CENICS 2014

The Seventh International Conference on Advances in Circuits, Electronics and Micro-electronics

ISBN: 978-1-61208-379-7

November 16 - 20, 2014

Lisbon, Portugal

CENICS 2014 Editors

Vladimir Privman, Clarkson University - Potsdam, USA

Pascal Lorenz, University of Haute Alsace, France

CENICS 2014

Foreword

The Seventh International Conference on Advances in Circuits, Electronics and Micro-electronics (CENICS 2014), held between November 16-20, 2014 in Lisbon, Portugal, continued a series of events initiated in 2008, capturing the advances on special circuits, electronics, and micro-electronics on both theory and practice, from fabrication to applications using these special circuits and systems. The topics cover fundamentals of design and implementation, techniques for deployment in various applications, and advances in signal processing.

Innovations in special circuits, electronics and micro-electronics are the key support for a large spectrum of applications. The conference is focusing on several complementary aspects and targets the advances in each on it: signal processing and electronics for high speed processing, micro- and nano-electronics, special electronics for implantable and wearable devices, sensor related electronics focusing on low energy consumption, and special applications domains of telemedicine and ehealth, bio-systems, navigation systems, automotive systems, home-oriented electronics, bio-systems, etc. These applications led to special design and implementation techniques, reconfigurable and self-reconfigurable devices, and require particular methodologies to be integrated on already existing Internet-based communications and applications. Special care is required for particular devices intended to work directly with human body (implantable, wearable, eHealth), or in a human-close environment (telemedicine, house-oriented, navigation, automotive). The mini-size required by such devices confronted the scientists with special signal processing requirements.

We take here the opportunity to warmly thank all the members of the CENICS 2014 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to CENICS 2014. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the CENICS 2014 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that CENICS 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of circuits, electronics and micro-electronics.

We hope Lisbon provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

CENICS 2014 Chairs:

Vladimir Privman, Clarkson University - Potsdam, USA

Sergey Y. Yurish, Technical University of Catalonia (UPC-Barcelona), Spain

Martin Horauer, University of Applied Sciences Technikum Wien, Austria
Adrian Muscat, University of Malta, Malta

CENICS 2014 Research/Industry Chairs
Ravi M. Yadahalli, PES Institute of Technology & Management - Karnataka, India

CENICS 2014 Industry Liaison Chairs
Falk Salewski, Muenster University of Applied Sciences, Germany

CENICS 2014 Publicity Chair
Sandra Sendra Compte, Universidad Politécnica de Valencia, Spain

CENICS 2014 Special Area Chairs
Formalisms
Peeter Ellervee, Tallinn University of Technology, Estonia
Application-oriented
Josu Etxaniz Marañon, University of the Basque Country / Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain
Sensors
Yulong Zhao, Xi'an Jiaotong University, China

CENICS 2014

Committee

CENICS Advisory Chairs

Vladimir Privman, Clarkson University - Potsdam, USA
Sergey Y. Yurish, Technical University of Catalonia (UPC-Barcelona), Spain
Martin Horauer, University of Applied Sciences Technikum Wien, Austria
Adrian Muscat, University of Malta, Malta

CENICS 2014 Research/Industry Chairs

Ravi M. Yadahalli, PES Institute of Technology & Management - Karnataka, India

CENICS 2014 Industry Liaison Chairs

Falk Salewski, Muenster University of Applied Sciences, Germany

CENICS 2014 Publicity Chair

Sandra Sendra Compte, Universidad Politécnica de Valencia, Spain

CENICS 2014 Special Area Chairs

Formalisms
Peeter Ellervee, Tallinn University of Technology, Estonia

Application-oriented
Josu Etxaniz Marañon, University of the Basque Country / Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain

Sensors
Yulong Zhao, Xi'an Jiaotong University, China

CENICS 2014 Technical Program Committee

Amr Abdel-Dayem, Laurentian University, Canada
Amir Shah Abdul Aziz, TM Research & Development, Malaysia
Adel Al-Jumaily, University of Technology, Sydney
Said Al-Sarawi, The University of Adelaide, Australia
Mohammad Amin Amiri, Iran University of Science and Technology, Iran
Lotfi Bendaouia, ETIS-ENSEA, France
Yngvar Berg, Vestfold University College, Norway
Madhu Bhaskaran, RMIT University, Australia
Manuel José Cabral dos Santos Reis, University of Trás-os-Montes e Alto Douro, Portugal
Javier Calpe, University of Valencia, Spain

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Novel High Speed and Robust Ultra Low Voltage CMOS NP Domino Carry Gate

Abdul Wahab Majeed*, Halfdan Solberg Bechmann †, Yngvar berg ‡
Departments of Informatics
University of Oslo, Oslo, Norway
*Email: abdulwm@ifi.uio.no
†Email: halfdasb@ifi.uio.no
‡Email: yngvarb@ifi.uio.no

***Abstract*—In this paper, a novel design of an Ultra Low voltage Carry Gate shall be presented. The main objective is to target the robustness of the presented ciruits. We shall also imply as to what extent these circuits can be improved and what their benefits, compared to conventional topologies, are. The design presented, compared to a conventional CMOS carry gate, is area efficient and high speed. The relative delay of a ULV carry gate lies at less than 3% compared to conventional CMOS carry gate. The circuits are simulated using the TSMC 90nm process technology and all transistors are of the Low Threshold Voltage (lvt) type.**

***Index Terms*—ULV; Carry Gate; NP domino.**

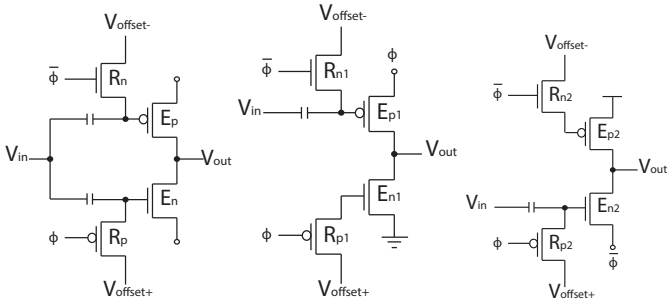
I. INTRODUCTION

As the semiconductor industry grows, the demand for Ultra Low Voltage (ULV) circuits is increasing. These circuits are being implemented in VLSI where different kind of functions are combined on one chip. The Arithmetic Logic Units (ALU)s are one of the many circuits that are implemented in the VLSI chips. Since an adder is an important part of the ALU, the speed of the adder used, is important for the ALU performance. The speed of the adder is determined by the propagation delay of the carry chain. Although high speed conventional carry circuits like Carry Look Ahead, Dual rail domino carry, CPL, etc., are well established design topologies, their performance at ULV suffers from degradation [1]. Several approaches are proposed for the improvement of performance [2][3] but the design presented in this paper is influenced by [4]. This paper shall present a new high speed NP domino ULV carry design. To highlight the improvement, the results shall be compared to conventional domino design such as Dual Rail Domino carry. Both the carry circuits are implemented in a 32-bit carry chain in order to show as to what extent one is better than the other, regarding their speed and power.

