

Received 9 March 2020; accepted 6 April 2020. Date of publication 10 April 2020; date of current version 7 August 2020.

The review of this article was arranged by Editor L. Lukasiak.

Digital Object Identifier 10.1109/JEDS.2020.2987084

Compact FeFET Circuit Building Blocks for Fast and Efficient Nonvolatile Logic-in-Memory

EVELYN T. BREYER^{1,2}, HALID MULAOSMANOVIC², JENS TROMMER² (Member, IEEE), THOMAS MELDE³, STEFAN DÜNKEL³, MARTIN TRENTZSCH³, SVEN BEYER³, STEFAN SLESAZECK², AND THOMAS MIKOLAJICK¹ (Senior Member, IEEE)

¹ Chair of Nanoelectronic Materials, Technische Universität Dresden, 01069 Dresden, Germany

² NaMLab gGmbH, 01187 Dresden, Germany

³ GLOBALFOUNDRIES Fab1 LLC & Company KG, 01109 Dresden, Germany

CORRESPONDING AUTHOR: E. T. BREYER (e-mail: evelyn_tina.breyer@tu-dresden.de)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 780302, in part by the German Bundesministerium für Wirtschaft (BMWi), in part by the State of Saxony in the frame of the "Important Project of Common European Interest (IPCEI)," and in part by the Open Access Funding by the Publication Fund of the TU Dresden.

ABSTRACT Due to their CMOS compatibility, hafnium oxide based ferroelectric field-effect transistors (FeFET) gained remarkable attention recently, not only in the context of nonvolatile memory applications but also for being an auspicious candidate for novel combined memory and logic applications. In addition to bringing nonvolatility into existing logic circuits (Memory-in-Logic), FeFETs promise to guide the way to compact Logic-in-Memory solutions, where logic computations are examined in memory arrays or array-like structures. To increase the area-efficiency of such circuits, a dense integration of FeFETs and standard FETs is essential. In this paper, we show that the ultra-dense co-integration of FeFETs and nFETs (28nm HKMG) with shared active area does not alter the FeFET's switching behavior, nor does it affect the baseline CMOS. Based on this, we propose the integration of a FeFET-based, 2-input look-up table (memory) directly into a 4-to-1 multiplexer (logic), which is utilized directly in a 2TNOR memory array or stand-alone circuit. The latter one dramatically reduces the transistor count by at least 33% compared to similar FeFET-based circuits. By storing values of the look-up table in a nonvolatile manner, no energy is consumed during standby mode, which enables normally-off computing. To take another step towards novel Logic-in-Memory designs, we experimentally demonstrate a very compact in-array 2T half adder and simulate an array-like 14T full adder, which exploit the advantages of the array arrangement: easy write procedure and a very compact, robust design. The proposed circuits exhibit energy-efficiency in the (sub)fJ-range and operation speeds of 1GHz.

INDEX TERMS Adder, ferroelectric FET (FeFET), hafnium oxide (HfO₂), logic-in-memory (LiM), look-up table (LUT), memory array, multiplexer (MUX), ultra-dense integration.

I. INTRODUCTION

Recently, as the amount of data to be processed is constantly increasing, not only the way of storing these data, but also their processing time and the associated power consumption become critically important for efficient computing. This is aggravated by the physical separation of memory and logic units in today's computer architectures, resulting in the necessity for temporary storage solutions. Ferroelectric field effect transistors (FeFETs) seem to be promising candidates to counter these challenges, as they unite the

nonvolatile (NV) storage ability of memory devices with the three-terminal logic operation of field effect transistors. Besides the area of neuromorphic computing [1]–[3], merging memory and logic in so-called Logic-in-Memory (LiM) applications is a promising way to bridge the gap between both [4], [5]. As shown in [6]–[8], especially 2TNOR (Fig. 1a) and AND memory arrays (Fig. 1b), comprising parallel connected FeFETs, can be utilized for LiM operations. Additionally, hafnium oxide based FeFETs were proven compatible with standard CMOS processing [9]. Aggressive

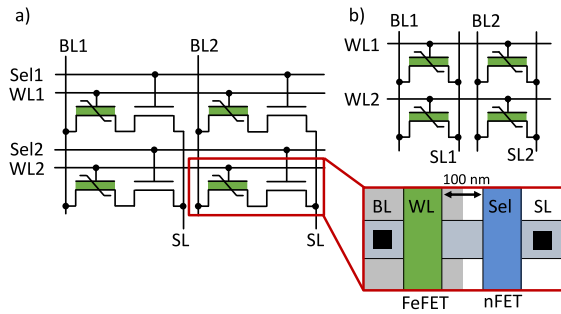


FIGURE 1. Section of the (a) 2TNOR memory array with selector n-FET (Sel-nFET) and n-type FeFET (layout as inset) and (b) the FeFET-only AND array. The complete arrays are made up of 90 and 63 memory cells, respectively.

scaling and 3D stacking led to an improved integration density [10], [11].

Regardless of whether conducting logic operations in memory arrays, or integrating NV elements into logic circuits, a very close-by integration of nonvolatile and volatile devices has to be ensured [6], [12]. Therefore, we first investigate FeFETs having the same size, incorporated into 2TNOR and AND memory arrays, to show that the ultra-dense co-integration of FeFETs and standard FETs does not alter the switching characteristics of the FeFETs, thus paving a way to spatially closely integrated LiM (Section III-A). Then, the switching behavior of FeFETs at low write voltages up to $\pm 2\text{V}$ is examined to exclude a switching of the polarization state by the applied logic readout voltage (Section III-B).

Based on this, we propose a 4-to-1 multiplexer (MUX) with an integrated nonvolatile 2-input look-up table (LUT) (Section IV). Similar to [13]–[15], the LUT values are stored in a nonvolatile manner, thus keeping their state while consuming no energy during standby. Advantageously, the proposed design saves at least 33% of transistor count compared to other FeFET-based designs [13], features a low energy consumption during write/readout, and can also be implemented in a 2TNOR memory array [6]. To further expand the range of LiM and exploit the aforementioned memory arrays, we introduce a very compact, array-like 1-bit half adder and 1-bit full adder (Section V), which calculate carry and sum separately. One input value of these circuits is stored in a nonvolatile manner within the FeFET, which is especially useful if the addition uses one constant summand for a certain amount of time. But different from [5], where the FeFET is integrated into an existing logic full-adder, the write operation of FeFET in the memory array is more straightforward, as known memory array write schemes can be used. These circuits can be utilized for computations directly within arrays as well as array-like stand-alone applications. Transient logic measurements and SPICE simulations prove the functionality of the proposed concepts.

