Why m-valued circuits are restricted to a small niche?

Daniel Etiemble

ECE department
University of Toronto
de@eecg.toronto.edu

Abstract

Although many m-valued circuits have been presented, very few have been successful. We first examine the technical and economical reasons that prevent the success of m-valued circuits. We show that these circuits implemented with voltage mode, current mode and charge-mode circuits are generally less efficient than the corresponding binary circuits according to VLSI criteria (speed, chip area, power dissipation). When slower access time are allowed, m-valued DRAM can provided a significant reduction in chip area: this is the case for flash memories. M-valued circuits are thus restricted to small niches in the binary world. We also examine if m-valued circuits can solve the interconnection issues.

Keywords

M-valued circuits, VLSI circuits, flash memories, interconnection issues

1. Introduction

For more than fifty years now, researchers have been trying to design efficient m-valued integrated circuits. In the early 70's, K.C. Smith has been a pioneer in the design of m-valued circuits compatible with integrated technologies [1-3]. Many have been proposed in various technologies (TTL, ECL, CMOS) in the 70's [4-8]. A good survey of these circuits can be found in [9]. Some m-valued integrated circuits have ever been used or are being used in commercial products [10]. However, the promise for m-valued circuits has never been so dark: the performance gap between today binary integrated circuits and the proposed multivalued ones continuously increases.

The "circuit" papers that have been presented in the Proceedings of the International Symposium on Multiple Valued Circuits since 1970 can be roughly sorted into three different categories: 1) M-valued circuits that are implemented with currently used integrated circuit technologies. If bipolar (TTL, ECL, I^2L), MOS and GaAs technologies were used in the old times, only CMOS technology must be considered now. 2) M-valued devices exhibiting intrinsic multivalued behavior that are implemented in uncommon technologies. Circuits using resonant tunneling diodes (RTD) are a good example of this approach. 3) M-valued circuits based on new computing models or using non electrical devices to compute data. Molecular computing is the most recent example. Prof. Higuchi group in Tohoku University has delivered many papers considering new computing models for "post-binary" circuits.

In this paper, we outline the fundamental technical and economical reasons that explain why m-valued circuits can only represent a small niche in the overwhelming world of digital circuits. These reasons apply to each one of the three categories defined just before. We also review the common misunderstanding about interconnection issues and m-valued circuits. After this introduction, the second section examines the technical and economical situation of m-valued circuits compared to the binary circuits (2-valued ones) and the analog circuits. The third section explains why m-valued circuits are fundamentally less efficient than the corresponding binary ones and specifies the exceptions that correspond to the commercial m-valued circuits actually used. The fourth section considers the interconnection issues.

2. M-valued, binary and analog circuits

2.1. Digital and analog circuits.

The invention of integrated circuits and the dramatic progresses that they have shown since the mid 60's have led to their universal use in many domains: computers, communication, control, etc. These circuits use voltage, current or amount of charges as support of the information. Electrical devices can be classified as digital or analog circuits. A system in which signals vary in a continuous fashion is called an analog system. A system in which the signals are discrete is said to be digital. According to this definition, any digital system has m different signal levels and can be called m-valued. Binary

circuits are thus a special case of m-valued circuits with m=2 and analog circuits correspond to the case m=∞ . However, it is more pertinent to consider the fundamental difference between m-valued and analog circuits. The typical situation for m-valued circuits is shown in Figure 1. If the total voltage swing is V, the different m values are respectively 0, V/(m-1), 2V/(m-1)… V. The situation is similar when using current or charges instead of voltages. When the same voltage gap V/(m-1) is allocated to each value, the maximum noise margin for the extreme values 0 and m-1 is V/2(m-1). Basically, m-valued circuits have noise margins. If the corresponding m-valued gates are correctly designed, they are level-restoring. M-valued circuits are fundamentally non-linear circuits with noise margins. The level-restoring property means that they don't propagate noise signals. On the opposite, analog circuits are fundamentally linear circuits that add and propagate the noise signals.

