconfiguration time is about 32ns. After that, we send a short pulse with 2ns duration to transistor MN6 through the control signal “SEN,” Qm will be set to a new data. At last, MNB is active to update the data for the logic computing in as low as about 100ps, logic delay from MNB and the LUT begins to compute in the new function.

The very low reconfiguration delay and low data write power allow the function of logic blocks to be reconfigured frequently and, thereby, greatly improve their operating efficiency. This method promises to accelerate the computing speed because there is less interconnection delay between the different logic blocks. It may also economize the chip physical area for the same computing task by reducing the number of logic blocks, as shown in Figure 12. For
example, in the SRAM-based FPGA, three LUTs should be used for three functions operated in series, by using STT-LUT, only one LUT is required and it will be reconfigured with the three functions in continuous time. During the function transition, all the internal data can be stored nonvolatilely in the STT-MRAM–based flip-flop (STT-FF) [Zhao et al. 2008; Sakimura et al. 2008].

The stack structure of STT-MTJ and low switching current permit the integration of multicontext operation into one LUT without much additional area. Figure 13 demonstrates the programming circuit for a two-input STT-LUT with two contexts. For the computing structure, one additional NMOS in parallel with MNA (see Figure 6) is required for the second context. This multicontext architecture is very promising to relax the dynamic power consumption for the FPGA operating with limited configurations. In this case, no MTJ programming is executed during the reconfiguration, and the function of LUT can be changed by simply switching the contexts. The initialization of this multicontext architecture is simple as well as, same as the reconfiguration shown in Figure 11, but multiply the data write period with the number of contexts.

3.4 STT-MRAM–Based Programmable Interconnection (STT-PI)

Programmable interconnection based on six switches is another configuration component for the FPGA circuit (see Figure 14(a)), which plays an important role in the whole die area [Brown et al. 1992]. An FIMS-MRAM–based programmable interconnection has been proposed [Zhao et al. 2006a], which uses a pair of MTJs in a differential state, one sense amplifier and one current source for each switch. Its area is much larger than the conventional interconnection in FPGA due to the large current source for FIMS technology and impossibility to program multibits. By using STT-MRAM, we first benefit the minimizing die area thanks to the low switching current, and, moreover, one current source is used to program multswitch (e.g., 32) or multiinterconnection in series (see
Fig. 14. (a) conventional programmable interconnection, there is a SRAM cell per switch, (b) STT-MRAM–based programmable interconnection, there are two MTJs and one sense amplifier (SA) per switch.

Fig. 15. Mask of STT-LUT–based FPGA prototype.

Figures 4 and 13). However, unlike the STT-LUT with only one output, multiple switches in the interconnection may be active at the same time. Therefore, one 6-transistor–based sense amplifier (Figure 10) with a pair of MTJs in differential state is still required per switch, and the die area of STT-PI is nearly the same with the conventional one (see Figure 14(b)). STT-PI operates in the similar way as STT-LUT in the aspects of reconfiguration, data retrieving, multicontext configuration, and the like. etc.

4. CONCLUSION AND PERSPECTIVES

STT-MRAM based–FPGA logic circuits are presented in this article, which can perform runtime reconfiguration, multicontext configuration, and “instant-on” start-up with low power dissipation, small area, and at high speed. This non-volatile FPGA logic circuit has great potential to replace all the types of current FPGA circuits in the high-performance computing and those embedded in mobile equipments powered by the battery. They could be advantageously used in the field of aviation and space where the data hardness to radiation is one of the most important considerations. The prototype (see Figure 15) of this
hybrid magnetic CMOS logic circuit is under development and realization in our laboratory in collaboration with Belfield University (magnetic process) and STMicroelectronics (CMOS 90nm low-power technology).

REFERENCES


GALLAGHER, W. J. AND PARKIN, S. S. P. 2006. Development of the magnetic tunnel junction MRAM at IBM: From first junctions to a 16mb MRAM demonstrator chip. IBM J. Res. and Dev. 5–23.


Received June 2008; revised December 2008; accepted February 2009