Section I-A presents a general introduction to the ULV circuits presented in [5]. Section II presents different configurations of ULV carry designs and gives an explanation on how it works. Section III presents the performance of the proposed ULV carry gate compared to the conventional carry gate.

A. ULV Inverter

1) Evaluation and Precharge Phase: A simple ULV inverter model is presented in the Figure 1a. A ULV Semi-Floating gate circuit design consists of two phases, an evaluation phase, determined by the evaluation transistors E_n and E_p, and a precharge phase determined by the precharge transistors R_n and R_p. As seen in the Figure 1a $\overline{\phi}$ is applied to R_n and



(a) A simple SFG-ULV inverter (b) P-type inverter (c) N-type inverter

Fig. 1. SFG NP ULV domino inverter

ϕ is applied to R_p. In such a circuit, the precharge phase occurs when ϕ=0 and the circuit enters evaluation phase when ϕ=1. During the precharge phase, the input floating nodes are charged to a desired level i.e logical 1 or V_{DD} for the E_n floating gate and logical 0 or Ground (GND) for the E_p floating gate. No input transition occurs during the precharge phase. However, once the clock shifts from logical 0 to 1 and has reached a stable value of 1, an input transition may occur, which determines the logical state of the circuit's output. We can configure the circuit in an NP domino fashion by engaging E_n to $\overline{\phi}$ (where $\overline{\phi}$=1 during the precharge phase) and E_p to V_{DD}. Such a configuration yields a precharge level of logical 1 and is called an N-type circuit. On the other hand, if we engage E_n to GND and E_p to ϕ we can obtain a precharge level of 0. Such a configuration is called a P-type circuit.

Considering the example of N-type inverter, we know that the output of N-type is precharged to 1. Once ϕ shifts from 0 to 1, circuit enters the evaluation phase. During the evaluation phase, there are two possible situations. If no input transition occurs, the output shall remain unchanged and hold its value to 1. Indicating that no work is to be done. However, if an input transition occurs and input is brought to 1 the E_{n2} shall be turned on and the output shall be brought to logical 0 or close to 0. This indicates that the only work to be done during the evaluation phase is to bring the output from 0 to 1 when an input transition occurs.

We have seen that the only work that is to be done, during the evaluation phase, is to bring the output to the logical 0 when an input transition occurs. This suggests that E_{p2} does

 ISBN: 978-1-61208-379-7

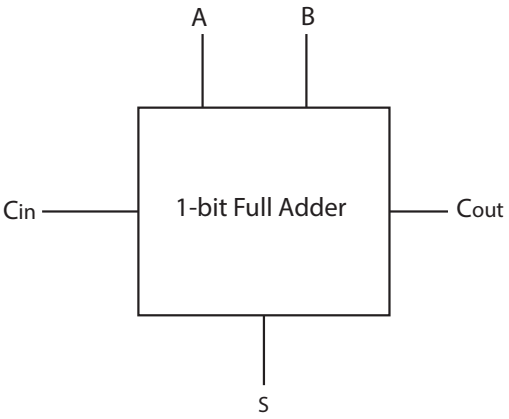


Fig. 2. 1 bit full adder

not require an input transition at any stage. Therefore, we can remove the input capacitor of E_{p2}. Such a configuration can be called pseudo SFG ULV inverter and is shown in Figure 1c. An equivalent P-type Pseudo SFG ULV inverter is shown in Figure 1b. This will lead to load reduction and hence higher speed. However, we may encounter some robustness issues with respect to noise margin due to leakage current.

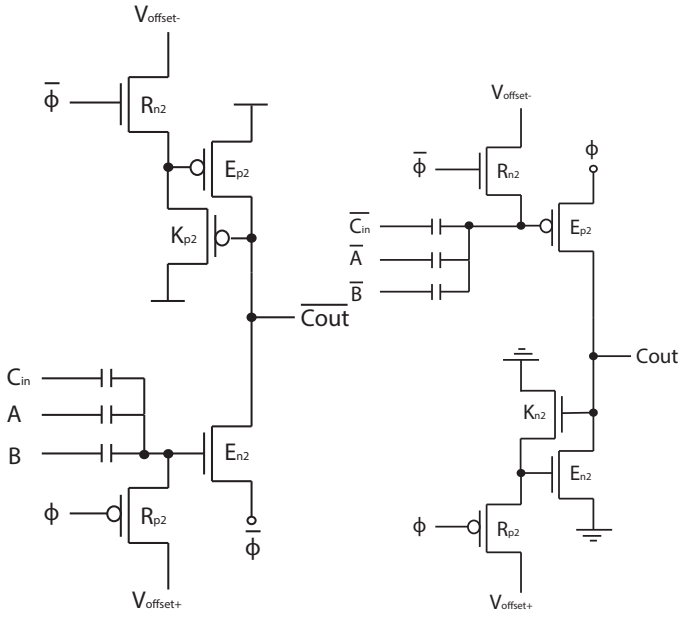
II. Methods

$$C_{out} = A \cdot B + (C_{in} \cdot (A \oplus B)) \quad (1)$$

The output of a carry circuit is generated using two inputs and a carry bit from the previous stage, if available (carry bit at the least signficant bit is always zero so it has no previous carry), as shown in figure 2. Equation (1) shows an arithmetic approach to carry generation where A and B is the input signal and C_{in} is the carry bit from the previous stage. There are two parts of this equation, one is generated internally, $A \cdot B$, and can be called carry generation (CG), the other one is dependent on the carry bit from the previous stage, $(C_{in} \cdot (A \oplus B))$, and is known as carry propagation (CP). The speed of any carry chain depends on the second part of this equation, because it has to wait for the carry bit from the previous stage to arrive. Inputs A and B both arives simultaneously at any stage of an N bit carry chain. Most conventional designs use two seperate parts for CG and CP but the design presented in this paper differ from the most designs as it is able to generate both CG and CP by applying all the inputs to a single transistor. This technique is called Multiple valued Logic (MVL) where classical truth value, logical 1 and 0, are replaced by finit or infinite logical values. It has a potential to decrease the chip area and total power dissipation[6].

A. Non-Differential Carry Gate

The Static Ultra Low Voltage Carry (SULVC) is a modified version of the ULV N-P domino inverter shown in section I-A. The carry circuit uses a keeper as proposed in [5] and



(a) N type SULVC (b) P type SULVC

Fig. 3. SULVC circuit

3 capacitors in parallel at the input gate providing the input logic for the circuit. The circuit is designed to make the A and B signal cancel each other out when A and B have contrasting values to allow the carry input signal to determine the carry output in this case. Because of the cancellation requirement between the A and B signals they need to arrive as equally sized rising or falling transitions, this can be acheived by utilizing level-to-edge converters or a logic style with a VDD/2 precharge level.