Figure 1: Totally ordered set of values

In this context, binary circuits are a special case of m-valued circuits, with m=2. In the next section, we will detail the differences between m=2 and m>2. But some simple considerations are sufficient to understand the technical and economical triumph of binary circuits. With m=2, the maximum noise margin is 50% of the total swing for each level, which is a great advantage in any case. But with today CMOS technologies that use a 1V voltage swing, this advantage becomes decisive. The other big advantage with m=2 is the simplicity of the basic circuits. The n-input NAND or NOR gates only need 2n transistors, which both means high speed and reduced chip area. Our purpose is not to detail basic CMOS techniques that can be found in any good VLSI textbook. We just add that some "analog" techniques (pass transistors) can be used to simplify digital circuitry while conserving the main features of digital circuits.

2.2. The exponential function

Progresses in VLSI technologies have been dramatically fast in the last 40 years. They are best expressed by the famous Moore's law: In his original paper, Moore predicted that the number of transistors per integrated circuit would double every 18 months. Even if this law has often be reformulated differently, all the formulation expresses the exponential progresses.

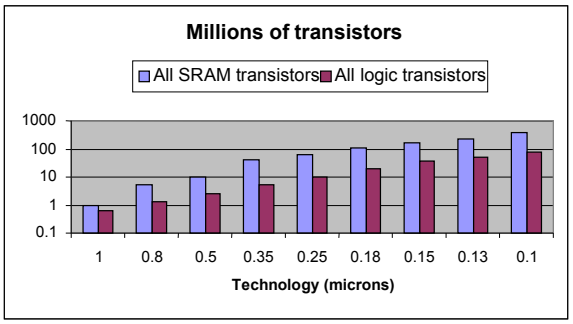

Figure 2 presents the successive generations of CMOS technologies with the corresponding number of transistors for memory and logic applications.

The Moore's law has consequences on many different aspects of computing technologies: the most spectacular one is the exponential increase of performance of VLSI devices (speed, capacity...).

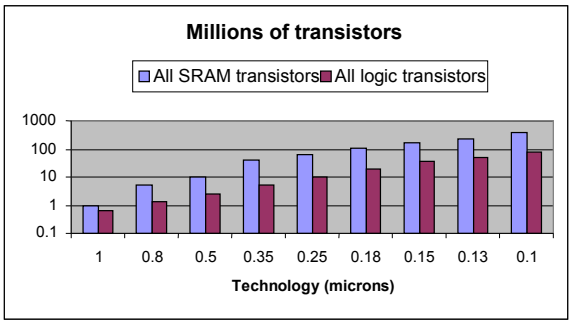

Figure 2: Successive generations of CMOS technologies

Figure 3 shows the yearly increase for CPUs (overall performance and clock frequency). Figure 4 shows the yearly increase for memory devices (DRAM and Magnetic disks). All the parameters don't evolve at the same rate. The latency issue of DRAM leads to the continuously increasing gap between CPU and main memories and explain the increased complexity in computer memory hierarchies. It is more difficult to illustrate performance increase of embedded systems, but their progress rate is also exponential.

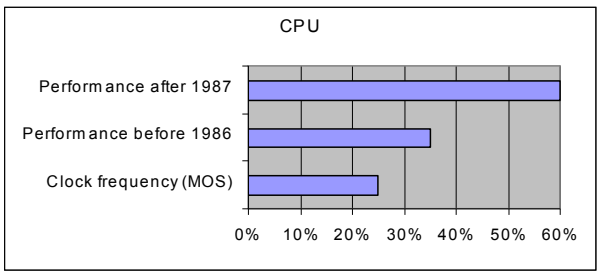

Figure 3: Evolution of CPU parameters (% increase per year)

While digital circuit trends are exponential both for computing and memorizing, progresses have been slower for analog circuits. One well know consequences is that digital circuits have conquered new application areas that previously used analog circuits: voice processing and storage, communications, TV, etc. Digital circuits continuously extend their scope of applications.

The exponential trends have also favored the standardization process. Basically, it is more a "de facto" process than the results of standardization boards. The PC standard, the DRAM and disk standards, the bus and network standards, the convergence towards a limited set of programmable devices (PLDs, FPGAs, etc) are some examples of the most common standards.