II. DEVICE STRUCTURE AND METHODS

The field effect transistors used in this work were fabricated using the 28nm high-k metal gate process (HKMG) [16].

The n-type FeFETs comprise a $\text{TiN}/\text{Si}:\text{HfO}_2(8\text{nm})/\text{SiON}/\text{Si}$ gate stack. Two device sizes (length L , width W) were used: FeFETs of $L = W = 100\text{ nm}$, arranged in a 2TNOR (Fig. 1a) and AND (Fig. 1b) memory array [6], [12], and FeFETs of $L = W = 500\text{ nm}$, arranged in an AND memory array. The AND and the 2TNOR arrays comprise 63 and 90 FeFETs, respectively. FeFETs in the 2TNOR array are connected to a selector nFET in series.

Electrical measurements were conducted using Keithley 4225PMU pulsed measurement units (PMUs), which were connected to a Keithley 4200-SCS Semiconductor Analyzer. To program (PRG) or erase (ERS) the FeFET, i.e., to set it into the low or high threshold voltage state, voltage pulses of different width t_p and height V_p were applied at its gate. Standard values were $V_p = 4.5\text{V}$ (PRG)/ $V_p = -3.5\text{V}$ (ERS) at $t_p = 10\mu\text{s}$. To determine the threshold voltage V_t [17], the readout operation comprised a fast gate voltage (V_g) sweep, at which the drain current (I_d) was measured. In the logic applications (look-up table with integrated multiplexer, array-based half adder), FeFETs were written by a block erase of the whole array, followed by programming selected cells. To avoid unintentional write of other devices in the array, unselected cells were sufficiently inhibited. Logic readout pulses were of $50\mu\text{s}$ or $500\mu\text{s}$ length.

Transient SPICE simulations were carried out using the Cadence SPECTRE simulator. A behavioral ferroelectric capacitor model, based on the time-dependent Preisach model of hysteresis [18], [19], connected in series to a standard FET from the GLOBALFOUNDRIES 28 nm PDK, constituted the FeFET model.

III. CHARACTERIZATION OF FEFETS

A. SWITCHING BEHAVIOR AND ULTRA-DENSE CO-INTEGRATION

One vital quest to accelerate the development of FeFET-based LiM circuits is the ultra-dense co-integration of logic FETs and FeFETs, while ensuring that this kind of procedure does not alter the switching behavior of the FeFET itself. Therefore, we compare the switching behavior of FeFETs ($L = W = 100\text{ nm}$) co-integrated with MOSFETs into 2TNOR arrays (Fig. 1a) with those in standard AND arrays (Fig. 1b).

The cumulative distribution functions (CDF) of the programmed state (Fig. 2a, $V_p = 4.5\text{V}$, $t_p = 10\mu\text{s}$) and the erased state (Fig. 2b, $V_p = -3.5\text{V}$, $t_p = 10\mu\text{s}$) of FeFETs in the AND and 2TNOR memory array are in good agreement. When observing the voltage difference between the low and the high V_t state, the formation of a distinct memory window of 0.88 V (AND) and 0.94 V (2TNOR) is seen, similar to what was observed in [20]. This confirms the FeFETs' functionality as non-volatile storage devices. Although the CDFs exhibit a steep slope in both cases, corresponding to devices of very similar threshold voltage, they also contain tails. The small channel measures of 100nm lead to a stronger impact of single grains within the ferroelectric layer, which in turn results in an increased V_t variability as shown in [10]. The

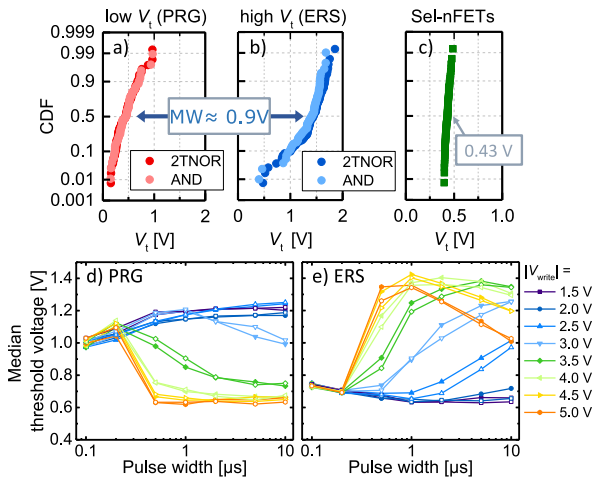


FIGURE 2. Comparison of ultra-densely co-integrated nFeFETs from a 2TNOR array with stand-alone nFeFETs of the same dimensions ($L = W = 100$ nm) from an AND array. Their cumulative distribution function (CDF) of the (a) low V_t (PRG) state and (b) high V_t (ERS) state are well matching. (c) CDF of the selector devices of the 2TNOR array. (d) and (e) V_t vs. t_p graphs reveal a very similar switching behavior of FeFETs in the 2TNOR (closed symbols) and AND memory array (open symbols).

selector devices’ (Sel-nFeFETs) CDF is centered at 0.43 V and shows a very steep distribution with no significant tails (Fig. 2c). As such, it is very suitable for logic applications in the lower V_{dd} range. For the FeFETs, in turn, a readout voltage of 0.9 V seems to be a suitable choice to distinguish between the high and low V_t state. This value also constitutes a lower limit for V_{dd} during logic operation.

To evaluate the switching behavior of FeFETs integrated into the AND and the 2TNOR array, the median threshold voltage of the CDF after the write operation was determined for different write pulse heights (from ± 1.5 V to ± 5 V) and widths (from 0.1 μ s to 10 μ s), as shown in Figs. 2d,e. Write voltages as low as ± 3.5 V are sufficient to erase/ program the FeFET, however, the pulse width has to be set to at least 10 μ s in this case (tradeoff between write pulse height and width [21]). With increasing V_p and t_p , the threshold voltage of the erased state (high V_t , Fig. 2e) decreases again after reaching a maximum value. This effect is attributed to the injection of holes from the Si substrate into the gate stack of the FeFET, which overlays the effect of polarization charge, and as a result, diminishes the threshold voltage [22], [23]. Larger FeFETs with a channel size of $L = W = 500$ nm, integrated into an AND memory array, show the same behavior (Fig. 3a and Fig. 3b). The curves of the 2TNOR and the AND memory array match very well (Fig. 2d and 2e), suggesting that the ultra-dense co-integration of FeFET and nFET within the 2TNOR memory array has no significant influence onto the FeFETs’ switching behavior. Thus, novel Logic-in-Memory concepts (described in Sections IV and V) are enabled.

