Evaluating Ternary Adders using a hybrid Memristor / CMOS approach

Dietmar Fey
Friedrich-Alexander-University Erlangen-Nürnberg (FAU)
Chair Computer Architecture
Martensstr. 3, 91058 Erlangen, Germany
dietmar.fey@fau.de

Abstract

This paper investigates the potentials of using a hybrid memristor CMOS technology, called MeMOS, for the realisation of ternary adders. Ternary adders exploit the qualitative advantage of multi-value storage capability of memristors compared to conventional CMOS flip-flops storing only binary values in one cell. Furthermore they carry out an addition in $0(1)$ and are therefore considered. The MeMOS approach is compared to a CMOS solution for the ternary adders using multi value memristors as registers concerning the achievable latency and the energy consumption. It is shown that using the TEAM, VTEAM model and a model considering commercially available memristors from known the approach of using CMOS ternary adders using memristors as multi-value register memory is to prefer. MeMOS circuits have advantages for a static operation mode, i.e. if they are operated after a reset.

1. Introduction

New memory technologies, like e.g. memristors, offer the possibility of so-called in-memory computing concepts. Characteristic for such concepts is a paradigm change in the way how data processing is done. Whereas in a pure von-Neumann computer data is brought to the instructions in a computing architecture that uses in-memory computing the instructions are brought to the data. Concrete does that mean the memory storage element is not purely used for storage, moreover the storage is an integral part of the processing.

Using memristors as storage elements in an in-memory computing architecture means that logical resistor networks are built up and the resistor values can be changed to new ones according to the result of a processing step in which the old resistor value was included.

One of the first concepts in this sense in the world of memristive computing was the so-called IMPLY logic [1], [9]. Two parallel connected memristors got as input voltage levels that are either shortly before a switching level of the memristor or are above that limit. In dependence of the content of the memristor pair, which is the input of the function, a certain resistor network is built up. Depending on the adjusted resistances in that resistor network the state of one memristor, operating as the output, can be changed or not if a voltage above the threshold limit is applied. By that a kind of operation mode a quasi inherent implementation of the Boolean implication operator is realised.

First made proposals with that IMPLY logic include basic Boolean functions like NOT, NAND or NOR. However, the disadvantage is that the realisation of such elementary Boolean functions are mapped onto several subsequent executions of IMPLY operations what causes a higher latency. This problem does not occur in a pure CMOS logic in which operations like NAND or NOR are quasi inherent to that processing which is based on complementary switching of PMOS and NMOS transistors. This means that the principal advantage of in-memory computing concepts over CMOS like less energy consumption and possibly less latency, which is given due to the fact that storing and processing takes place in the same device, is lost.

New expanded proposals like so-called MAGIC [12] and MAD gates [5] or also the concepts presented in [3] and of MeMOS [11], the last one stands for combining of memristors and CMOS, overcome the mentioned disadvantage and allow to carry out basic Boolean operators in one elementary step. This paper is focused on using the MeMOS concept since the requirements to the control path for the memristors seems to be less complex and therefore one can expect even a short- or mid-term solution concerning the realisation of real circuits compared to the other solutions. The control path in MeMOS is more or less reduced to CMOS inverter gates what is much easier compared to the more difficult driver circuitry that MAD gates require. Of course, the price one has to pay for the easier realisation of MeMOS is a higher number of memristors as needed for MAD gates and MAGIC.

All of the three concepts mentioned above allow to do the next necessary step in digital memristive computing for the community, namely to move away from simple Boolean gates towards more sophisticated memristive arithmetic circuits. According to the conviction of the authors this step is absolutely necessary if memristors will have a chance to be used as digital logic any time in the future in order to exploit its principal benefits like less energy consumption and may be also less latency.

In addition, we would like to focus that the success of digital memristive computing requires also to exploit features offered by new memory technologies like memristors which are not possible with pure CMOS storing elements. This concerns e.g. the multi-value storing capabilities of some NVMs in general, and in particular of memristors. We favour in this presented work a ternary number representations realised in memristors.
as another focus aside the MeMOS concept.

