SDPS-2021
Printed in US, December 2021

TERNARY COMPUTING; THE FUTURE OF IOT?

Henning Gundersen, Steven Bos
Department of Science and Industry systems,
University of South-Eastern Norway
henning.gundersen@usn.no steven.bos@usn.no

ABSTRACT

Ternary logic, in which the number of discrete logic levels are restricted to three, has been a subject to excessive research over several years. In this position paper we discuss advantages and consider the future impact for IoT devices in seven categories: computational power, communication, compression, comprehension, cyber-security, design complexity and energy consumption.

INTRODUCTION

The use of CMOS-transistors has forced both hardware and software developers to use binary solutions as these transistors inherently have two stable states; "on" or "off". Between these two stable states is an undesired analog state of behavior which is designed to be as short as possible. Other bases have been used for computation as well such as the 1958 ternary computer Setun (Brousentsov et al. 2002). In practice this computer could not make 3 stable states, but used a multiple of 2, namely 4 stable states and discarded one state for computation (Ware, 1960). By using balanced ternary numbers (-1,0,+1) it is possible to add both negative and positive numbers without using a sign bit. Note that this is purely a logical encoding with implications for logic gate design. At the signal level three positive voltage levels are still possible. Balanced signals around ground are also possible but require a dual power supply design.

Donald Knuth, a computer scientist known for his seminal work *The Art of Computer Programming* stated (Knuth, 1981):

> "If it would have been possible to build reliable ternary architecture, everybody would be using it."

Recent advances on memristors (Chua, 1971) and CNTFETs (Lin & kim & Lombardi, 2011) now make it possible to design and fabricate reliable ternary hardware (Moaiyeri et al. 2013; Nancy Soliman et al. 2019).

Ternary logic vs. Multiple valued logic

In the last few decades Multiple-Valued logic (MVL) has been proposed as a possible substitute to binary logic. While binary logic is limited to only two states, "true" or "false", multiple-valued logic can replace these with finitely or infinitely numbers of values. A MVL system is defined as a system operating on a higher radix than two (Smith, 1988). A radix-n set has *n* elements, {0, 1, ..., n-1}. The practicality of MVL depends on the accessibility of the devices constructed for MVL operations (Etiemble, 1992). The engineering challenge is thus to fabricate devices that are able to switch between the different logical levels, and preferably be less complex (eg. with equal or less components, cost of fabrication, die area, power consumption and signal propagation delay) than their binary counterparts. Ternary logic is MVL compliant. However, it only uses three logic states, "0", "1" and "2". Higher radices give more complexity, so is there an optimal radix?

The radix economy

Historically, the radix economy of a number N was proposed as a cost metric to compute the optimal radix. This number is often the computer architecture, the largest number that the processor can handle such as N = 16 bits = 65536. The metric uses the *rw-product* to estimate the cost or hardware complexity. Here *r* is the radix or amount of unique symbols. The *w* is the width of the word or amount of positions needed to encode a random n from N using the symbols available. For instance, the number 5 is encoded in binary as 101_2, in unbalanced ternary as 12_3 and balanced ternary as $+0-_{3b}$. Therefore, to encode the number 5, binary and balanced ternary needs 3 positions while unbalanced ternary needs just 2. This is a direct consequence of computers using a positional numbering system. The *rw-products* are respectively 6, 9 and 6. If N=6 we should compute the *rw-product* of the other possible *n* and average (assumed uniform) the number to be able to compare the radices. The derivation of the function *rw* gives a minimal point at 2.71828 (Hayes, 2001), where the radix and width are treated as continuous variables. This point is remarkably the napierian base (Appleyard, 1913). The closest discrete number to this minimal point is radix 3, hence it being a more optimal numbering system than base 2 (Hayes, 2001).