B. FEFET CHARACTERIZATION FOR LOGIC-IN-MEMORY

If memory arrays are used in LiM applications, storing a result of a logic operation within the memory cell might

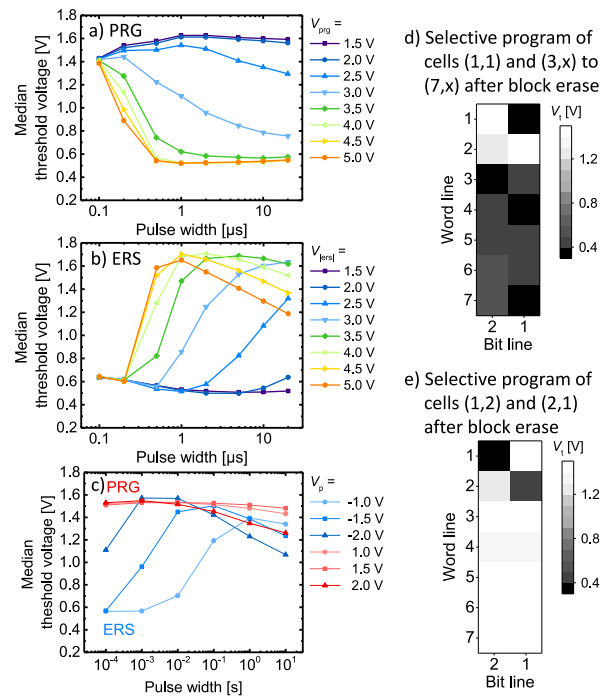


FIGURE 3. Threshold voltage V_t of nFeFETs in the AND-array (channel size: $L = W = 500$ nm). V_t is set by applying different program pulse heights (V_p) and widths (t_p). The programmed state (a) and erased state (b) show a gradual switching behavior with respect to t_p . (c) For V_p below $|2V|$, t_p extends up to seconds to switch the polarization state of the FeFET. In (d) and (e), the V_t of FeFETs connected to bit line 1 to 2 and word line 1 to 7 are shown after a block erase (white cells) and subsequent selective program (black cells).

be of interest. In order to develop compact designs, lower write voltages are desirable to relax the constraints when mixing thin and thick oxide devices dedicated to logic or write operation. However, the program voltages should not be as low as the logic voltage level (V_g at the FeFET) in order not to overwrite the FeFET during logic operation. For this, FeFETs ($L = W = 500$ nm) in an AND memory array, as they are also used in the half adder of Section V, are investigated. Fig. 3c depicts the threshold voltage of a FeFET after the write operation for $V_p = \pm(1.0; 1.5; 2.0)$ V. t_p varied from 100 μ s to 10s. The low V_t state (PRG) is not yet completely programmed even after a write time of 10s. The high V_t state (ERS), on the other hand, is written at lower pulse widths (e.g., in 0.01s at $V_p = -1.5$ V). Thus, purposely writing FeFET memory cells at such low voltages is not reasonable due to the long write times. In turn, read disturbs are very unlikely during normal operation, since logic voltages, like 0.9V as used in this work, are not sufficient to rewrite the FeFET in the targeted time frame (GHz regime). However, it has to be taken into consideration that due to the accumulative switching behavior of FeFETs [24] a polarization reversal might occur even for smaller V_p if several voltage pulses of the same type were applied subsequently.

During write operation of FeFETs in memory arrays, care has to be taken about a sufficient inhibit of other

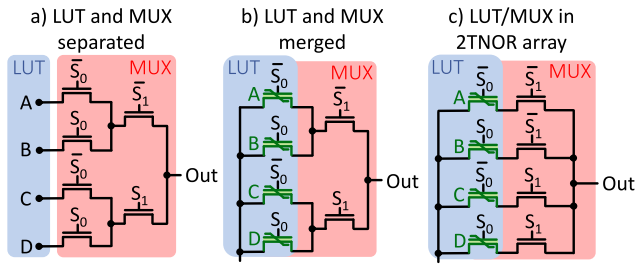


FIGURE 4. (a) Standard 2-input LUT with adjoined multiplexer, (b) proposed merging of LUT and multiplexer (LUTMUX), and (c) LUTMUX integrated into a 2TNOR memory array.

memory cells. In order to prove a successful inhibit operation, Figs. 3d and 3e show an extract of an AND array. Here, a block erase ($V_p = -3.5V$) was followed by a selective programming of cells ($V_p = 4.5V$). During the programming, unselected cells were inhibited with a voltage $V_{inh} = 3V$, applied to the source line and bit line. This voltage is sufficiently high in order to inhibit cells but is low enough not to overwrite programmed cells (white and black cells in Figs. 3d, e, respectively). The shown patterns at BL1/BL2 and WL1/WL2 are utilized for the half adder of Section V-D in case its input A is logic 0 (Fig. 3d) or logic 1 (Fig. 3e).

IV. MULTIPLEXER WITH INTEGRATED LOOK-UP TABLE

A. BASIC CONCEPT

In standard FPGA cells, look-up tables (LUT) serve as a first stage connected to a multiplexer as the second stage, which accomplishes the routing of the LUT signals (Fig. 4a). The LUT can be constituted of nonvolatile memories (e.g., FeFETs [13], see also Table I) or volatile storage devices (e.g., SRAM cells [25]). This widely used concept is receptive to improvement in the following ways, when integrating the LUT stage directly into first selection stage of the MUX (Fig. 4b):

- signal delay between LUT and MUX can be decreased
- overall size can be reduced by saving one transistor stage
- signal routing does not require transmission gates, as pass gates (only nFETs) are sufficient

Within this merged “LUTMUX” structure (in Fig. 4b similar to the standard structure, in Fig. 4c using a 2TNOR array), the values of the LUT are stored in the FeFETs in a non-volatile manner, where the high V_t state represents a logic “0”, while the low V_t state corresponds to a logic “1”. Thus, the logic output function of the look-up table is determined by setting the polarization state of the FeFETs. The same FeFETs also act as logic transistors, and thus, replace the first stage of selector FETs of the MUX that are controlled by the selecting signal S_0 . As proposed previously [4], the FeFET executes a logic AND operation between its internally stored value and the applied selecting signal S_0 . The other stages of the MUX consist of standard FETs with the applied selecting signals S_1, S_2 etc.