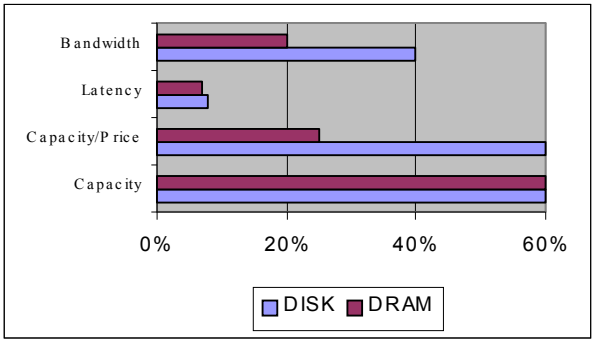

Figure 4: Evolution of storage components (% increase per year)

2.3. M-valued circuits faced to the exponential trends

Being digital circuits, one could think that the evolution of performance of m-valued circuits should be equivalent when m=2 or m>2. But the actual question is: do the multivalued circuits with m>2 have any serious chance 1) to compete against the binary circuits and 2) to compete successfully again binary circuits.

To compete successfully again their binary equivalent circuits, the m-valued circuits should 1)prove some evident performance advantage and 2) exhibit the same exponential performance trends. The second constraint means that using more than 2 distinct levels should not hurt the scaling of technologies.

For m-valued circuits using currently used technologies, we will show in the next section that they are generally less efficient that the corresponding binary circuits according to VLSI criteria (speed, chip area, power dissipation).

With technologies using devices with natural m-valued transfer characteristics such resonant tunneling diodes, it is obviously easier to build m-valued circuits that with current CMOS technologies. These technologies are generally far less mature than the current technologies and less cost effective. However, this in not the main reason why many operational prototypes have not led to commercial circuits. The reason comes from the design techniques to build m-valued circuits. We will summarize these techniques in the next section and show that they raise the same issues when using m-valued devices than when using 2-valued devices to design m-valued circuits.

New technologies and new computing models are faced to the same issues. They must exhibit some performance advantage versus the current technologies. More precisely, they must exhibit a performance/cost ratio advantage. They must also exhibit similar or better exponential trends. The current CMOS technologies will face physical limits in the future. It would be totally stupid to claim that the current exponential trends will continue for ever because ultimate physical limits do exist: Figure 5 shows the solid structure of silicon. The lattice is defined and its dimensions cannot be reduced. However, as long as the exponential progresses of current technologies continue, there will be few opportunities for concurrent technologies to emerge, as they will remain less efficient and more costly. The questions are similar with new computing paradigms. If new technologies emerge for which the basic NOR and NAND gates are available, the principles of logic design that are used today will continue to prevail, as they are simpler and already provide a lot of CAD tools to use. If they cannot, new computing paradigms will be necessary. The present research works on new computing paradigms will then be useful. Unfortunately, they are few chances that they will be used today because of the so large domination of digital integrated circuits.

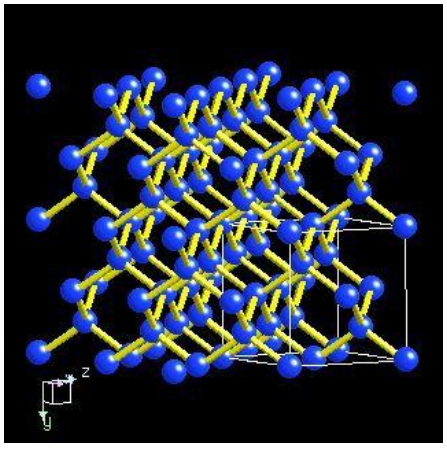

Figure 5: Solid structure of silicon

Many humorists have outlined that it is difficult to forecast, specially for the future. The discussion on the future of the new computing paradigms is out of the scope

of this paper. However, considering the past competition between dominant paradigms and new paradigms can give some insights. For several decades, parallel processing has competed with "sequential" computing. At the hardware level, many different models of machine have been proposed. Many programming paradigms and many parallel languages have been proposed. In the last decade, everything has converged towards simple models compatible with the main stream (PCs and workstations). Clusters of multiprocessors at the hardware levels, simple programming models embedded into the MPI (message passing) and OpenMP (share memory model) have emerged. Simplicity and compatibility with the dominant model are the key words.

3. Design of m-valued circuits

More than 10 years ago [11], we have described the implementation of m-valued circuits with binary circuits by using voltage-mode circuits and current-mode circuits. Only the "charge-mode" circuits should be added to this description.