In this sense we would like to investigate if memristor technology has the potential to realise ternary computers in the future. The reason why we favour ternary arithmetic than binary is the advantage of ternary structures to carry out an addition in a constant number of steps independent of the operand’s word length, i.e. in $O(1)$. On the contrary a binary adder works best at $\log(N)$ by an area increase that is limited to $(N \log(N) / \log_2(N))$, if $N$ is the number of used bits.

In order to prove the potential and to quantify what has to be done in the future to bring digital memristive computing to success we made a comparative investigation concerning required processing time and energy consumption for ternary adders realised in three different technologies. (i) A pure CMOS solution of the adders using two flip-flops to store one operand, i.e. a ternary digit; (ii) A non in-memory computing approach using memristors, i.e. memristor are just used as multi-value memory in a digital CMOS circuit; (iii) A MeMOS solution for those adders.

In order to obtain the energy and latency values for memristor based adder architectures, i.e. (i) and (ii), we used an own simulation environment written in C++ that solves the memristor behaviour by a discrete solution of the differential equations for different models proposed in literature, namely the TEAM [8] and the VTEAM [6] model. The difference between TEAM and VTEAM model is that the last one considers explicitly a threshold at which the I/U curve for a memristor starts to show a slope. Furthermore, we consider a statistical model for one of the first commercially available memristors, namely the memristors from Knowm [10].

The remainder of the paper is organized as follows. Chapter 2 presents an introduction in ternary arithmetic and explains the specific adders we selected for this investigation. Next chapter briefly introduces the MeMOS concept and shows the MeMOS solution for our preferred ternary adder. Chapter 4 introduces the simulation system we have developed for the evaluation of MeMOS circuits. Chapter 5 continues with an evaluation for the ternary adders in MeMOS and discusses the pro and cons. Finally we end the paper with a conclusion.

### 2. Ternary adders

We have investigated two adders operating on ternary numbers which are processed by binary logic for a realisation in MeMOS technology. The two ternary adders are operating both to a base $r = 2$. One adder, denoted in the following as base$_2$step$_3$, requires three discrete time steps for the processing, but the price is that this adder shows a more complex logic per steps compared to the other adder, denoted as base$_2$step$_4$, which requires however 4 discrete time steps. These two adders are more promising than other adders offering also non-binary operand presentation and constant addition time. E.g. there exist adders, that show only two discrete steps but require a much more complex logic what is caused by the fact that these adders are using operands to a base $r \geq 2$. Details concerning this computer arithmetic stuff can be found in [4], [2].

Both adders are using for their operands a so-called signed digit (SD) number representation, i.e. they allow not only positive but also negative digits. Concretely, a so-called balanced ternary number representation to base $r = 2$ comes into use. That means, generally for each digit $sd_i$ of a SD number $sd$ holds $sd_i \in \{-r+1, \ldots, -r\}$ and for the case $r = 2$, $sd_i \in \{-1, 0, 1\}$. The value of $sd$, having a digit length $N$, is then calculated according to (1).

$$sd = \sum_{i=0}^{N} sd_i \cdot r^i$$

Due to the signed digits we receive redundant representations, i.e. two different SD numbers possess the same value, e.g. $10\overline{1}_3 = (7)_2 = 10\text{I}_2$, $\overline{I}_2 = -1$.

As next we explain briefly the working principle of the two adders. We will start with the adder base$_2$step$_4$ which carries out an addition in four steps independent of the word length.

For the binary processing of the ternary SD number we have to use a coding for each trit. The coding we are using is a so-called (negative, positive) coding, abbreviated as (n,p)-coding, shown in Table 1. The (n,p)-coding has the advantage that the inverse $-sd_i$ of a digit $sd$ can be easily formed by exchanging the negative and the positive part of $sd$.

| Ternary SD digit (trit) | digital 2-bit coding (neg pos) |
|------------------------|-------------------------------|
| -1                     | 1 0                           |
| 0                      | 0 0                           |
| 1                      | 0 1                           |

Table 1: Binary coding of the trits used for the digital processing. The first bit is interpreted as positive, the second one as negative.