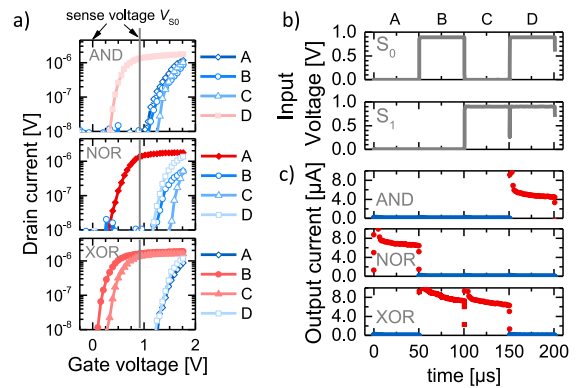


FIGURE 5. (a) Measured $I_d - V_g$ curves of four nFeFETs in the 2TNOR array, which store the values A, B, C, and D of the proposed look-up table. Transfer curves of the four involved FeFETs are shown for the AND, NOR, and XOR case. (b) Logic operation of the LUTMUX. By applying the selecting input signals S_0 and S_1 at the multiplexer, the value of a specific storage cell (A, B, C, D) is read out. This value is represented by the measured drain current (c).

B. MEASUREMENT AND SIMULATION

In this paper, we concentrate on a 4-bit LUTMUX (stored values A, B, C, and D, Fig. 4c), which is capable of all 16 two-in-one-out logic functions by purposely setting the values of A, B, C, and D [6]. Interestingly, the structure of a 2TNOR FeFET memory array (Fig. 4c) naturally maps a 4-bit LUTMUX. Due to the basic structure of the 2TNOR array, it comprises two additional selector FETs (signal S_1) compared to the structure shown in Fig. 4b. By evaluating the FeFETs’ $I_d - V_g$ curves, suitable values for the voltages V_{si} , corresponding to the selector signal values of $S_i = 0$ and $S_i = 1$, are found to be $0V (= V_{ss})$ and $0.9V (= V_{dd})$, respectively. At the latter, low and high V_t state of the FeFET can be clearly distinguished when evaluating the drain current I_d (vertical line in Fig. 5a). Due to the size of the FeFETs ($L = W = 100$ nm), a certain variation of the threshold voltages (around 200mV) is observable, which is also reflected in the slight variation of the high output current levels (logical “1”) from $4 \mu A$ to $8 \mu A$. Thus, a realistic situation is emulated. In comparison to the on/off ratio of a ferroelectric transistor, which is $>10^3$, this variation has a negligible influence on the logic functionality. Fig. 5c depicts the transient output current measurements of the LUTMUX in three different configurations with applied selector signals S_0 and S_1 (Fig. 5b): the LUT stores an AND, NOR, and XOR function. Thus, every FeFET is set into the high and low V_t state at least once.

To complement the experimental results, we conducted behavioral SPICE circuit simulations on a 4-bit LUTMUX. As in the measurement, the supply voltage was set to 0.9V. In the simulation, RC delays of the measurement setup can be avoided, and intrinsic circuit delays during logic operation and the FeFET switching delay during the write of the FeFET can be assessed. Moreover, when eliminating the undesirable noise of the measurement setup in the circuit simulation, the drain current (output current) spans over

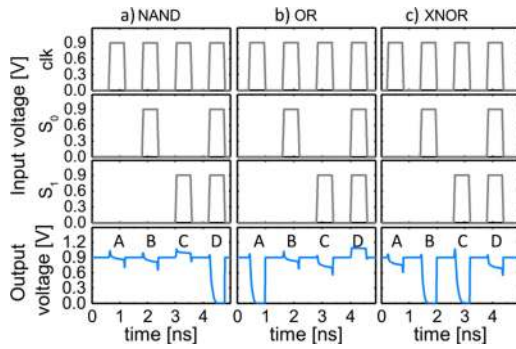


FIGURE 6. Results of the SPICE simulation of the proposed LUTMUX. A clock frequency of 1 GHz (denoted by “clk”) was used for the dynamic readout of the LUT, which emulated a logic (a) NAND, (b) OR, and (c) XNOR gate.

TABLE 1. Comparison of the proposed 4-to-1 LUTMUX to literature.

Device	No. of devices ¹	Write energy	Read energy
<i>This work</i> (4:1) ^{2,3}	4 FeFETs + 2 (or 4) nFETs	1.36 fJ (~4V, all 4 cells) ³	0.27 fJ (1 bit) ³
<i>FeFET</i> (4:1) [13] ³	4 FeFETs + 10 nFETs	53.88 fJ (up to 4V, per 1 bit)	NA
<i>SHE-MTJ</i> (4:1) [14] ³	4 MTJs + 6 nFETs	~10s of pJ	NA
<i>Memristor</i> (4:1) [15] ³	4 Memristors + 4 nFETs	26.78 pJ	2.405 fJ
<i>CMOS</i> (4:1) [25] ³	≥ 22 nFETs + 8 pFETs	1.54 fJ [13]	NA

¹ only LUT plus MUX ² measured ³ results from simulation

six orders of magnitude (from 10^{-12} A to 10^{-6} A), resulting in an on-to-off-ratio of 10^6 . During the logic readout, a clocked pull-up pFET, situated between V_{dd} and the output node, allows to transform the FeFET current into a voltage signal. Fig. 6 depicts the clock signal (clk) and the selector input voltages (V_{s0} , V_{s1}), as well as the resulting output voltages for NAND, OR, and XNOR operation of the LUT. The maximum operation frequency is at least 1 GHz, while the largest dynamic energy consumption is as low as 1.36 fJ for the write operation (all cells) and 0.271 fJ for read operation (one cell). Table 1 shows a comparison with other non-volatile implementations of LUT with separate MUX.

C. DISCUSSION

Compared to other pass gate implementations of multiplexers, the input (i.e., gate) and output (i.e., drain/source) signals are decoupled in the proposed structure. Thus, the full V_{dd} voltage swing is exploited. When the LUTMUX serves tasks within an FPGA, the proposed structure (Fig. 4b) has the advantage over other FeFET-based nonvolatile LUTs (e.g., [13]) to reduce the footprint drastically. Skipping the first stage of the multiplexer, corresponding to, e.g., four (eight, sixteen) FETs in case of a 2-input (3-input, 4-input) look-up table saves 40% (36.4%, 34.8%) of transistors. Finally, since the number of MUX transistors T corresponds to a geometric row, where N is the number of LUT-inputs

TABLE 2. Inputs (voltages) and output (current) of the array half adder (black) and the array full adder (black and blue).