If both A and B are rising, the floating node will rise causing a falling transition on the carry output of the N-type circuit regardless of the carry input signal. If they are both falling, the carry input signal can not elevate the floating node voltage enough to cause a transition, leaving the carry output at precharge level. If A and B are not equal, their two transitions cancels each other out and the floating node remains at precharge level until a possible rising edge occurs on Cin. A P-type equivalent of the circuit is shown in Figure 3 (b) where all signals and logic are the inverse of those in the N-type circuit. For both circuits, a transition on the output indicates carry propagation and they can both be characterized as a carry generate circuit corresponding with the truth table shown in Table I, the transition logic for the N-type circuit can be seen in Table II.

During the precharge phase, the voltage level of the floating node is set to ground for the P-type circuit and V_{DD} for the N-type and can only be changed by the inputs through the capacitors in the evaluation phase. In these circuits, when used in CPAs (Carry Propagate Adder), C_{in} can arrive later than A and B when the carry bit has to propagate through the chain of carry circuits. This introduces the challenge of keeping the output precharge value during the evaluation phase in case no

 ISBN: 978-1-61208-379-7

TABLE I. TRUTH TABLE FOR A CARRY CIRCUIT

Inputs			Output
A	B	C_{in}	C_{out}
0	0	0	0
0	1	0	0
1	0	0	0
1	1	0	1
0	0	1	0
0	1	1	1
1	0	1	1
1	1	1	1

TABLE II. TRANSITION TRUTH TABLE FOR N-TYPE SULVC

Inputs			Output
A	B	C_{in}	$\overline{C_{out}}$
↓	↓	0	1
↓	↑	0	1
↑	↓	0	1
↑	↑	0	↓
↓	↓	↑	1
↓	↑	↑	↓
↑	↓	↑	↓
↑	↑	↑	↓

carry signal arrives. As Figure 4 shows, the floating node of the P-type circuit is precharged to 0V. This causes the transistor E_{p2} in Figure 4 (b) to conduct and the output will drift and may eventually cause an incorrect output value as shown in Figure 5 at 70ns. The drifting effect is countered with the K_{n2} and K_{p2} keeper transistors but the effect limits the length of the evaluation pase and therby the number of carry circuits that can be put in a chain and the maximum number of bits an adder based on the circuit can process in one clock cycle. The maximum achieved number of bits acheived varies with the supply voltage as shown in Figure 6 and at 300mV a 32-bit carry chain can be implemented.

The transistor sizing is adjusted to accommodate the change in NMOS/PMOS mobility difference with changed supply voltage. In these simulations the NMOS evaluation transistor size is kept minimum sized and the PMOS evaluation transistor

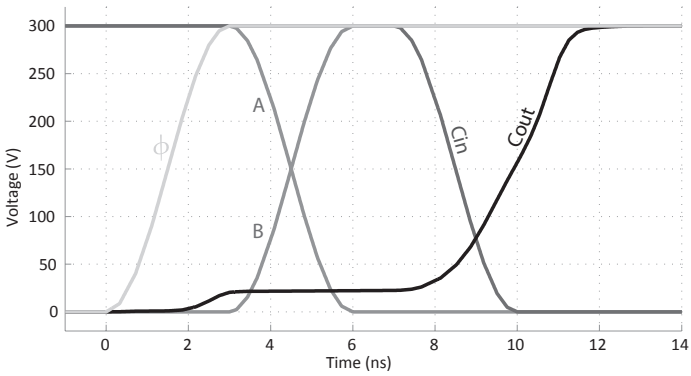


Fig. 4. *Carry input and output for SULVC gate. Supply voltage at 300mV*

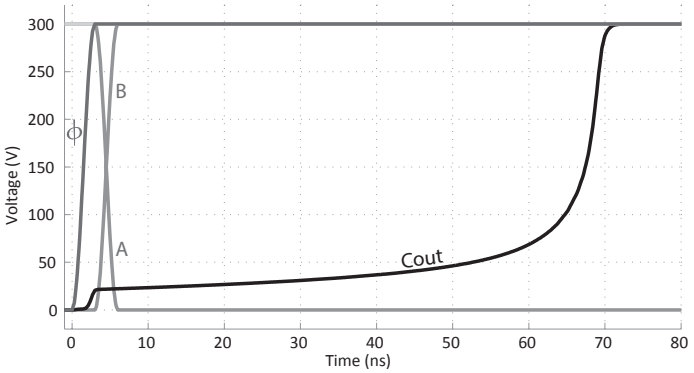


Fig. 5. Drifting problem of the SULVC output.

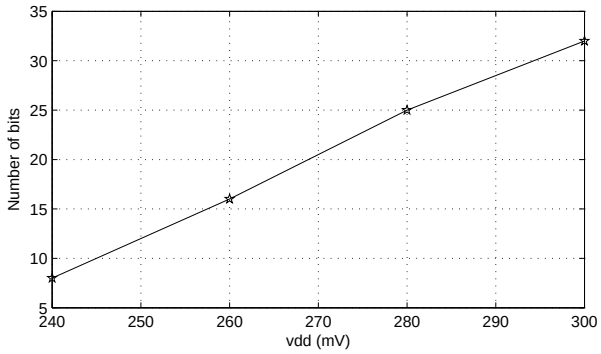


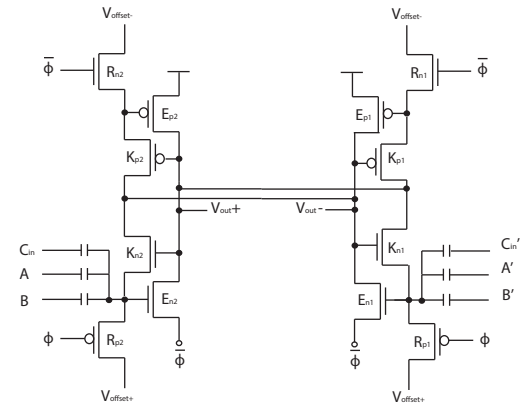
Fig. 6. Number of carry circuits or bit obtained from carry chain when supply voltage is varied

length is changed to match the NMOS drive strength.