3.1. Voltage mode circuits

Voltage mode circuits can be implemented according to Figure 6. Each m-valued input is decomposed into m-1 binary signals by m-1 threshold detectors. Each threshold detector compares the input value with 0.5 + p, where p are the successive integers from 0 to m-2. The binary signals are computed by m-1 binary functions that deliver the binary coding of the m-valued output. The output encoder transforms the m-1 binary signals corresponding to the binary functions into the m-valued output. The approach that we described corresponds to the monotonic system of Post algebra[11]. Many different forms of Post algebras can be used, but there are all equivalent as shown by Nutter et al [12]. We have implicitly used the form that leads to the simplest circuit implementation.

We must first observe that the complexity of the procedure comes from the fact that the m different values of the Post algebra are totally ordered: 0<1<2…<m-2<m-1. Using Post algebra is mandatory because the different voltage values that are used are totally ordered. The situation is the same when using currents or amounts of charge to define the m different values. As the m different values are totally ordered, m-1 threshold detectors are needed. The problem is a physical one and inventing new forms of Post algebras cannot help because they will be all based on the total order set.

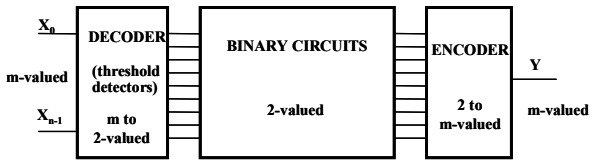

Figure 6: General scheme of m-valued circuits

To compare with the binary case, let's consider $m=2^k$. A direct comparison between Post algebras and Boolean algebras is then possible. The implementation of the m-valued function (and the corresponding m-valued circuit) requires m-1 binary sub-functions plus the input decoder and the output encoder. For the Boolean algebra, the m values will be coded with $\log_2(m)$ bits and neither the decoder nor the encoder are needed. The comparison considers the general case: for some specific m-valued functions, some simplifications are possible. But generally, even without considering the implementation of the decoder and the encoder, the complexity of m-valued circuits is proportional to m-1, when the complexity of binary circuits is proportional to $\log_2(m)$. m-1 is greater than $\log_2(m)$, except when m=2! All m-valued designers that have used voltage mode circuits to design 4-valued circuits and compared the performance with the binary circuits implementing the same functions have got greater propagation delays and greater chip areas. As the current trend is the reduction of the voltage swing to minimal values to reduce power dissipation, this makes very difficult to efficiently use more than 2 levels.

What could be the interest of m-valued circuits in this situation? The only possible advantage would be to reduce the number of interconnections between circuits while having more complicated and slower circuits. If reducing the number of interconnections was a key issue, it could be seriously considered. In the next section, we will show that better solutions exist to reduce the number of connections. More fundamentally, we will show that the interconnection issue in VLSI design is not the number of connections, but the propagation delays in the connections.

3.2. Current-mode implementation

Current mode circuits can also be implemented according to the general procedure illustrated in Figure 6. However, there exist specific implementation problems. While many voltage inputs can be connected to one voltage output (fan out >1), only one current input can be connected to one current output. Current mirrors are needed to duplicate outputs or to reverse a current flow to get bi-directional currents. I^2L circuits and CMOS current mode circuits use these basic

components. The CMOS version of the current sources and threshold detectors [13] is shown in

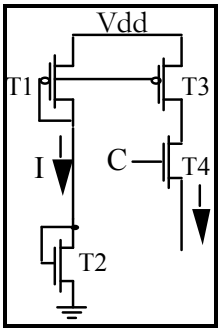

Figure 7 and Figure 8. The threshold detectors include current mirrors. The general considerations that we considered for voltage mode circuits (noise margins, complexity) still apply. But there are additional constraints due to the supplementary "analog operations" of the current mirrors. We have considered the dispersion issues of I²L circuits and current mode CMOS circuits in [14]. The main issue is the tolerance on the current levels.