Input stored in FeFETs		Input applied at FeFET gate		Input applied at sel-FET		Output (current)	
A	!A	B	!B	C_i	! C_i	C_o	S
0	1	0	1	0	1	0	0
0	1	1	0	0	1	0	1
1	0	0	1	0	1	0	1
1	0	1	0	0	1	1	0
0	1	0	1	1	0	0	1
0	1	1	0	1	0	1	0
1	0	0	1	1	0	1	0
1	0	1	0	1	0	1	1

and n is the number of transistors in the first stage (= number of LUT cells), it follows for the limit of infinitely large LUTs:

$$T = n \cdot \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} \left(\frac{1}{2}\right)^k = 2 \cdot n \quad (1)$$

Thus, the overall transistor count savings are the number of transistors skipped by the proposed design (n) divided by the number of transistors in the conventional design ($2n$ for the MUX tree, n for the LUT): $n/(2n + n) = 33.3\%$. In the current design, this not yet corresponds to the overall area savings, as FeFETs usually have a larger footprint than conventional FETs.

V. ARRAY BASED ADDERS

A half adder consists of two simple logic gates to conduct an addition operation between its two inputs A and B. An XOR gate calculates the sum ($S = A \oplus B$) of both inputs, and an AND gate computes the overrun carry ($C_o = A \cdot B$) (see Table 2). To extend the half adder to a full adder, another input, the carry-in C_i , has to be processed to get $C_o = A \cdot B + B \cdot C_i + A \cdot C_i$ and $S = A \oplus B \oplus C_i$ (blue section of Table 2). Usually, the corresponding circuits are built of CMOS transistors in the volatile case [26], or of a hybrid of CMOS transistors and nonvolatile devices like magnetic tunnel junctions (MTJ) [27], [28] or FeFETs [5]. In the latter case, the nonvolatile device replaces the devices, which process a specific input (e.g., A), and stores the value of A in a nonvolatile manner. The second input B and C_i are still applied to the corresponding other logic transistors. This kind of nonvolatile circuit structure is closely connected to the volatile circuit structure, although it combines its logic ability with nonvolatility (Memory-in-Logic). Further, the write operation of the nonvolatile devices within the logic circuit might raise the risk of either undesirably high voltage drops over and/or currents through the standard logic transistors. In-array approaches were pursued in memristor-crossbar memory arrays, requiring several operation steps [29]. To circumvent the mentioned issues, we propose adder structures, which exploit the advantages of standard FeFET based memory arrays with respect to writing procedures of these nonvolatile devices. Additionally, they

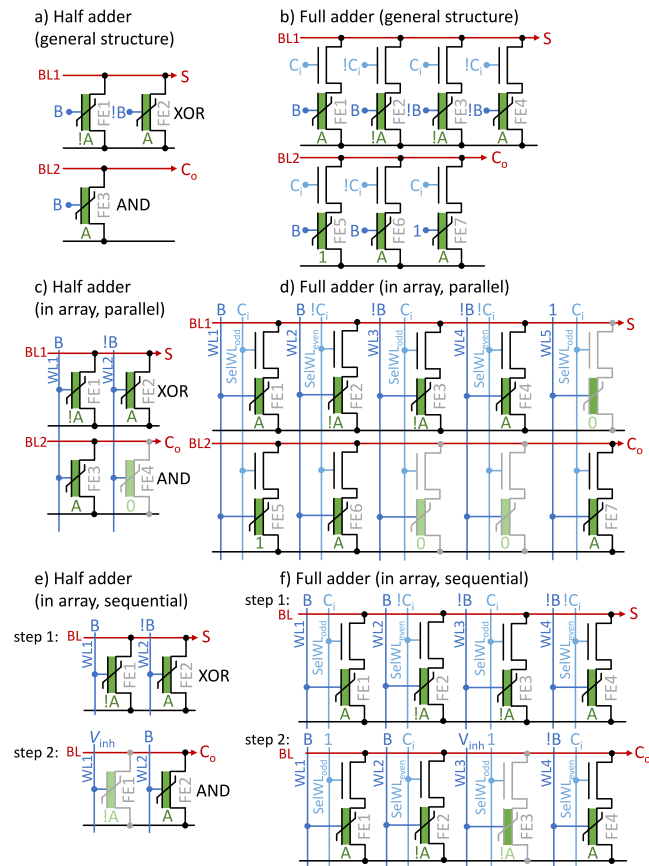


FIGURE 7. FeFET based adders. (a) and (b) show the basic structure of the array based half adder and full adder (“stand-alone”). BL and WL are bit and word lines. FE1 to FE7 denote FeFET1 to FeFET7. (c) and (d) depict the half adder (c) and the full adder (d), mapped into the AND or 2TNOR array, respectively, in parallel operation mode. Unused and therefore inhibited cells are grayed out. C_0 and S are calculated simultaneously. In the sequential mode of (e) the half adder and (f) the full adder, C_0 and S are computed sequentially at the same bit line. Step1 results in the output of the sum of inputs A and B , while step2 corresponds to outputting the carry of A plus B .

conduct binary summations between values stored internally in the FeFETs and values applied externally at the gate terminal of the same FeFETs. Such an approach is particularly interesting for applications, in which one operand is a constant to be adjusted at times, as it is the case in digital filter applications.

A. BASIC CONCEPT

The basic *half adder* structure (Fig. 7a) comprises three FeFETs and works as follows. To compute the sum bit $S = A \oplus B$, two FeFETs (FeFET1 and FeFET2) are connected at their drains. First, input A is stored as a complementary pair (\bar{A} and A) in FeFET1 and FeFET2, respectively. Then, the complementary inputs B and \bar{B} are applied at the gate terminal of FeFET1 and FeFET2, respectively, which act together as a logic XOR gate between inputs A and B , similar to the approach described in [7]. FeFET3 is chosen to compute the carry bit $C_0 (=A \cdot B)$. For this, the

first input value (A) is stored in the polarization state of FeFET3 and the second summand (B) is applied to the gate terminal of FeFET3. FeFET3 itself acts as a sequential logic AND gate between both inputs as described in [4]. S and C_0 are reflected by the level of the output drain current (low current = logical 0, high current = logical 1) at their respective bit line. The proposed structure perfectly maps to an AND memory array.