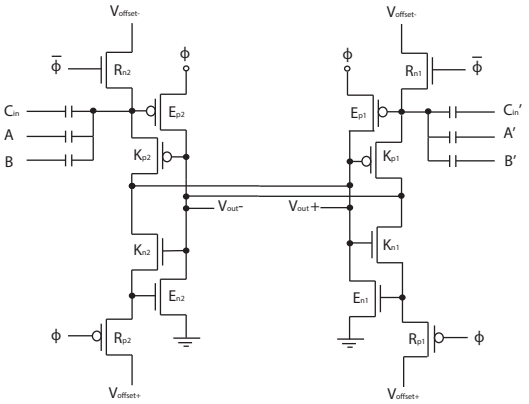
B. Differential Carry Gate

In order to overcome the challenges with robustness and drifting of the SULVC circuit, a differential approach is a possible solution. A Static Differential Ultra Low Voltage Carry (SDULVC) as shown in Figure 7 is designed in exactly the same manner as the SULVC, however with differential inputs and outputs. The differential nature of the circuit makes it less prone to drifting and eliminates the need for level-to-edge converters it can be sized to allow a single edge without causing an output transition. The outputs of the proposed circuit are precharged to the same level during the precharge phase, however it yields a differential output during the evaluation phase. So, instead of employing an inverter to obtain the carry bit we can read it from the opposite end of the circuit, i.e. in an N-type SDULVC if inputs A B and C are applied to E_{n2} output can be read from V_{out-}. Figure 7 demonstrates the design of an SDULVC circuit. The backgate of the keeper transistors of these circuits are connected to the floating gate to achieve maximum robustness.

$$V_{fg} = V_{initial} + k_{in} \cdot V_{in} \text{where } k_{in} = \frac{\sum_{i=1}^{n} C_{innHigh_i}}{C_{total}} \quad (2)$$



(a) N type SDULVC



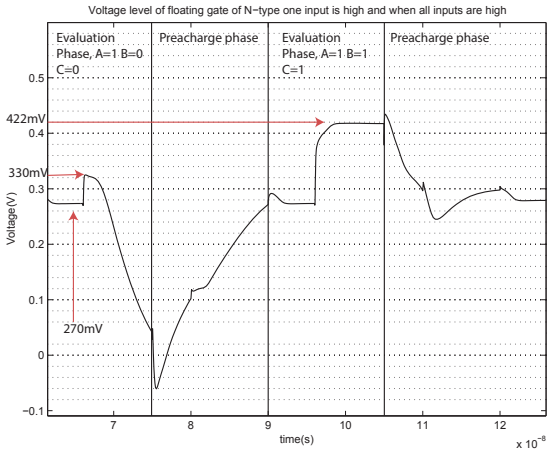
(b) P type SDULVC

Fig. 7. SDULVC circuit

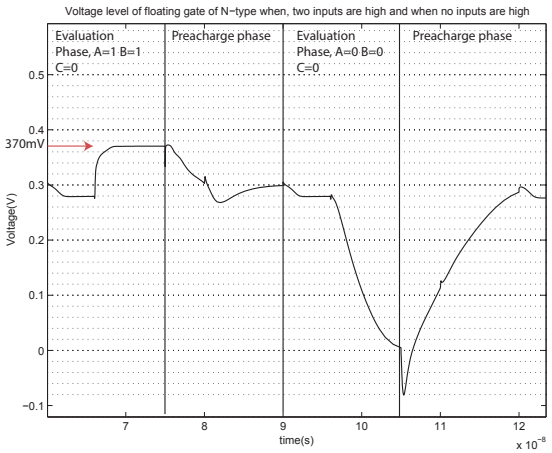
The variable 'i' in (2) denotes the index of the input and the 'n' denotes fan-in. $V_{initial}$ is the precharge voltage level of the floating gate. C_{inHigh} is a combination of input capacitors with a high (rising) input.

Considering an example of an N-type SDULVNC we can calculate the voltage level of the floating gate using (2). We assume that the diffusion capacitance is equal to the input capacitance and that the supply voltage is equal to the input voltage. The load capacitance introduced by the keeper's back-gate connection to the floating node should also be considered and is in this paper assumed to be equal to the input capacitor as well.

Our calculation in (2) gives us a theoretical idea of the voltage level at V_{fg} (floating gate voltage). In the real world the capacitance size might not be exactly the same as our assumption and depends on transistor size and many other factors like process variation and mismatch. The simulation results of the voltage levels for the floating gate in the Figure 8 shows that the floating node is precharged to 270mV. Equation (2) yields an analytical result for the floating gate input of 330 mV, 390mV and 450 mV for one, two and three high inputs, respectivly. The simulation results in Figure 8a shows that the voltage level of the floating node gets to 330mV for a single rising input transition and to 420mV when all inputs are high. These results are marginally different from the calculated



(a) Voltage level of input floating gate of an N-type SDULVC/SDULVC when A=1 B=0 and C=0, and when A=1 B=1 C=1



(b) Voltage level of input floating gate of an N-type SDULVC/SDULVC when A=1 B=1 and C=0, and when A=0 B=0 C=0

Fig. 8. Voltage level of floating gate of N-type at supply voltage of 300mV

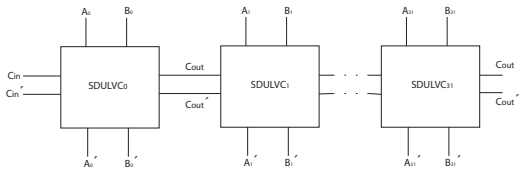


Fig. 9. 32 bit ULV carry chain

values. This is possibly due to the assumptions on capacitance sizes. Figure 8a shows that if only one input gets high, the keeper transistor turns on and discharges the floating node. The reason for this is that the transition at the input, i.e. 60mV, is not sufficient to produce enough current at the output. Figure 8b shows the results for two high inputs and all low inputs.

 ISBN: 978-1-61208-379-7

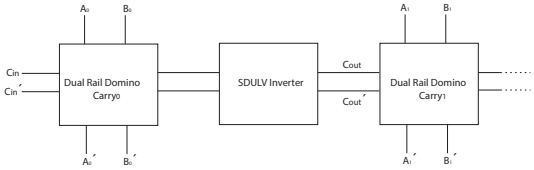
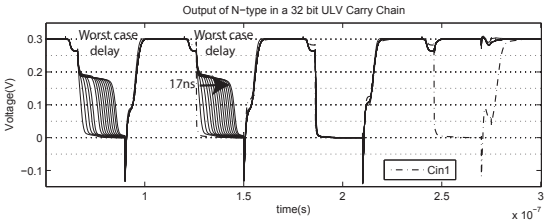
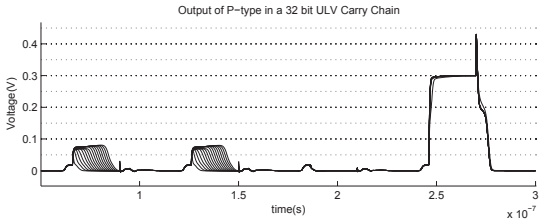