Although possibly the best metric we currently have, there are several points of critique one should be aware of. The first is related to *r,* which can also be interpreted as the amount of devices needed to store (eg. memristors, capacitors, transistors) and process (eg. transistors) all symbols. If we need more than log(3)/log(2) * 100% ≈ 58.5 % devices (for a large N such as 2^{16} advantages discussed below might be void. This average number can thus be seen as the theoretical design overhead. Another point is related to the equal proportionality of the *rw-product*. One could interpret the *r* being more important as binary has a 70 year advantage in fabrication cost and functional efficiency (eg. caching, branch prediction) thus directly influencing the first point. In this work we assume that ternary is more efficient than binary because we expect the fabrication cost to go down and designs be more efficient when adopted. Since the

rw-product gives a sense of the amount of actions or transistor switches required to process *n*, we argue that the amount of transistors per area compared to binary is a good estimate when designing ternary chips.

TERNARY LOGIC AND IOT

Ternary computing has several advantages compared to binary computing of which are grouped in 7 categories below. As IoT devices have strict requirements with regards to power consumption, heat generation, footprints and costs, this class of devices will especially benefit from the transition to pure ternary or mixed binary-ternary signals.

Computation

Where does the 58.5 % information advantage of ternary over binary come from? To represent or store in a register a number N, less positions and thus storage devices are needed for ternary. When focusing on performance or logic operations on numbers less transistor switches are needed. What does this mean for computation? Let's assume we design an adder logic gate to add and subtract numbers. It belongs to the most fundamental functional units in a CPU and optimizing it results in various improvements across the whole CPU. If a 1 bit binary adder is constructed with 28 transistors, a ternary variant (with the same functionality) can use up to 28 x 1.585 = 44 transistors and still be more efficient. Since many of these adders exist in one design, this advantage scales rapidly. A 22 bit binary adder will use 22x28=616 transistors with a resolution of 2^{22} = 4194304. A 14 trit ternary adder will use 14x44=616 transistors with a resolution of 3^{14}= 4782969, so it can handle slightly larger numbers than a binary design for the same transistor count. Note that 14 trits is the closest discrete approximation of $2^{22} = 3^n$.

With an efficient synthesis tool (Lin & kim & Lombardi, 2011) it is possible to make optimal ternary designs that will lead to area benefits as shown in Fig. 1 where a larger digit size gives better advantage (S. Kim, 2018). IoT devices will benefit of more computational power on less area. They will also benefit from using the balanced ternary numbering system as the arithmetical computation will be faster. For subtraction operations 2's complement is not needed in a balanced numbering system as the same adder can do both adding and subtracting. In addition, you only need one logic gate (Gundersen, 2006) to perform a comparison for more, less and equal. This can be used in search tree structures to increase the speed for search operations.

Communication

In wireless communication today different modulation methods are used. For instance, in telecommunication and digital cable TV, a common method is Quadrature Amplitude Modulation (QAM) (Lajos L. Hanzo et al. 2004). Digital QAM uses a combination of Phase Shift Keying (PSK) and Amplitude Shift Keying (ASK) for information interchange. Quadrature refers to a 90° phase difference. Each combination of phase and amplitude, the *modulation state* and the *symbol*, represents a combination of two or more bits. This means that a multi-valued encoded signal is already used in communication.

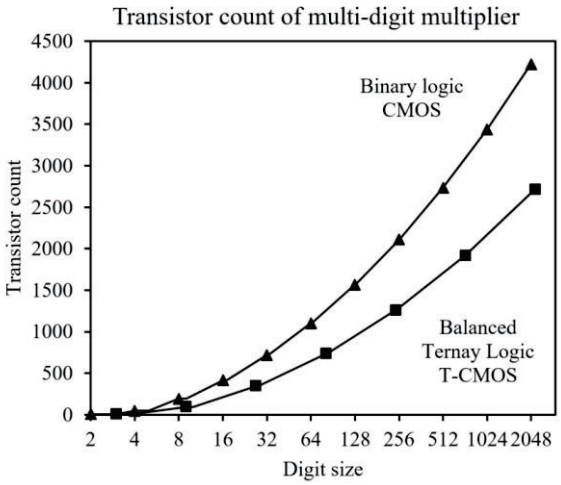

Fig. 1 Number of Transistor vs. digit size of a signed multiplier (S. Kim, 2018)