With m-valued current mode circuits, there is one particular case which makes these circuits attractive. It correspond to the "threshold logic" that was popular in the 50's. A subset of the binary functions can be expressed as f = 1 when $\Sigma w_i x_i >= T$ and 0 otherwise. w is some weight and T is the threshold. The general scheme for the implementation of threshold functions is shown in Figure 9. Many arithmetic functions can be expressed by threshold functions. For instance the carry out of a 1 bit full adder is the threshold function $(x_i+y_i+c_i) >= 2$. More generally, all carry functions are threshold functions. As currents can be added or subtracted for free (simple connection), the threshold logic implemented by m-valued current mode circuits look attractive and regain some popularity in the m-valued community in the 80's [15,16]. A very spectacular multiplier has been presented with outstanding performance[17]. But, no more circuits with such performance have been published since this one. The current mode m-valued circuits face the same basic issue as the threshold circuits of the 50's: keeping some noise margin between current levels require very low dispersion tolerance on the basic parameters of the key circuits: current mirrors and threshold detectors.

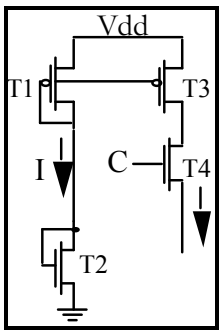

Figure 7: CMOS current sources

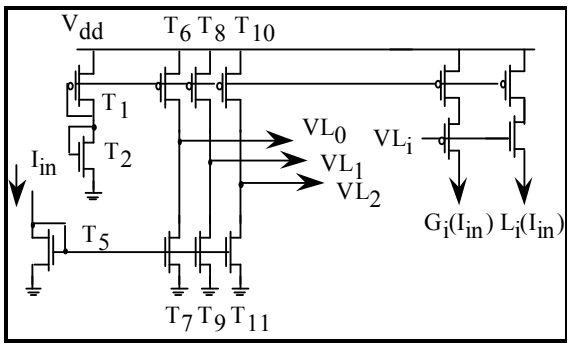

Figure 8: CMOS threshold detectors

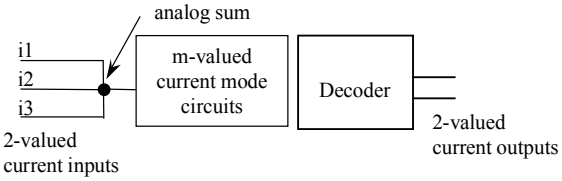

Figure 9: Implementation of threshold functions with current-mode circuits

3.3. Charge-mode implementation

The 1 bit cell used in binary DRAMs consists in a capacitor that is charged (or not) with a charge $Q=C\Delta V$ to store 1 (resp. 0) and a single transistor. The extension to m-valued case consists in having charge per level signal identical to the binary case, but using m levels instead of 2 levels. The time to charge or discharge the capacitor to m-1 is proportional to m-1 [18]. Charging a m-valued cell will be m-1 times longer than charging a 2-valued cell. Most of the time, this approach gives a chip area advantage assuming that the m-valued cell uses the same area that the binary one. But the disadvantage is slower access times. This is extensively discussed in [19]. The

scope of applications of memory is rather large and a low access time is not always a constraint. Increasing memory capacity could be far more important. For instance, this is the case when the goal is to replace back-end magnetic memories with semi-conductor memories. The access time, which is so critical for DRAMs used as computer main memories, becomes far less important. Here is a niche for m-valued circuits. It is significant that many of the most interesting m-valued results correspond to m-valued memory technology and that the rare currently used m-valued circuits are memory circuits. However, these results don't contradict the general rules. Even in this most favorable case, the m-valued circuits are faced to the tolerance issues and the slower operational speed.

4. The interconnection issues

In most cases, the first motivation that is claimed for using m-valued circuits is to solve the interconnection problems. As a matter of facts, only the voltage mode m-valued circuits could be used for reducing the number of wires between different logic blocks. 4-valued signaling divides by 2, 8-valued signaling divides by 3 the number of wires. However, it is interesting to consider how the semi-conductor industry has faced the interconnection issues in the past.

4.1. More interconnection layers

The first generations of CMOS technologies have used one metal layer. Then, two metal layers have been used. Now, up to 4 metal layers are being used. Increasing the wiring capability has been realized with more metal layers and not by transmitting more bits by wire.