The operation principle of the *full adder* (Fig. 7b) is similar to the half adder, i.e., one input value (A , or its inverse \bar{A}) is stored in the FeFETs, while the second input value (B , or \bar{B}) is applied to gate of the same FeFETs. One FeFET itself acts as logic AND gate between inputs A/\bar{A} and B/\bar{B} as proposed in [4]. Additionally, a serially connected nFET processes the carry input C_i/\bar{C}_i , conducting another logic AND operation. In that way, at the upper bit line in Fig. 7b provides the output sum S :

$$S = A \oplus B \oplus C = (A \oplus B) \cdot \bar{C}_i + (A \oplus B) \cdot C_i \quad (2)$$

The lower bit line provides the output carry C_0 :

$$C_0 = A \cdot B + B \cdot C_i + A \cdot C_i = (A + B) \cdot C_i + A \cdot B \cdot \bar{C}_i \quad (3)$$

In summary, seven FeFETs and seven FETs constitute the nonvolatile, array-based full adder. The proposed structure can be appropriately mapped into a 2TNOR memory array.

Thus, with the proposed full and half adders, very compact, nonvolatile adders are constructed, where the sum bit and the carry bit are naturally decoupled, promising similar delay times for computing S and C_0 . In general, both structures feature an stand-alone operation, e.g., to replace traditional adder circuits with fixed input connections, but also an operation directly in a memory array as required for LiM applications (i.e., “in-array”).

B. PARALLEL IN-ARRAY OPERATION

In order to map the proposed adders into regular memory arrays, the gate terminals of the FeFETs (and nFETs in case of the full adder) are connected to the word lines (and select lines). One possible operation mode is the parallel fetch, where S and C_0 are computed simultaneously at two separate bit lines.

As the base frame of the *half adder* (Fig. 7c) we select two FeFETs with their drains connected to one bit line (FeFET1 at BL1/WL1, and FeFET2 at BL1/WL2) and two FeFETs of another bit line, but same word lines (FeFET3 at BL2/WL1, and FeFET4 at BL2/WL2). First, operand A/\bar{A} is written into the FeFETs by a bulk erase and ensuing selective programming operation. The second summand, B (at WL1), and its inverse, \bar{B} (at WL2), are applied to the gate of the FeFETs for readout. FeFET1 and FeFET2 are selected to calculate the sum bit $S (=A \oplus B)$, reflected by the level of the output current at bit line 1. While FeFET3 computes $C_0 = A \cdot B$, FeFET4 stays in the erased state, corresponding to an internal value of logic “0” (grayed out in Fig. 7c), in

order not to influence the output current at bit line 2. This implies that it is not needed for the calculation of C_o but is still existent in the structure, which is mapped into the regular AND array. Thus, bit line 2 carries the output current corresponding to C_o . All FeFETs not involved in the half adder circuit are either not contributing to the output current, since they are connected to separate bit lines (e.g., bit line 3), or are sufficiently inhibited during readout in case they are connected to the same bit lines as FeFET1 to FeFET4.

The in-array operation of the *full adder* is very similar to the half adder as depicted in Fig. 7d. The full adder structure comprises more branches, whose inputs B/ \bar{B} not necessarily overlap for C_o and S. That is, the number of sufficiently inhibited unused branches is increased (grayed out in Fig. 7d) in order to avoid unintentional readouts of cells that are connected to the same word line (WL3, WL4, WL5).

The parallel readout therefore requires more transistors: 2 FeFETs and 2 FETs for the half adder, and 10 FeFETs and 10 FETs for the full adder, but it is twice as fast as the sequential fetch and ensures simultaneous computation of C_o and S. As the input arrangement does not change during readout, an in-array and stand-alone operation is possible.

C. SEQUENTIAL IN-ARRAY OPERATION

In the sequential operation mode of the adders (half adder in Fig. 7e, full adder in Fig. 7f), only FeFETs connected to one single bit line are used, and C_o and S are fetched sequentially at this bit line. The circuit speed is slowed down by a factor of two compared to the parallel fetch, with the advantage of using only few FeFETs for a complementary, TCAM-like storage of input A (stored in the FeFETs) [7], [8]. Hence, in case of the half adder (Fig. 7e), inputs \bar{A} and A are stored in FeFET1 (BL1/WL1) and FeFET2 (BL1/WL2), respectively. The full adder (Fig. 7f) stores \bar{A} and A as a double complementary pair in FeFET1 to FeFET4.

The sequential fetch is divided into two steps. In step 1, the sum S is calculated by applying B and \bar{B} at the gates of the FeFETs as shown in Figs. 7e,f, and additionally applying C_i and \bar{C}_i at the gate of the selector nFETs as depicted in Fig. 7f (only full adder). In step 2, C_o is computed by re-applying the inputs B and \bar{B} at the gates of the FeFETs, as well as C_i and \bar{C}_i at the selector nFETs (only full adder), in order to comply with the basic adder structures of Figs. 7a,b. Similar to the parallel operation, any uninvolvement FeFET is sufficiently inhibited in order not to contribute to the output current at the bit line. During both steps of sequential operation, the internal polarization states of the FeFETs (corresponding to A or \bar{A}) remain unchanged.

Compared to the parallel fetch, the execution speed halved due to two operation steps required to calculate C_o and S. However, the readout disturb of other bit lines in the in-array operation, as a potential issue in the parallel fetch, is avoided, as only one bit line is read out. As the inputs B (and C_i) vary for both computation steps, a stand-alone operation

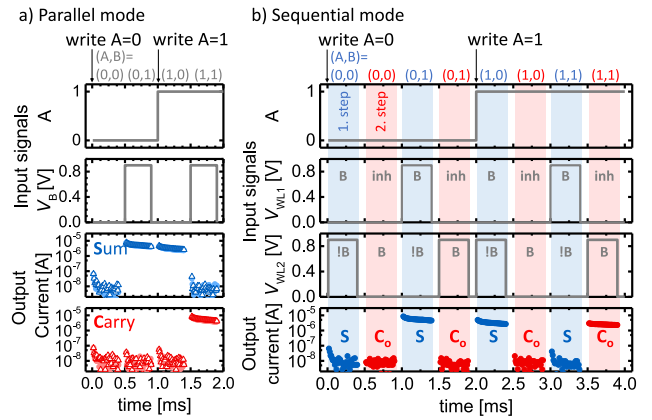


FIGURE 8. Measurement results of the half adder in parallel (a) and sequential operation mode (b), directly integrated in a 9x7 cell mini array. Input A (polarization state of the FeFET), input B (voltage applied to word lines 1 and/or 2) and the output current, corresponding to the sum S (blue) and carry C_o (red), are shown. Open and closed symbols of the output current in (a) correspond to the best-case scenario (all noninvolved FeFETs erased) and worst case scenario (all noninvolved FeFETs programmed), respectively.

would require a steady re-routing of inputs and periphery, therefore an in-array operation should be preferred.