So why does the sending and receiving hardware have to encode and decode to binary values? For IoT devices less conversion is better as each conversion consumes energy. Instead of sending modulated binary signals, why not send modulated multi-valued signals? A 8 trit word compared to a 8 bit word can carry $3^8/2^8$=25.6 times more information which directly leads to higher communication speed and less delay which is an important issue in communication between servers and for internal network communication between IoT devices.

Consumption

Energy consumption is an issue for IoT devices due to their small footprint, battery reliance and limited cooling options. Often these devices have embedded various sensors which benefit from low power operation. In binary when flipping bits, a signal toggles between 0 and 1, the full length of a signal bandwidth. In balanced ternary the states are -1, 0 and +1. Four out of possible six state changes have half signal lengths, from -1 to 0 or +1 to 1 and reverse. Dynamic power constitutes of the transient power consumption and will thus be smaller in ternary for the same amount of transistors.

With CNTFETs it is possible to make energy efficient MVL-circuits (Ramzi A. Jaber et al. 2019). There are currently not built any ternary circuits using CNTFET-technology, but simulations show a low power-delay product consumption (Sunmean, 2018), which shows the potential of building power efficient ternary designs. Ternary Tunneling CMOS transistors on commercial wafers are able to show these consistent power advantages (Jeong et al. 2019). In 2019 a modern microprocessor built from complementary CNTFETs was built (Gage Hills et. al. 2019). They have overcome major intrinsic CNT fabrication process challenges and demonstrate it is possible to build a complete microprocessor using complementary CNTFETs. The same technology can also be used to build ternary computing hardware.

Compression

While discussing the origin of the information advantage we made a clear distinction between memory and logic, the

two fundamental building blocks of a computer. For memory this means less needed memory devices and can thus be seen as compression of information. For example, a 16 bit number has a resolution of 2^{16}=65536 while a sixteen-trit number has a resolution of 3^{16}=43046721. The closest approximation that can encapsulate all values of that resolution with ternary would be an eleven-trit number. This has resolution of 3^{11}=177147 which means the handling of +170 % larger numbers and an effective reduction in devices of -31 %. In multimedia applications, this can lead to improved audio, photo and video compression algorithms.

For storing data, we showed that off-the-shelf memory resistors or memristors are excellent devices for low power and non-volatile (persistent) MVL data storage (Bos, 2020). They can be made as small as a few nm^2 (S. Pi et al. 2018). Other researchers have demonstrated that a single device can hold up to 6.5 bits of information (96 states) in a single memristor (Stathopoulos et al. 2017). Like in modern data communication, data storage in solid state drives already use non-binary techniques. Using capacitors, a single memory element can store 16 (4 bits) different charge levels/states (Toshiba, 2017). The trend is to cram more stable charge levels into devices, thus further increasing the information density or compression ratio. As binary encoding/decoding (multiples of 2) are used, read and write time will either increasingly take longer in serial or take more silicon space when done in parallel.

Comprehension

The world is analog and modelling it in binary result in known and unknown simplifications, especially in software engineering. The closer we can get to the analog equivalent, the less error prone our software can be. In a great exposé Charley Bey [1] gives several examples of why writing software is made easier with ternary computers. For example, in low level programming it is important to know your data type when working with integers being either signed or unsigned. Compilers and instructions set require additional code to deal with negative numbers. Most microcontrollers have control registers with sign and sign overflow flags.