FPGA circuits [20] are a good example of circuits for which the interconnection resources are critical. FPGAs generally provide two types of connections. High speed direct connections are used to propagate signals through the whole chip. Then programmable switches are provided to interconnect the different logic blocks. Some switches are used to connect logic cells to the wires in a channel. Some switches are used for connections between crossing channels. Using "pass transistors" for connections introduce propagation delay issues. Too many pass transistors in series lead to significant increase of rise and fall times due to the line effects. Using 4 values (or 8) instead of 2 means increasing the voltage swing (the 2 valued voltage swing is generally minimal and cannot be split further). Increased voltage swing means increased switching times and propagation delays. As the FPGA frequencies are significantly slower than the custom VLSI circuits using a similar technology, degrading further the clock frequency may not be the good solution. The FPGAs for which the interconnection issues are the most critical are the biggest ones and they are the ones that also require good timing performance.

4.2. Serial transmission

Reducing the number of wires is a problem that has been addressed recently by the memory designer. The solution that has been used is high speed serial transmission. Obviously, doubling the clock frequency to transmit a bit is equivalent to transmit 2 bits per clock period with 4-valued signals. The RDRAM interface (Figure 10) shows the approach used by computer industry [21] to reduce the number of off-chip connections and improve the data transfer bandwidth.

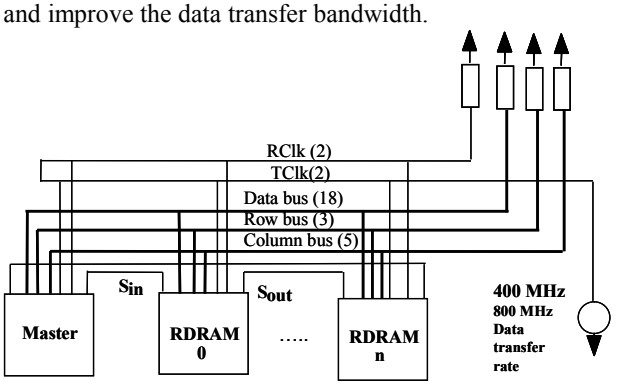

Figure 10: The RDRAM interface (serial transmission)

Let us compare the SDRAM and RDRAM cases. Switching from a 100 MHz 8-bit parallel transmission (SDRAM interface) to a 800 MHz serial transmission (RDRAM interface) is equivalent to use 256-level signaling at 100 MHz, 16-level signaling at 200 MHz or even 4-level signaling at 400 MHz! But high speed serial links are far easier to implement.

4.3. Again on interconnection issues

Ten years ago, Stone and Cocke already considered that the problem for off-chip transmissions was with "the speed limitation of metal interconnections" [22]. In the recent years, this issue has confirmed to be the biggest problem related to interconnections. During the first decades of VLSI development, the propagation delays in the interconnections have always be considered as negligible compared to the propagation delay into the basic blocks (logic gates). The situation has dramatically changed. With the scaling of the technologies, the RC effect of the interconnection wire becomes more and more important and the propagation delay corresponding to the propagation of the signals in the interconnections becomes more important than the propagation delays through the gates. Advances in manufacturing technologies are the key

factor to overcome this problem. IBM, by introducing copper connections recently, contributed to postpone the problem for some time. But the trends is certain. M-valued circuits cannot help for this problem. While expecting new progress in connection technology, the best solution at circuit level is to use the smallest possible voltage swing because larger voltage swing will be more sensitive to RC effects. Interconnections do become a critical issue in advanced VLSI technologies. But it is absolutely different from the traditional view of m-valued designers.

5. Concluding remarks

In several decades of tremendous development of VLSI technologies, several significant m-valued designs have been presented: arithmetic circuits, ROMs and RAMs. M-valued memory circuits are presently commercialized by Intel. But everybody can observe that m-valued circuits are marginal among electronic circuits while binary circuits continue to progress exponentially and extend their scope of applications.

There are fundamental reasons: the voltage levels, the current levels or the amount of charges that are used to represent the m different values are totally ordered. Because they are totally ordered, some flavor of Post algebra should be used. All the needed math support is available and the circuit designer will always need m-1 threshold decoders plus current mirrors with current-mode circuits. Most of the time, the circuits will be slower and more complicated than the corresponding binary ones. Tolerances with m larger than 2 will remain the big issue. These remarks, that looks obvious when using standard technologies, are also valid for devices or technologies that exhibit m-valued characteristics.

M-valued designers were (are) right when claiming that the interconnections are a critical issue for modern VLSI technologies. But reducing the number of off-chip interconnections is realized efficiently by high-speed serial links and m-valued signaling doesn't seem to be the solution for reducing the number of on-chip connections. The most fundamental issue is the increasing part of the interconnections in the overall propagation delays and m-valued signaling cannot help.