The base$_2$step$_4$ adder consists of two subsequent connected subadders, which expect one operand as binary input and the other one as an SD number. This makes the gate logic more easier. An example for an addition that is carried out with such a subadder is shown in Table 2.

Assuming the (n,p) coding we will get the following Boolean equations, (2) to (5), for the base$_2$step$_4$ adder that have to be applied to calculate the intermediate binary values $c^+_i$, $z^-_i$, $s^+_i$, and $s^-_i$ in each digit position of the adder.

$$c^+_i = x^+_i \lor y_i \land \overline{x^-_i}$$

(2)

$$z^-_i = (x^+_i \lor x^-_i) \oplus y_i$$

(3)

$$s^+_i = z^-_i \land c^+_i$$

(4)

$$s^-_i = c^+_{i-1} \land z^-_i$$

(5)
x = (1 -1 1 1)_2 = (7)_{10}
+y = (0 1 1 0)_2 = (6)_{10}

step 1:
-1 0 0 -1 = z

step 2:
1 -1 1 1 -1 = s = 1 \cdot 2^4 - 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 - 1 \cdot 2^0 = 13

Table 2: Addition of SD number x with a binary number y.

Fig. 1 shows a gate layout for one basic building block for this adder for the processing of one digit. It is operated on one digit of SD operand sd1, i.e. on its positive and negative part denoted as x_i_plus, resp. x_i_minus, and on the positive part of a digit of the other SD operand sd2, denoted as y_i.

Figure 1: Gate layout for the processing of one digit in the ternary Radix_2_step_4 adder.

In order to process two SD numbers as inputs the addition process has to be be split in two phases (6), due to the fact that one subadder of the the base_2_step_4 adder has to be fed with one SD input and one binary input. Therefore, first an addition of the first operand sd1 takes place with the positive part sd_+ of the second input operand sd2. Then, a subsequent subtraction with the negative part sd_− has to follow. Both input operands for the subadders are binary.

\[ sd_1 + sd_2 = (sd_1 + sd_+^2) - sd_− \]  

Eq. (7) shows how a subtraction of an SD number, sd, and a binary, B, can be easily reduced to an addition. The positive and the negative part of sd have to be exchanged, then the addition with B can be carried out. The result is an SD number, whose positive and negative part have to be exchanged to form the inverse.

\[ sd - B = (-1) \cdot ((-1) \cdot sd + B) \]  

Fig. 2 shows a schematic on block level for the complete addition of two SD numbers with a width of four digits. Since two subadders are needed we require four steps in total.

The base_2_step_3 adder requires only three steps, but needs a more complicate Boolean logic what is to explain since both input operands for this adder are SD numbers, i.e. a vector of four bit width is read in. Table 3 shows an example how the addition takes place in three steps. The corresponding Boolean equations are given in (8) to (10).

\[ t_{i+1} = sd_1^+ \bar{sd}_2^- \lor \bar{sd}_1^- sd_2^+ \]  
\[ z_i^+ = sd_1^+ \land sd_2^- \]  
\[ z_i^- = \bar{sd}_1^+ sd_2^- \lor \bar{sd}_1^- sd_2^+ \lor \bar{sd}_1^- sd_2^- \lor sd_1^+ sd_2^+ \]  
\[ t_{i+1}' = (t_i \land z_i^-) \lor z_i^+ \]  
\[ z_i' = (\bar{t}_i \land \bar{z}_i^-) \lor (t_i \land z_i^-) \]  
\[ s_i^+ = \bar{t}_i \land z_i' \]  
\[ s_i^- = t_i' \land z_i' \]

Fig. 3 shows a schematic on block level how the complete adder works. In the blocks denoted step1, step2, and step3 the equations (8), (9), and (10) are carried out.

The idea behind the base_2_step_3 adder to avoid a carry chain is the following one. In the first step the digits, z_i, of the
intermediate value, and the transfer bits, , of the so-called transfer vector, , are produced. It holds for if and for if . Consequently, it is impossible that in the second step a digit with a value equal to or can be produced, i.e. no carry can occur. It requires a final third step to consider the final necessary shift and to produce again an SN number.