Fig. 10. Implementation of hybrid Dual rail domino carry



(a) output ULV carry chain N-type



(b) output ULV carry chain P-type

Fig. 11. Simulation result of 32 bit ULV carry chain at a supply voltage of 300mv

III. Simulation

A. 32 bit SDULVC chain

A 32 bit ULV carry chain is implemented using 32 SDULVC circuits connected in a chain or NP domino fashion shown in Figure 9. Figure 11 shows the simulation response of a 32 bit ULV carry chain. The propagtion delay of this carry chain is 17ns. In order to compare the SDULVC to other carry gate topologies, a dual rail domino carry gate designed in a hybrid fashion, i.e. instead of utilizing conventional inverters at the output, the Static Differential ULV inverter presented in [7] and a conventional NP Domino Dual Rail carry is used. Compared to the hybrid dual rail domino carry (HDRDC) chain shown in the Figure 10 the SDULVC chain is almost 10× faster and compared to a Conventional Dual Rail Domino Carry (CDRDC) this is closer to 35×. These numbers are based on the propagation delay for the carry bit through the chain, which is 166ns for the hybrid dual rail domino carry and 636ns for the conventional dual rail domino carry, all at 300 mV.

The robustness of the SDULVC can be analyzed by looking at the simulation response shown in Figure 11b. The plot for the worst case delay scenario, i.e. A=1, B=0, C=0, exhibits that due to a delayed carry bit and the early arrival of inputs, A and B, a marginal transition at the output occurs. However, once the carry bit has arrived, the output shifts to its final value. Average transition at the output for a P-type and N-type SDULVC when waiting for the carry bit is between 70mV and 100mV. This can be seen as a problem for the noise margin and power consumption. The output manages to return to the right final value due to synchronisation of keeper signals with the input. Therefore, the issue of noise margin can be ignored by concluding that the final value can be read at the end of the evaluation phase.

TABLE III. DIMENSIONS OF HYBRID DUAL RAIL DOMINO CARRY GATE

	Supply voltage variation	Width of dual rail domino evaluation transitor/Width of SDULVN evaluation transitor	Length of dual rail domino evaluation transitor/Length of SDULVN evaluation transitor	Width of dual rail domino precharge transitor/Width of SDULVN precharge transitor	Length of dual rail domino precharge transitor/Length of SDULVN precharge transitor
Size 1	270mv-400mv	4×	3.3×	1×	1×
Size 2	220mv-400mv	6.67×	8.3×	3.33×	35×

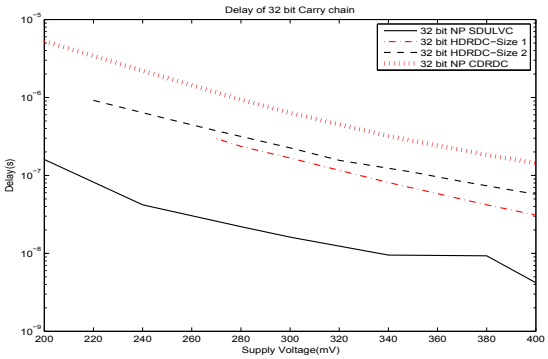


Fig. 12. Delay of 32 bit SDULVC and hybrid dual rail domino at varried supply voltage

Figure 12 shows the delay of an SDULVC chain compared to an HDRDC and a CDRDC chain. Table III shows that the transistor size has to be increased in order to increase the ON current of the device [8] and be able to decrease the supply voltage for HDRDC. Table IV shows the minimum operating frequency required for the clock to simulate SDULVC, HDRDC and CDRDC at different supply voltages.

B. PDP and EDP of SDULVC chain

PDP charachteristics of a circuit highlights its efficiency with respect to power consumtion. A low PDP means a more

TABLE IV. TABLE IV: MINIMUM CLOCK OPERATING FREQUENCY FMIN REQUIRED BY THREE TOPOLOGIES

Supply Voltage (mV)	f_{min} for SDULVC (MHz)	f_{min} for HDRDC-Size 1 (MHz)	f_{min} for HDRDC-Size 2 (MHz)	f_{min} for CDRDC
200	1.6	-	-	0.08
220	-	-	-	-
240	3.125	-	-	0.217
250	-	-	0.83	-
270	-	1.66	1.225	-
280	6.25	-	-	0.5
300	8.33	2.3	2	0.769
320	-	-	2.27	-
340	16.66	5.5	3.33	1.562
380	21	10	5.55	2.5
400	23.8	60	7.692	3.33

 ISBN: 978-1-61208-379-7

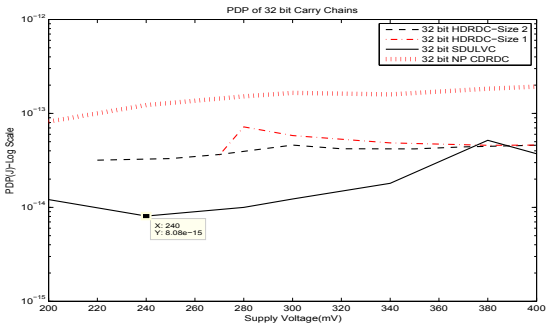


Fig. 13. PDP of 32 bit carry chains

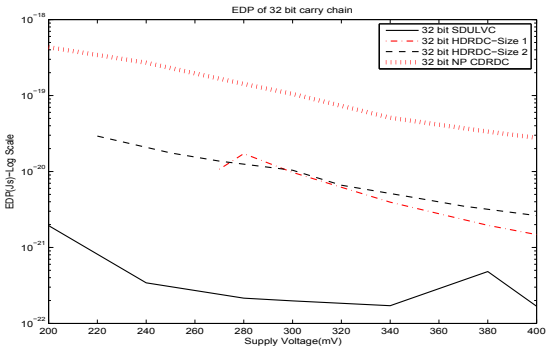


Fig. 14. EDP of 32 bit carry chains

energy efficient circuit. Although the ULV circuits presented in this paper are power hungry, it still manages to maintain its PDP at approximately the same level as conventional circuits where the power consumption is lower. The average power of the HDRDC and the SDULVC is $0.347\mu W$ and $1.28nW$ respectively at a supply voltage of 300 mV. This indicates that the power consumption of HDRDC is up to $3\times$ better than ULV circuits. However, at the same supply voltage the ULV circuit is $10\times$ faster than the HDRDC. Therefore, the ULV circiuts are still more energy efficient. Figure 13 shows PDP of three different 32 bit carry chain topologies at varied supply voltage. The minimum energy point of the 32 bit SDULVC carry chain is found at 240 mV.

Another important charachteristic of any circuit is EDP. It demonstrates enhanced speed of any circuit with respect to its energy effieciency. It is obvious that circuits with better propagation delay shall stand out in this characteristic. Figure 14 shows the EDP of three carry chains and the evident performance advantages of SDULVC circuits.