6. References

[1] Z.G. Vranesic and K.C. Smith, "Engineering Aspects of multi-valued logic systems", Computer, 7, 1974, pp. 964-971.

[2] Z.G. Vranesic and K.C. Smith, "Electronic circuits for multi-valued digital systems", in Computer Science and Multiple-Valued Logic, Theory and Applications, North Holland, 1977, pp. 397-419

[3] K.C. Smith, "Multiple-Valued Logic: A Tutorial and Appreciation", Computer, April 1988, pp. 17-27

[4] D. A. Freitas and K.W. Current, "A quaternary logic encoder-decoder circuit design using CMOS", in Proc. Int'l Symp. Multiple Valued Logic, pp. 190-195, May 1983.

[5]| S.P. Onneweer and H.G. Kerkhoff, "Current-Mode High Radix Circuits", Proc. Int'l. Symp. Multiple Valued Logic, pp. 60-69, May 1986

[6] H.T. Mouftah and I.B. Jordan, "Integrated circuits for ternary logic", in Proc. 1974 Intern. Symp. On Multiple-Valued Logic, "Morgantown, W. Va, 1974, pp. 285-302.

[7] D. Etiemble and M. Israel, "A new concept for ternary logic elements", in Proc. 1974 Intern. Symp. On Multiple-Valued Logic, "Morgantown, W. Va, 1974, pp. 437-458.

[8] T.T. Dao, " Threshold I^2L and its application to binary symmetric functions and multivalued logic", IEEE J. Solid-State Circuits, vol. SC-12, pp463-472, Oct.1977

[9] S.L. Hurst, "Multiple-valued Logic—its Status and its Future", in IEEE Trans. Computers, Vol C-33, 1984, pp 1160-1179.

[10] Intel, "What is flash memory", http://developer.intel.com/design/flash/articles/what.htm

[11] D. Etiemble and M. Israël, "Comparison of binary and multivalued integrated circuits according to VLSI criteria", IEEE Computer, April 1988, pp. 28-42.

[12] R.Y. Nutter, R.E.Swartwout and D.C. Rine, "Equivalences and transformations for Post Multivalued Algebras", in IEEE Trans. Computers, March 1974, pp. 294-300.

[13] K. Navi, A. Kazeminejad and D. Etiemble, "Performance of CMOS Current Mode Full Adders", Proc. 24th Int'l. Symp. Multiple Valued Logic, pp. 27-34, May 1994

[14] D. Etiemble, "On the performance of multivalued integrated circuits: Past, Present and Future", Proc. Int'l. Symp. Multiple Valued Logic, pp. 156-164., May 1992

[15] S. Kawahito, M. Kameyama and T. Higuchi,, "VLSI-Oriented Bi-Directional Current Mode Arithmetic Circuits Based on the Radix-4 Signed Digit Number System", in Proc. Int'l Symp. Multiple Valued Logic, pp. 70-77, May 1986

[16] T. Yamakawa, "CMOS multivalued Circuits in Hybrid Mode", in Proc. 1985 ISMVL, Kingston, Ontario., pp 144-151.

[17] S. Kawahito, M. Kameyama, T. Higuchi, H. Yamada, "A

32 x 32 bit Multiplier Using Multiple-Valued MOS Current-Mode Circuits", IEEE J. Solid-State Circuits, vol. SC-23, pp124-132, Feb.88

[18] M. Horiguchi, M. Aoki, Y. Nakogome, S. Ikenaga and K. Shimohigashi, "An Experimental Large-Capacity Semiconductor File Memory Using 16-Levels/Cell Storage", in IEEE J. Solid-State Circuits, Vol SC-23, N° 1, February 88, pp 27-33

[19] P. G. Gulak, "A Review of Multiple-Valued Memory Technology", Proceedings 1998 ISMVL, Fukooka, Japan, pp 222-231.

[20] A Vranesic, "The FPGA challenge", Proceedings 1998 ISMVL, Fukooka, Japan, pp 121-126

[21] RAMBUS web site: www.rambus.com

[22] Harold S. Stone and John Cocke, "Computer Architecture in the 1990s", Computer, September 91, pp 30-3