Using binary result in classical semantic problems as it must be either true or false. Maybe, partial or unknown are not valid possibilities. This illusion of simplicity effects our code structures and interpretation. For example, in binary if a bit is not 0 it must be the opposite, a 1. This is logically and functionally not always true. An accidental bit flip due to eg. cosmic radiation is hard to prevent. With a logically "unknown" middle state the change for a state inversion can be reduced. Binary control structures like if/else also force the programmer to disregard the option of more options. For humans modelling a state as unknown is common and gives the system (eg. the O/S or hardware) the possibility to notice this unknown, reflect and resolve it at a later time.

Cyber Security

Today we have smart homes and smart cities. We use more and more smart devices connected together in large IoT based networks. This makes us vulnerable for cyber-attacks. What can we do against hacking and other threats? One popular attack vector are the communication channels,

[1] Presentation at the C++Now Conference 2016. Slides and presentation at https://youtu.be/gLJrOTFw6J0

especially in an IoT network. Current technology uses binary solutions, but as we have been discussing in earlier sections, ternary has more information density. By making a 128 trit encryption key we can generate $3^{128}=1.18x10^{61}$ different combinations, compared with a 128 bit encryption key which has $2^{128}=3.40 \times 10^{38}$ combinations, hence by using a ternary secure key (Bertrand Cambou et al. 2018) we decrease the possibilities for cracking the encrypted key substantially. In addition, we can use a ternary Physical Unclonable Function (PUF) circuit using CNTFETs (Zhengyang He et al. 2018; Nitish Kumar et al. 2019) and also implement Ternary Addressable PUF Generator (T-APG) utilized as a hardware layer for protection of databases (Mohammadinodoushan et al. 2019). By mixing binary with ternary signals for eg. an encryption scheme, logic states are obscured or lost, especially when reading with binary hardware. A hard problem for cryptographic circuits is hiding its output at the power level. With differential power analysis (DPA) the output of such circuits can be read. One of the solutions is by using "multi-bit data representations" or in other words MVL (Kocher et al. 2011).

Design Complexity

Design complexity exist on many levels. When designing logic gates, we can choose from $2^{(2*2)}$ =16 different logic functions with two inputs and one output in binary. Examples are the AND, OR, XOR gates. In ternary we can make $3^{(3*3)}$ =19683 logic functions for a similar two-input gate. While in binary each gate has found its usefulness, in ternary many of these logic gates are undiscovered. The amount of expressiveness is much greater, meaning that complex designs in binary can be a single gate solution in ternary such as a half adder gate, carry or data latch. A standard library of these logic gates would mean a reusable way to build larger circuits with more overview due to fewer individual logic gates.

Another level of design complexity is at the transistor geometry level, especially wire layout. These interconnects can take up to 90 % of the dynamic power consumption and often requires manual labor to do correctly (Magen et al. 2004). With smaller technology nodes (currently at 5nm) more transistors can be crammed, and thus more interconnects need to be connected. These interconnects are used in binary fashion while they inherently can handle analog signals. This inefficiency causes the switch to dissipate or charge all power in the wire before a transistor can switch. With ternary a design can save in both amount of interconnects, as there are less transistors needed, and in interconnect usage with on average smaller power transitions.

At the highest design level every chip is designed with power constraints. Cooling solution have not improved since the 2000's meaning that every technology node introduces more dark silicon. This term signifies that modern processors can have more transistors per cm^2 but that less of them can be active or must run at reduced speed at the same time to prevent overheating. It is estimated that barely 10 % of the transistors can be active in modern designs. With a higher utilization of interconnects and with less transistors needed

an increase of efficiency can be realized by switching to a mixed binary/ternary or full ternary design.

For IoT devices a simpler and smaller design mean more possibilities for form factors and faster development cycles. For example, without the need for a sign bit, die space is saved, logic gate functionality is made closer to how humans normally handle negative numbers and programming the gate with assembly becomes simpler.