IV. Conclusion

In this paper, a new ULV carry circuit has been presented and performance enhancements have been demonstrated. The ULV carry circuits are better than conventional topologies in both speed and energy efficiency, shown by comparing the SDULVC to the HDRDC and CDRDC circuit topologies. A credible conclusion is that a static differential dynamic ULV carry circuit is a favorable choice when speed and robustness at low voltages are important.

References

[1] M. Alioto and G. Palumbo, "Impact of supply voltage variations on full adder delay: Analysis and comparison," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 14, no. 12, pp. 1322–1335, 2006.

[2] ——, "Very high-speed carry computation based on mixed dynamic/transmission-gate full adders," in *Circuit Theory and Design, 2007. ECCTD 2007. 18th European Conference on*, 2007, pp. 799–802.

[3] Y. Berg and O. Mirmotahari, "Ultra low voltage and high speed cmos carry generate circuits," in *Circuit Theory and Design, 2009. ECCTD 2009. European Conference on*, 2009, pp. 69–72.

[4] Y. Berg, "Ultra low voltage static carry generate circuit," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 2010, pp. 1476–1479.

[5] Y. Berg and O. Mirmotahari, "Ultra low-voltage and high speed dynamic and static cmos precharge logic," in *Faible Tension Faible Consommation (FTFC), 2012 IEEE*, june 2012, pp. 1 –4.

[6] Y. Berg, S. Aunet, O. Naess, O. Hagen, and M. Hovin, "A novel floating-gate multiple-valued cmos full-adder," in *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*, vol. 1, 2002, pp. I–877–I–880 vol.1.

[7] Y. Berg and O. Mirmotahari, "Static differential ultra low-voltage domino cmos logic for high speed applications."

[8] M. Alioto, "Ultra-low power vlsi circuit design demystified and explained: A tutorial," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 59, no. 1, pp. 3 –29, jan. 2012.

 ISBN: 978-1-61208-379-7

Simultaneous Measurement of Temperature and Expansion on Radio Frequency Power Electronic Components

Eric Joubert, Olivier Latry
GPM UMR 6634
Université de Rouen - France
eric.joubert@univ-rouen.fr
olivier.latry@univ-rouen.fr

Jean-Philippe Roux
CEVAA - France
jp.roux@cevaa.com

***Abstract*—This paper presents a new approach for measuring physical variables on micro-electronic components. An optical system is used to simultaneously quantify the surface temperature of a component and its expansion. This double acquisition is achieved by a Michelson interferometer coupled with a Charge Coupled Device (CCD) line device. To validate this method, the temperature measurements were directly compared with the results obtained by an infrared camera and by a measurement of variation of I(V). The displacement measurements were compared with those obtained by a laser 3D vibrometer, whose physical principle is completely different. Consistent results were obtained regarding the different techniques.**

Keywords-thermal measurement; laser; electronic components.

I. INTRODUCTION

Modern telecommunications and radar systems require the development of increasing Radio Frequency (RF) power. The new generation of electronic power transmitters uses solid state technology (GaAs, LDMOS, GaN) transistors. To properly manage parameters, such as size, weight and cooling requirements, these components have to improve many other parameters, such as electrical efficiency and power density while maintaining a high level of reliability and miniaturization. Because of these constraints, the behavior of these components that are subject to high thermal stress, is increasingly difficult to characterize. There is currently a high demand to fully understand the impact of thermal parameters on overall performance and lifetime. This is why it has become essential to develop measurement tools that characterize these kinds of phenomena.

In literature, many technologies are very efficient to extract different types of parameters separately. Some methods use a measurement of V_{be} on a small cycle of the signal to extract the average junction temperature [2]. Others place thermal sensors directly in power Print Circuit Board [5]. The micro-Raman spectroscopy combined with infrared technology provides accurate thermal values [1][7]. Using nematic liquid crystal thermography is very efficient for specific emissive components [4]. Thermoreflectance imaging using specific wavelengths measurement allows thermal changes measurements over large time scales [8]. It is possible to reach very high thermal resolutions with an Optical Time Domain Reflectometer coupled with a Michelson interferometer [3]. Finally, speckle interferometry provides data on the physical deformation of a component subjected to thermal stress [6].

In this state of the art, all the methods aim at providing unique physical data for a device. However, integrating temperature and expansion makes it possible to estimate the dilatation factor of the constitutive material of a power electronic device. In fact, this parameter provides valuable information in the study of the degradation of electronic component under stress.

This paper also suggests a method for measuring simultaneously temperature and expansion of micro-electronic components. The first part presents the optical structure of our bench and the physical data measured to estimate temperature and expansion. The second part describes the particular optical calibration process. The third part details the specific sensor we used to transfer the interferometric information to the acquisition device. At last, test devices and results obtained are proposed and discussed in relation with other thermometric/expansion methods.

II. OPTICAL SYSTEM

The system is based on a Michelson interferometer [6], as presented in Figure 1. The beam produced by a laser is split into two separate paths. The measuring path uses the device under test and the reference path is reflecting on a plane mirror. The interferometric pattern, which is the sum of the two branches, passes thereafter through a long focal lens.

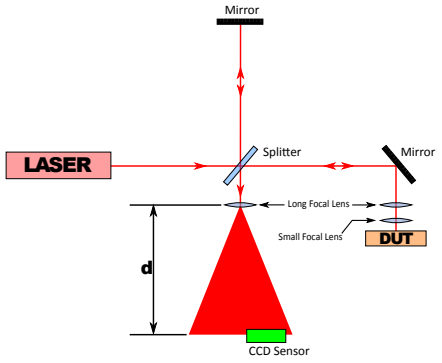


Figure 1. Structure of the interferometer.

 ISBN: 978-1-61208-379-7

A Charge Coupled Device analyses the interference pattern, contrary to conventional interferometers that are using a single point measurement. This feature allows to measure two parameters: the displacement and amplitude of the interference fringes. This makes our method completely unique. Thereafter, these parameters are used to determine respectively the expansion and temperature of the component (see Figure 2).

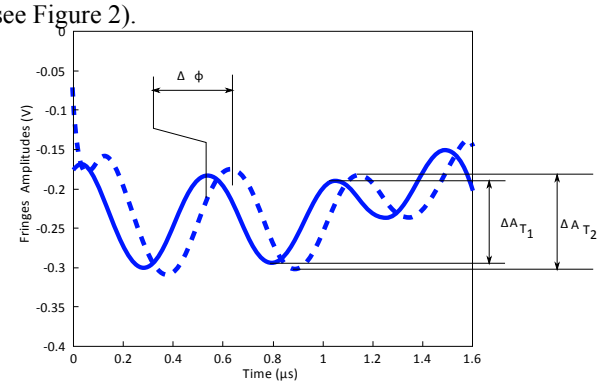


Figure 2. Fringes changes according to expansion and temperature.