The idea behind the adder to avoid a carry is much simpler to understand. The vector contains only 1’s and 0’s, the vector contains only T’s and 0’s. Therefore, it is also impossible that in the second step two T’s or to 0’s meet each other at a certain digit position and no carry can arise.

Both presented adders have been intensively investigated on technological side by a layout synthesis evaluation by other authors [2] concerning their latency and energy consumption behaviour in comparison to a binary carry-look-ahead adder (CLA). The both adders were connected to a multi-value memristor based register file. As expected for both ternary adders a constant run time was received, about 3 ns for the adder and about 3.5 ns for the adder for a 130 nm CMOS process from LFoundry.

As expected, too, the run time for a carry-look-ahead (CLA) adder was higher. It started at about 3.6 ns for a bit length of 16 and ended at about 6 ns for a word length of 512 bits. However, concerning for the energy-delay product the break-even point where the ternary adders were better than the CLA adder was just at 40 digits. We are interested in this paper to see if the energy-delay product would improve if these adders are realised in MeMOS technology.

3. MeMOS implementation of ternary adders

Singh [11] made a proposal to combine memristors and CMOS in order to solve the problem of signal deterioration of the voltage signal level when using a pure ratioed logic [7] with memristors. Basic of ratioed logic are voltage dividers based on memristors, which make it difficult to cascade circuits since voltage drops off at the memristors. The idea is to work with CMOS inverters at certain distances in a digital logical memristor circuit to refresh the voltage signals to the level of the supply voltage. Singh denoted this technique as Memristor-

CMOS (MeMOS) logic. Further inverters are used if inverted signals are necessary for memristor inputs, which are used as logic gates.

Basic building blocks of MeMOS are OR and AND circuits in ratioed logic with closing CMOS inverters for signal restoration what establishes a NOR-/NAND logic.

Singh showed benefits for a full adder realised in MeMOS in comparison with a CMOS full adder concerning energy consumption and latency for an assumed TEAM model for the memristors. We want to apply that approach here to more sophisticated circuits, namely ternary adders to profit from further qualitatively benefits offered by memristor technology.

What we need for a perfect mapping onto MeMOS building blocks are Boolean equations formulated as NAND/NOR expressions. In order to receive such expressions it is necessary to expand the Boolean equations by a double inversion. We demonstrate that exemplarily for the calculation of the signal that is generated in the first step of the adder (11).

\[
c_i^+ = x_i^+ \lor (y_i \land x_i^-)
\]

\[
= x_i^+ \lor (y_i \lor x_i^-)
\]

\[
= x_i^+ \land (y_i \land x_i^-)
\]

The gained expression can be directly mapped in a 1-to-1 fashion onto a functionally equivalent MeMOS circuit shown in Fig. 4. We repeated this procedure for all other equations for both ternary adders, cascaded the gained circuits to build complete ternary adders as MeMOS circuits. These circuits have been simulated by an own written simulator.

4. Simulation system for MeMOS circuits

In order to identify the possible processing speed and the necessary energy consumption of our found MeMOS circuits we are using simulation. For that we wrote a corresponding C++ program. The timely memristor behaviour was modelled by an implementation of the Euler solution of the differential equation to determine the memristance of the memristor devices.
The same procedure is used in the TEAM and the VTEAM model published by Kvatinsky et al. [8], [6]. In addition our simulation system accesses to the physical parameters of the memristors’ given in the TEAM and VTEAM model.

Whereas the TEAM and the VTEAM model are deterministic models our simulation system considers also a statistical model that was introduced by Knowm [10]. The last model is of interest in the view of the fact that real, commercially available memristor devices are standing behind this model.

All these three models can be conveniently considered by our simulation system thanks to the object-orientated approach of C++. A memristor in our simulation system is a class, in which different functional behaviour can be instantiated by selecting a specific model. In order to simulate the functional behaviour of cascaded memristor circuits, as e.g. in the hybrid MeMOS logic, a tree of cascaded and interconnected memristors is built up. This tree is parsed and by that the memristor’s inputs can be determined and the output is calculated in discrete time steps according to the Euler procedure.