D. MEASUREMENT AND SIMULATION

The measurement results of both methods (parallel and sequential fetch) of the half adder reveal a successful in-array operation at $V_{dd} = 0.9V$ (Figs. 8a,b), which also corresponds to the optimum readout voltage of the utilized FeFETs (compare to Fig. 5a). The array featured 9 word and 7 bit lines. For the parallel fetch, a worst-case scenario (except FeFET1 to FeFET4 all other FeFETs are programmed) and a best-case scenario (except FeFET1 to FeFET4 all other FeFETs are erased) was examined, exhibiting no significant differences in the output current. This confirms a successful inhibit of cells, which are not involved in the half adder operation. The I_{on}/I_{off} ratio was in the order of 10^3 , while the off current was around the lower detection limit of the setup, limited by the pulsed measurement method. Nonetheless, a clear distinction of output states is possible. Speed is limited by measurement setup restrictions.

In order to verify the functionality and to examine metrics of the full adder (structure as Fig. 7b), SPICE simulations were conducted on FeFETs in the 28nm HKMG technology (Fig. 9). The supply voltage was set to 0.95V. A clocked readout scheme with a pull-up transistor was used in order to determine the output voltage. At operation frequencies as high as 1 GHz, the full adder still operates reliably and energy-efficient (average read energy: 1.42 fJ, average write energy: 15.9 fJ, see Table 3).

E. DISCUSSION

The proposed array-like adder structure is more compact than CMOS-like approaches, e.g., as described in [5]. However,

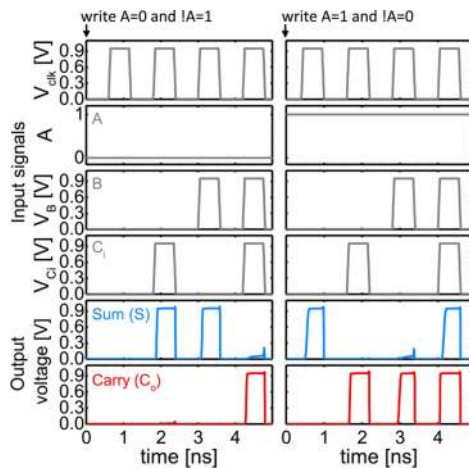


FIGURE 9. SPICE simulation results of the stand-alone full adder (parallel operation mode). V_{dd} was set to 0.95V. Input A/ \bar{A} corresponds to the FeFETs' polarization state, input B/ \bar{B} is applied to the FeFETs' gate terminal and input C_i is applied at the selecting transistor. Clock frequency was 1 GHz.

TABLE 3. Comparison of the proposed 1-bit full adder to literature.

Device	No. of devices	Write energy (or power)	Read Energy (or power)
<i>This work</i> (FeFET, 28nm)	7 FETs + 7 FeFETs + 6 FETs ^{1,2}	15.9 fJ (average)	avg.: 1.42 fJ min.: 0.42 fJ
FeFET [5] (22nm)	28 FETs + 4 FeFETs ^{2/} 17 FETs + 3 FeFETs ³	not stated not stated	0.54 fJ (0.27 μ W) 0.42 fJ (0.21 μ W)
STT+SHE [27] (28nm)	52 FETs + 4 MTJs ²	80 fJ (for 1 bit)	1.23 fJ
SHE-MTJ [28]	23 FETs + 3 SHEs ³	\sim 148 μ W	40.68 fJ (\sim 13.6 μ W)
Memristor [30]	9 memristors ⁴	accumulated: 6.4 nJ	
CMOS [26]	> 14 FETs	n/a	60.6 μ W ²

¹ includes 2 pull-up FETs for C_o and S, 4 FETs (inverter for C_o and S)

² completely separate computation of sum and carry

³ calculation of sum uses results from calculation of carry

⁴ requires 43 sequential steps

when compared to conventional 1bit-1cell memory storage, as it ensues in AND and 2TNOR memory arrays, the transistor count and area to store one bit are at least doubled (half adder) or even quadrupled (full adder). Generally, the way of data storage and readout is more similar to a TCAM array as described in [8], where data bits are stored complementarily. Nonetheless, the compact and robust array-like design has advantages: easy write (e.g., bulk erase of the whole array, followed by a programming of selected cells) with known write schemes [31], the possibility to map it into conventional memory arrays (AND/2TNOR), and a separate calculation of carry C_o and sum S. Additionally, the 1-bit full adder (Fig. 7b) can be cascaded in order to build n-bit full adders. For this, the output carries C_o / \bar{C}_o are directly connected to the C_i / \bar{C}_i inputs, without influencing the FeFET input A (stored state).

VI. CONCLUSION

An ultra-dense co-integration with logic transistors does not degrade the switching behavior of the FeFET, thus it constitutes a promising method for nonvolatile memory and logic co-integration. Moreover, logic voltages around 1V do not disturb the stored state of the FeFET when operating faster than kHz. Building upon this, we suggest suitable Logic-in-Memory operations for two memory array types (AND, 2TNOR), that can be conducted directly within the memory array. The proposed multiplexer with integrated look-up table (LUTMUX), 1-bit half and full adders work in a memory array environment as well as in stand-alone applications. The array-like circuit structure enables the use of known array write/ operation schemes. The stand-alone applications feature a competitive transistor count compared to recent research – at least 33.3% less in case of the proposed LUTMUX, 2T in case of the sequential half adder and 8T in case of the sequential full adder. Due to a very simple structure, the energy consumption during read out was no more than 1.42fJ (full adder) and 0.27fJ (LUTMUX) for a logic operation, while operation speed is at least 1GHz.

ACKNOWLEDGMENT

The authors gratefully acknowledge support by GLOBALFOUNDRIES Fab1 LLC & Co. KG, Dresden, Germany.