CONCLUSION

In this position paper we gave an overview of 7 categories where a transition from pure binary to a mixed binary/ternary or pure ternary signal would be beneficial for general computing and IoT devices specifically due to their requirements. This trend is ongoing for a number of years with communication technology being non-binary since inception and solid-state storage solution being non-binary since the last decade. We find that MVL is often disguised in technology or adapted to binary just for the sake of compatibility. Now with new advances in logic processing using CNTFET, memristors and TCMOS, the missing piece is finally addressed. This finally enables the pursuit of post-binary solutions and practical experimentation. Especially the area of reduced noise margins that follow from a higher radix and the state stability at high frequencies is a topic that need more research.

Binary technology is not inferior. The 70 year legacy will require many fabrication and engineering challenges to be solved to bring ternary to the same level such that the assumptions in the radix economy hold. But when it is solved, IoT devices being low power, small and low cost will benefit the most as they reap all benefits discussed.

ACKNOWLEDGMENT

The authors would like to thank the Ternary Research Group and the SMART research group at the University of South-Eastern Norway for supportive and constructive feedback.

REFERENCES

Appleyard, R. (1913) The Direct Measurment of the Napierian Base, *Proc. Phys. Soc. London, 26*, page(s) 178-182.

Bertrand Cambou, Paul G. Flikkema, James Palme, Donald Telesca and Christopher Philabaum (2018) Can Ternary Computing Improve Information Assurance? *Cryptography 2018, 2(1), 6.*

Bos, S.; Gundersen, H.; Sanfilippo, F. (2020) uMemristorToolbox: Open source framework to control memristors in Unity for ternary applications, *2020 IEEE 50th International Symposium on Multiple-Valued Logic (ISMVL2020),* pp. 212-217, doi: 10.1109/ISMVL49045.2020.000-3.

Brousentsov, N. P.;Maslov, S.P.;Ramil Alvarez,J.; Zhogolev, E. A. (2002). Development of ternary computers at Moscow State University, *Russian Virtual Computer Museum.*

Chua, L. (1971). Memristor - The missing circuit element, *IEEE Transactions on Circuit Theory Volume: 18 , Issue: 5* , Page(s): 507-519, September 1971.

Etiemble, D. (1992) On the Performance of Multiple-valued Integrated Circuits: Past, Present, and Future. *Proceedings of the 22th international symposium on Multiple-valued logic*, page(s) 56-164.

Gage Hills, Christian Lau, Andrew Wright, Samuel Fuller, Mindy D. Bishop, Tathagata Srimani, Pritpal Kanhaiya, Rebecca Ho, Aya Amer, Yosi Stein, Denis Murphy, Arvind, Anantha Chandrakasan and Max M. Shulakerm. (2019) Modern microprocessor built from complementary carbon nanotube transistors, *Nature 572*, page(s) 595-602.

Gundersen H.; Berg, Y. (2006) A novel ternary more, less and equality circuit using recharged semi-floating gate devices *2006 IEEE International Symposium on Circuits and Systems*, pp. 4 pp.-3172, doi: 10.1109/ISCAS.2006.1693298.

Hayes, B. (2001). Computing Science: Third Base. *American Scientist, 89(6),* page(s) 490–494.

Jeong, J.W., Choi, YE., Kim, WS., Park, JH., Kim, S., Shin, S., Lee, K., Chang, J., Kim, SJ. and Kin, K.R. (2019) Tunnelling-based ternary metal–oxide–semiconductor technology. *2019 Nature Electron 2*, pp. 307–312, doi: 10.1038/s41928-019-0272-8.

Knuth, D. (1981). The Art of Computer Programming, Second edition, Seminumerical Algorithm Vol 2, *Addison-Wesley Publishing Company*, page(s) 190-193.