The CMOS inverters in MeMOS circuits necessary for signal restoring are functionally modelled as follows. If the input applied to such an inverter is more than 0.7 of the supply voltage, \( V_{dd} \), then the complete voltage hub is switched through and the complete voltage supply level is accessible at the inverter’s output. An analogue behaviour is assumed for the case that the input is below 0.3 \( \times V_{dd} \). In this case the output is pulled down to ground signal. By that it is possible by means with the parameters of the memristor models to evaluate MeMOS and in principle also other memristor circuits to determine latency and energy consumption by simulation.

Table 4 gives an excerpt of the listing of the source code for the simulation of a MeMOS circuit. It shows the definition of a MeMOS circuit and how it is possible to simulate this circuit with different models, just by defining one of the parameters TEAM, VTEAM, or KNOWM during the instantiation of a memristor MeMOS circuit (line 4). The MeMOS circuit itself is defined by the call of corresponding AND or OR memristor functions (line 3) which simulate the building blocks of MeMOS (line 1 and 2). The Boolean parameter at the end of the function call (line 3) determines if the output of the MeMOS circuit has to be inverted or not.

With that program and the known values of a ternary CMOS adder attached to memristor registers, which have been published in [2], we can now carry out a comparison study between ternary adders using MeMOS and conventional CMOS.

5. Comparative evaluation of memristors

First we investigated the memristor behaviour after an initial reset of the memristor, i.e. the both memristors in the building blocks had a state variable 0.5. This corresponds to the situation that the memristance is exactly adjusted at the zero point of its I/V hysteresis curve. We tried to find out at which operating frequency a correct switching would fail because the ions have not enough time to move to another location in the memristor cavity to change the memristance clearly either to the high or to the low resistance mode. Table 5 shows the gained results for an applied moderate voltage level of 1.8 V for three different memristor models.

| Memristor Model | Operating Frequency (GHz) | Energy Consumption (pJ) |
|----------------|----------------------------|-------------------------|
| TEAM           | 1.2                       | 40                      |
| VTEAM          | 1.5                       | 50                      |
| KNOWM          | 2.0                       | 60                      |

Table 4: Extract of the C++ simulation program. Syntactic definition of the circuit shown in Fig. 4
32, and 64 digits. With the simulation system we calculated the average energy consumption per addition. As mentioned, the simulation software allows to specify an operation frequency. The reciprocal of this operating frequency corresponds to a certain amount of time that is applied to the memristor’s inputs.

It turned out by the simulation results that for all three models the energy consumption doubled if the digit width is doubled, too. This was to expect since the ternary adders have a complete regular setup concerning their logic blocks and the the spreading of a carry signal is clearly limited. The latency depends on the number of MeMOS blocks that have to be passed for the two ternary adders. We assume the reciprocal of the operating frequency, \( f \), as run time through one MeMOS building block including therein contained CMOS inverters. Therefore, we need a latency of \( 3 \times \frac{1}{f} \) for the base_2_step_3 adder, and of \( 4 \times \frac{1}{f} \) for the base_2_step_4 adder.

Table 6 shows a kind of cut-off frequency, e.g. 350 MHz for the base_2_step_3 adder, up to which we have got error-free simulation results. For higher frequencies the time the input signals were applied to the memristor inputs is not sufficient long to switch from a previous stored HIGH resistance to a LOW resistance or vice versa. It is interesting to see that the base_2_step_4 adder shows a more stable behaviour than the base_2_step_3 adder in the realistic dynamic situation. For the static case shown in Table 5 this was reversed. Furthermore, it is important to mention that this high speed processing times could only be achieved if the voltage is increased to 6 V for the TEAM and to 4 V for the VTEAM model. This could possibly become a problem for a hybrid integration with CMOS since such high values are no more standard. With the simulated energy consumption values it is now also possible to determine the energy delay product for the ternary adders and to compare them with the CMOS solution that uses memristors as multi-value registers what is done in the next chapter.

### 6. Summary and conclusion

We investigated in this paper the realization of ternary adders using the possibilities of the MeMOS approach, i.e. combining logic circuits based on memristors with CMOS inverters used for signal refreshing. The presented ternary adders exploit the potential of multi-value memristors.