The distance between the lens and the sensor is adjusted to take the inter-pixel distance into account. A polynomial interpolation is applied to ensure an efficient filtering of the recovered electrical signal (see Figure 3).

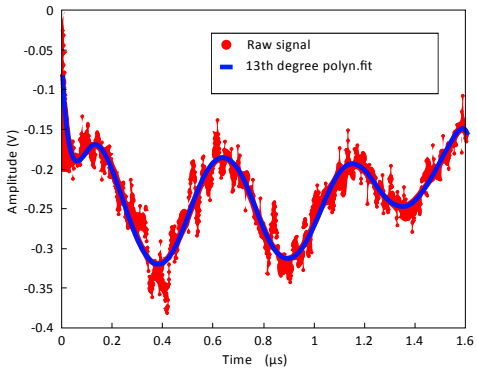


Figure 3. Fitting to 13th degree polynomial

A. Temperature measurement

The temperature is calculated from the amplitude of the interference fringes obtained. Indeed, the reflectance of the material depends on the temperature encountered through its refractive index according to the equation:

$$R(T) = \left(\frac{n_1 - n_2(T)}{n_1 + n_2(T)} \right)^2 \quad (1)$$

where $R(T)$ denotes the reflection coefficient, n_1 and n_2 are the refractive indices of the materials 1 and 2, respectively. Under these conditions, the intensity for each point of the interference pattern (perpendicular to the optical axis z) is affected by the reflection coefficient:

$$I(x,y) \propto A_1^2(x,y) + R(T).A_2^2(x,y) + 2..A_1(x,y).\sqrt{R(T)}.A_2(x,y) .\cos\left(\phi_2(x,y) - \phi_1(x,y)\right) \quad (2)$$

where $\phi 1\ (x, y)$ and $\phi 2\ (x, y)$ are the local phases of the waves 1 and 2. The amplitude of the interference (the third term of equation 2) is increasing linearly as a function of the reflection coefficient. By measuring changes of contrast of the interference fringes, or measuring the peak to peak values of the fringes, it is also possible to measure the temperature of the component.

B. Expansion measurement

The expansion of the component is determined by measuring the displacement of the fringes of the interference pattern. The phase shift induced by the expansion causes a shift in the interference pattern. The circular fringes move outwardly in the case of expansion and inwardly in the case of a contraction.

For a contraction of 10µm, the fringes move a few millimeters. Simultaneously, these fringes have their width changed. The expansion can either be determined by observing the displacement or the width of the fringes.

III. CALIBRATION METHOD

The system is calibrated with a raw wafer of silicon. The ideal surface of the material allows an easy implementation of the interferometer system.

The wafer is thermally excited by a Peltier module. The temperature induced by this module is directly measured by the voltage *Up* at its terminals. Indeed, the temperature difference ΔT provides by the Peltier module is in relation with its electrical characteristics according to the equation:

$$\Delta T = \frac{U_p - n.R.I_p}{n.\alpha} \quad (3)$$

where n is the number of cells in the module, R the resistivity of the module, *Ip* the module current and α the Seebeck coefficient.

The expansion is calibrated by comparing it with that obtained by an industrial laser Doppler vibrometer at a frequency in convenience with the characteristics of the interferometer (frequency not exceeding a few Hz).

IV. ACQUISITION

During the thermal excitation, the cross-sections of interference pattern are acquired by the CCD line sensor. This sensor transfers 1024 points within a period of 1.7 ms. The analog video signal is converted by a *National Instrument* acquisition unit that has a depth of 24 bits and a

 ISBN: 978-1-61208-379-7

sampling rate of 100kS/s. The recovered data are thereafter processed with Labview [9] to compute displacements and amplitudes of fringes at the maximum rate of the acquisition unit.

Speckle phenomenon in interference pattern brings noise in video signal and reduced overall measurement accuracy. A polynomial fitting of 13th degree is first performed to improve the quality of the signal. A detection algorithm follows this filtering to determine the peak amplitude of the fringes. The temperature is determined from this peak after a calibration process.

Measuring displacement of the fringes (and also the expansion of the component) is more complex, because fast moving fringes can be processed as appearing or disappearing. These changes have an important impact on the result. A suitable algorithm allows such a monitoring. The overall process is not real-time: the data are processed after acquisition for avoiding slowing acquisition with computing times.

V. TEST DEVICES

Two test components were investigated in order to highlight the characteristics of the temperature measurement on the one hand, and of the expansion on the other hand.

The component used for temperature measurements is an integrated preamplifier that was opened by laser ablation and chemical attack. This component comprises an ElectroStatic Discharge (ESD) diode that allows in situ measurement of the component temperature. As it is a complex integrated circuit, an externally controlled Peltier cell [10] was used to heat the component.

The component used for the expansion action was a 14x3cm aluminum blade simulating a real board used in radar systems. The size of this board allowed to use a miniature shaker.

VI. RESULTS

A. Temperature

The temperature measurements on the test component were validated by two measurement systems: an internal measurement with an ESD diode and a measurement carried out by a dedicated Infra-Red camera.

The ESD diode was polarized to a constant current so as to generate a voltage near the threshold voltage. This bias point limits the self-heating of the diode. Locally, the changes in the voltage V are a quasi-linear function of the temperature T in ° C:

$$T = 1280,6 \times V - 156,43 \quad (4)$$

This equation is corresponding to the calibration curve on the real component, obtained by comparison with a thermocouple (see Figure 4).

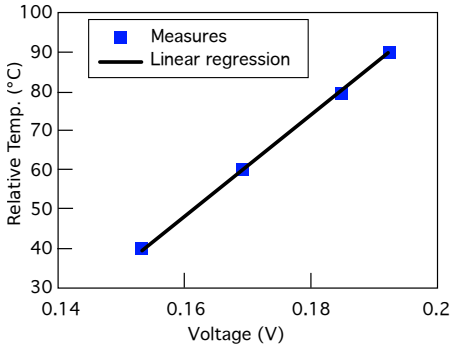


Figure 4. T(V) with constant current for an ESD diode

The second set of measurements was performed using an infrared camera for electronic components. The camera was corrected for emissivity at 20°C.

The interferometer measurement was not performed simultaneously due to the lack of space above the device under test. However, the physical conditions (heat rise) were identical.