REFERENCES

- [1] H. Mulaosmanovic *et al.*, "Novel ferroelectric FET based synapse for neuromorphic systems," in *Proc. Symp. VLSI Technol. (VLSIT)*, 2017, pp. T176–T177.
- [2] H. Mulaosmanovic, E. Chicca, M. Bertele, T. Mikolajick, and S. Slesazeck, "Mimicking biological neurons with a nanoscale ferroelectric transistor," *Nanoscale*, vol. 10, no. 46, pp. 21755–21763, 2018.
- [3] Y. Long *et al.*, "A ferroelectric FET based processing-in-memory architecture for DNN acceleration," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 5, no. 2, pp. 113–122, Dec. 2019.
- [4] E. T. Breyer, H. Mulaosmanovic, T. Mikolajick, and S. Slesazeck, "Reconfigurable NAND/NOR logic gates in 28 nm HKMG and 22 nm FD-SOI FeFET technology," in *IEDM Tech. Dig.*, pp. 1–4, 2017.
- [5] X. Yin *et al.*, "Exploiting ferroelectric FETs for low-power non-volatile logic-in-memory circuits," in *Proc. 35th Int. Conf. Comput. Aided Design (ICCAD)*, 2016, pp. 1–8.
- [6] E. T. Breyer *et al.*, "Ultra-dense co-integration of FeFETs and CMOS logic enabling very-fine grained logic-in-memory," in *Proc. 49th Eur. Solid-State Device Res. Conf. (ESSDERC)*, 2019, pp. 118–121.
- [7] E. T. Breyer, H. Mulaosmanovic, S. Slesazeck, and T. Mikolajick, "Demonstration of versatile nonvolatile logic gates in 28nm HKMG FeFET technology," in *IEEE International Symposium on Circuits and Systems*, 2018, pp. 1–5.
- [8] I. Bayram and Y. Chen, "NV-TCAM: Alternative interests and practices in NVM designs," in *Proc. IEEE Non Volatile Memory Syst. Appl. Symp.*, 2014, pp. 1–6.
- [9] J. Müller *et al.*, "Ferroelectric hafnium oxide: A CMOS-compatible and highly scalable approach to future ferroelectric memories," in *IEDM Tech. Dig.*, 2013, pp. 10.8.1–10.8.4.
- [10] S. Dünkel *et al.*, "A FeFET based super-low-power ultra-fast embedded NVM technology for 22 nm FDSOI and beyond," in *IEDM Tech. Dig.*, 2017, pp. 19.7.1–19.7.4.
- [11] J. Van Houdt, "3D memories and ferroelectrics," in *Proc. IEEE Int. Memory Workshop (IMW)*, 2017, pp. 1–3.
- [12] S. Beyer *et al.*, "Embedded FeFETs as a low power and non-volatile beyond-von-Neumann memory solution," in *Proc. IEEE Nonvolatile Memory Technol. Symp.*, 2018, pp. 28–29.

- [13] X. Chen, K. Ni, M. T. Niemier, Y. Han, S. Datta, and X. S. Hu, "Power and area efficient FPGA building blocks based on ferroelectric FETs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 5, pp. 1780–1793, May 2019.
- [14] J. Yang *et al.*, "Exploiting spin-orbit torque devices as reconfigurable logic for circuit obfuscation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 38, no. 1, pp. 57–69, Jan. 2019.
- [15] T. N. Kumar, H. A. F. Almurib and F. Lombardi, "A novel design of a memristor-based look-up table (LUT) for FPGA," in *Proc. IEEE Asia-Pac. Conf. Circuits Syst. (APCCAS)*, 2014, pp. 703–706.
- [16] M. Trentzsch *et al.*, "A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs," in *IEDM Tech. Dig.*, 2016, pp. 11.5.1–11.5.4.
- [17] *Procedure for Measuring N-Channel MOSFET Hot-Carrier-Induced Degradation Under DC Stress*, JEDEC Standard JESD28-A, 2001.
- [18] B. Jiang, P. Zurcher, R. E. Jones, S. J. Gillespie, and J. C. Lee, "Computationally efficient ferroelectric capacitor model for circuit simulation," in *Proc. Symp. VLSI Technol. (VLSIT)*, 1997, pp. 141–142.
- [19] K. Dragosits, R. Hagenbeck, and S. Selberherr, "Transient simulation of ferroelectric hysteresis," in *Proc. Int. Conf. Model. Simul. Microsyst. Techn.*, 2000, pp. 433–436.
- [20] J. Müller, T. S. Böske, U. Schröder, R. Hoffmann, T. Mikolajick, and L. Frey, "Nanosecond polarization switching and long retention in a Novel MFIS-FET based on ferroelectric HfO₂," *IEEE Electron Device Lett.*, vol. 33, no. 2, pp. 185–187, Feb. 2012.
- [21] H. Mulaosmanovic *et al.*, "Switching kinetics in nanoscale hafnium oxide based ferroelectric field-effect transistors," *ACS Appl. Mater. Interfaces*, vol. 9, no. 4, pp. 3792–3798, Jan. 2017.
- [22] W.-T. Lu *et al.*, "The characteristics of hole trapping in HfO₂/SiO₂ gate dielectrics with TiN gate electrode," *Appl. Phys. Lett.*, vol. 85, no. 16, pp. 3525–3527, 2014.
- [23] E. Yurchuk *et al.*, "Charge-trapping phenomena in HfO₂-based FeFET-type nonvolatile memories," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3501–3507, Sep. 2016.
- [24] H. Mulaosmanovic, T. Mikolajick, and S. Slesazeck, "Accumulative polarization reversal in nanoscale ferroelectric transistor," *ACS Appl. Mater. Interfaces*, vol. 10, no. 28, pp. 23997–24002, 2018.
- [25] P. Girard, O. Héron, S. Pravossoudovitch, and M. Renovell, "Delay fault testing of look-up tables in SRAM-based FPGAs," *J. Electron. Test.*, vol. 21, pp. 43–55, Feb. 2005.
- [26] M. Aguirre-Hernandez and M. Linares-Aranda, "CMOS full-adders for energy-efficient arithmetic applications," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 4, pp. 718–721, Apr. 2011.
- [27] E. Deng, Z. Wang, J. Klein, G. Prenat, B. Diény, and W. Zhao, "High-frequency low-power magnetic full-adder based on magnetic tunnel junction with spin-hall assistance," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–4, Nov. 2015.
- [28] A. Roohi, R. Zand, D. Fan, and R. F. DeMara, "Voltage-based concatenatable full adder using spin hall effect switching," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 12, pp. 2134–2138, Dec. 2017.
- [29] S. Hu *et al.*, "Reconfigurable Boolean logic in memristive crossbar: The principle and implementation," *IEEE Electron Device Lett.*, vol. 40, no. 2, pp. 200–203, Feb. 2019.
- [30] F. M. Puglisi, L. Pacchioni, N. Zagni, and P. Pavan, "Energy-efficient logic-in-memory I-bit full adder enabled by a physics-based RRAM compact model," in *Proc. 48th Eur. Solid-State Device Res. Conf.*, 2018, pp. 50–53.
- [31] S. Müller *et al.*, "Correlation between the macroscopic ferroelectric material properties of Si:HfO₂ and the statistics of 28 nm FeFET memory arrays," *Ferroelectrics*, vol. 497, no. 1, pp. 42–51, 2016.