Concerning the values for energy and latency the break even point for a ternary adder using multi-value memristor registers versus a CLA adder was about 40 digits according to the results in [2]. At 40 digits the presented ternary adder base_2_step_3 started to produce a better ED product, that was about 30 ns × pJ. We looked in this paper if better results can be achieved with ternary MeMOS adders. We can fix that is clearly not the case for a dynamic switching mode with MeMOS. The CMOS ternary adder with memristor registers in [2] showed a latency of 3 ns and an energy consumption of 10 pJ which is less than the corresponding values in Table 6 for both MeMOS ternary adders. Concerning the static case the latency for the base_2_step_3 offers with 1.2 ns an improvement of 60%. The energy delay product is with \( 1.2 \text{ns} \times 44 \text{pJ} = 52.8 \text{ns} \times \text{pJ} \) about 75% worse. The ratio concerning energy is even more worse because the simulation does not consider so far the energy consumption of the inverters in the MeMOS circuits.

Therefore we come to the following conclusion. For ternary adders it seems to be better to prefer CMOS circuits with multi-value memristor registers at the moment. However, we will investigate in the future potentials to optimize the MeMOS circuits. E.g., it could be possible to combine multiple steps of the adders, which are now strictly separated. Furthermore, it is to investigate if the approach using MAD and MAGIC gates produces better results since they need less inverters.

### 7. Acknowledgement

The author would like to thank his students Jonas Schmitt and Jonathan Martschinke who wrote the C++ simulation program.
References

[1] J. Borghetti et al., “Memristive switches enable stateful logic operations via material implication,” Nature, vol. 464, pp. 873–876, 2010.
[2] D. Fey et al., “Using memristor technology for multi-value registers in signed-digit arithmetic circuits,” in MEMSYS ’16: Proceedings of the 2016 International Symposium on Memory Systems. New York, NY, USA: ACM, 2016.
[3] L. Gao, F. Alibart, and D. B. Strukov, “Programmable cmos/memristor threshold logic,” IEEE Trans. Nanotechnol., vol. 12, no. 2, pp. 115–119, Mar. 2013. Available: http://dx.doi.org/10.1109/TNANO.2013.2241075
[4] A. F. González and P. Mazumder, “Redundant arithmetic, algorithms and implementations,” Integration, vol. 30, no. 1, pp. 13–53, 2000. Available: http://dx.doi.org/10.1016/S0167-9260(00)00015-8
[5] L. Guckert and E. Swartzlander, “Mad gates - memristor logic design using driver circuitry,” IEEE Transactions on Circuits and Systems II: Express Briefs, vol. PP, no. 99, pp. 1–1, 2016.
[6] S. Kvatinsky et al., “Vsteam: A general model for voltage-controlled memristors,” IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 62, no. 8, pp. 786–790, Aug 2015.
[7] S. Kvatinsky et al., “Mrl - memristor ratioed logic,” in Cellular Nanoscale Networks and Their Applications (CNNA), 2012 13th International Workshop on, Aug 2012, pp. 1–6.
[8] S. Kvatinsky et al., “Team: Threshold adaptive memristor model,” IEEE Trans. on Circuits and Systems, vol. 60-I, no. 1, pp. 211–221, 2013. Available: http://dblp.uni-trier.de/db/journals/tcas/tcasI60.html#KvatinskyFKW13
[9] S. Kvatinsky et al., “Memristor-based imply logic design procedure,” in Proceedings of the 2011 IEEE 29th International Conference on Computer Design, ser. ICCD ’11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 142–147. Available: http://dx.doi.org/10.1109/ICCD.2011.6081389
[10] M. A. Nugent and T. W. Molter, “Ahah computing–from metastable switches to attractors to machine learning,” PLoS ONE 9(2): e85175, 2014.
[11] T. Singh, “Hybrid memristor-cmos (memos) based logic gates and adder circuits,” CoRR, vol. abs/1506.06735, 2015. Available: http://arxiv.org/abs/1506.06735
[12] N. Talati et al., “Logic design within memristive memories using memristor aided logic (magic),” IEEE Transactions on Nanotechnology, vol. PP, no. 99, pp. 1–1, 2016.