Kocher, P.; Jaffe, J.; Jun, B. et al. (2011) Introduction to differential power analysis. *J. Cryptogr Eng* 1, 5–27. doi: 10.1007/s13389-011-0006-y

Lajos L. Hanzo; Soon Xin Ng; Thomas Keller; William Webb. (2004) Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems, *Wiley-IEEE Press*, pp.1-1, doi: 10.1109/9780470010594.part1.

Lin, S.; Kim , Y.; Lombardi, F. (2011). CNTFET-Based Design of Ternary Logic Gates and Arithmetic Circuits, *IEEE Transactions on Nanotechnology, vol. 10, no. 2*, page(s). 217-225, March 2011.

Magen, N.; Kolodny, A.; U. Weiser,; Hamir, N. (2004) Interconnect-power dissipation in a microprocessor, *Proceedings of the 2004 International Workshop on System Level Interconnect Prediction, ser. SLIP '04*. New York, USA: Association for Computing Machinery, pp. 7–13. doi: 10.1145/966747.966750

Moaiyeri, M.H., Mirzaee, R.F., Doostaregan, A., ET AL.: (2013) A universal method for designing low-power carbon nanotube FET-based multiple-valued logic circuits, *IET Comput. Digit. Tech., 2013, 7, (4)*, page(s) 167-181.

Mohammadinodoushan, M.; Cambou, B.; Philabaum,C.; Hely, D.; Booher, D. D. (2019) Implementation of Password Management System Using Ternary Addressable PUF Generator, *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pp. 1-8, doi:10.1109/SAHCN.2019.8824792

Nancy Soliman, Mohammed E. Fouda , Abdullah G. Alharbi, Lobna A. Said, Ahmed H. Madian and Ahmed G. Radwan, (2019) Ternary Functions Design Using Memristive Threshold Logic, *IEEE Access (Volume: 7,2019).*

Nitish Kumar; Jialuo Chen; Monodeep Kar; Suresh K. Sitaraman; Saibal Mukhopadhyay; Satish Kumar (2019) Multigated Carbon Nanotube Field Effect Transistors-Based Physically Unclonable Functions As Security Keys*, IEEE Internet of Things Journal Volume: 6, Issue: 1, 2019.*

Ramzi A. Jaber, Abdalla Kassem, Ahmad M. El-Hajj, Lina A. El-Nimri, Ali Massoud Haidar.(2019) High-Performance and Energy-Efficient CNFET-Based Designs for Ternary Logic Circuits ,*IEEE Access PP(99):1-1, July 2019.*

S. Kim, T. Lim and S. Kang,; Seokhyeong Kang,(2018) An optimal gate design for the synthesis of ternary logic circuits*, 2018 IEEE 23rd Asia and South Pacific Design Automation Conference (ASP-DAC).*

Smith, K. C. (1988) Multiple-Valued logic: A Tutorial and Appreciation. *IEEE Computers Vol. 21*, page(s) 17-27.

S. Pi, C. Li, H. Jiang, W. Xia, H. Xin, J. J. Yang, and Q. Xia. (2018) Memristor crossbars with 4.5 terabits-per-inch-square density and two nanometer dimensions. *arXiv.org April 2018.*

Stathopoulos, S.; Khiat, A.; Trapatseli, M. et al. (2017) Multibit memory operation of metal-oxide bi-layer memristors. *Sci Rep 7, 17532*. doi:10.1038/s41598-017-17785-1.

Toshiba (2017) Toshiba Develops World's First 4-bit Per Cell QLC NAND Flash Memory. *www.techpowerup.com June 28, 2017*, Retrieved 19. December 2021.

Ware, Willis H. (1960). Soviet Computer Technology—1959.*Communications of the ACM, March 1960, Vol. 3 No. 3*, Page(s) 131-166.

Zhengyang He ; Kai Ren ; Jiayan Chen ; Xinyi Dai ; Zhao Pan; Yuejun Zhang (2018) Design of Delayed Ternary PUF Circuit Based on CNFET, *2018 24th Asia-Pacific Conference on Communications (APCC).*