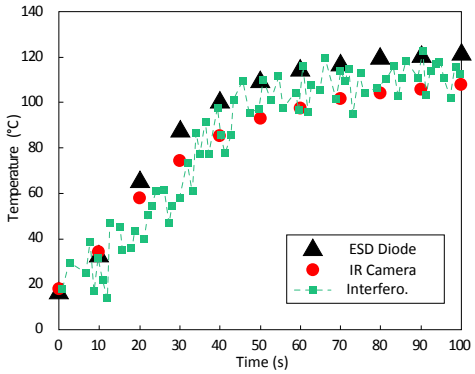


Figure 5. Result for a long temperature rise with the three methods

These results show a good correspondence between the three techniques, even if the interferometer has a measurement uncertainty (see Figure 5). These errors (±10°C) are due to measurement conditions that were not optimal on the component under test. A better mechanical isolation should limit these effects. Moreover, one must keep in mind that the detection area is very small for the interferometer, whereas detection areas for the other techniques are larger. A larger area obviously brings an averaging effect. The error of 15°C at 100s between IR camera and ESD diode is due to an emissivity compensation problem. This gap on the infrared camera could be removed efficiently by using black paint on the component to compensate the emissivity of the emitting surface.

B. Expansion

The goal here was to measure the ability of the interferometer to reach high frequencies and thus to provide a possible recovery operation range with Polytech vibrometer PSV400.

 ISBN: 978-1-61208-379-7

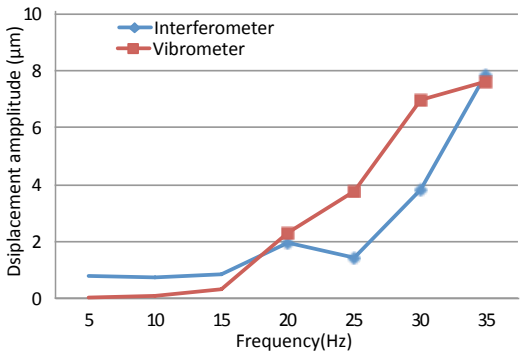


Figure 6. Comparison of the amplitudes measured with the two systems

Comparative measurements were made sequentially to solve practical issues of space above the component to be tested. For each measurement, conditions were the same: aluminum board was subjected to a shaker controlled by 1KHz sine signal bursts whose repetition frequency varied between 5Hz and 35Hz. The resulting signals were processed by time FFT and compared in Figure 6.

This figure shows the results for the two systems are quite consistent with the curve of mechanical response of the shaker. The average uncertainty of the vibrometer depends on working frequency and is about 1% on the frequency range 5hz-35Hz. Beyond this overall trend, both techniques show very similar results, despite very important differences in their physical principle. In fact, the interferometer is limited at high frequencies by the sampling due to CCD sensor. In the same time, the Doppler vibrometer principle implies that a minimum frequency is required for the detection to take place.

VII. CONCLUSION

The bench that we built enables to simultaneously acquire surface temperature and expansion of an electronic component. These measurements are made on the fringes of the interference pattern resulting from a Michelson interferometer.

In this pattern, two important parameters are extracted. On the one hand, the displacement of the fringes relative to the center of the interference pattern provides useful information regarding the thermal expansion of the component. On the other hand, the variation of the intensity of the fringes is due to changes in reflectance of the observed surface. These changes are due to the variations in the refractive index as a function of temperature. The intensity of the fringes therefore indirectly provides the temperature of the component. Our method allows to de-correlate two quantities linked physically. This makes it possible to evaluate specific mechanical parameters of the component like its expansion coefficient.

This simultaneity is obtained using a CCD line sensor. Adjusting the position sensor and optimally setting the focusing lens system allow adaptation to different contexts.

Our system was compared to other systems for measuring temperature and another system for measuring dynamic displacement. This separated temperature/ expansion comparison was inevitable because unlike ours no system can simultaneously measure these parameters. These results show a good correspondence of our method with others, which makes it valid and promising.

VIII. ACKNOWLEDGMENTS

The authors wish to thank the cluster Moveo partners, Audacity project and ESP Carnot Institute for the contribution to collaboration GPM/CEVAA.

REFERENCES

[1] R. Aubry et al., “Temperature measurement by micro-raman scattering spectroscopy in the active zone of AlGaN/GaN high-electron-mobility transistors,” Eur. Phys. J. Appl. Phys., vol. 27, 2004, pp. 293-296.

[2] B. M. Cain, P. A. Goud, and C.G. Englefield, “Electrical measurement of the junction temperature of an rf power transistor,” IEEE Trans. on Instr. And Meas., vol. 41, 1992, pp. 663-665.

[3] S. Dilhaire, T. Phan, E. Schauf, and W. Claeys, “Laser probes and methodology for thermal analysis at micrometric scale. Application to microelectronics,” . Rev. Gen. Therm., vol. 37, 1998, pp. 49-59.

[4] C. C. Lee and J. Park, “Temperature measurement of visible light-emitting diodes using nematic liquid crystal thermography with laser illumination,” IEEE PhotonicsTechnology Let., vol. 16, 2004, pp. 1706-1708.

[5] D. Mc Namara, “Temperature measurement theory and practical techniques,” Analog Devices, AN-892, 2006.

[6] K. Nassim, L. Joannes, A. Cornet, S. Dilhaire, E. Schaub, and W. Claeys, “High- resolution interferometry and electronic speckle pattern interferometry applied to the thermomechanical study of a mos power transistor,” Microelec. J., vol. 30, 1999, pp. 1125-1128.

[7] A. Sarua et al., “Combined infrared and raman temperature measurements on device structures,”. Mantech Conference Vancouver Canada, 2006, pp. 179-182.

[8] G. Tessier, S. Holé, and D. Fournier, “Quantitative thermal imaging by synchronous thermoreflectance with optimized illumination wavelengths,” Appl. Phys. Let., vol. 78, 2001, pp. 2267-2269.

[9] A. S. Morris and R. Langari, Measurement and Instrumentation – Theory and Application, Elsevier, 2012.

[10] D. Astrain, J. G. Vian, and J. Albizua, “Computational model for refrigerators based on Peltier effect application,” Applied Thermal Engineering, vol. 25, 2005, pp. 3149-3162.

 ISBN: 978-1-61208-379-7

Robustness Study of SiC MOSFET Under Harsh Electrical and Thermal Constraints

To an in-depth physical failure analysis

Safa Mbarek, Pascal Dherbécourt, Olivier Latry, François Fouquet*
University of Rouen, GPM UMR CNRS 6634
*IRSEEM, Esigelec
Rouen, France
mbarek.safa@gmail.com

Dhouha Othman, Mounira Berkani, Stéphane Lefebvre
SATIE UMR CNRS 8029 ENS Cachan
Cachan, France

***Abstract*— The improvement of power conversion systems makes SiC devices very attractive for efficiency, compacity and robustness. However, their behavior in response to short circuit mode must be carefully studied to ensure the reliability of systems. This study deals with a SiC MOSFET. After the description of the component structure and its electrical performances, the paper presents some preliminary results for robustness evaluation in harsh electrical and thermal conditions. Extensive studies are underway to try to correlate the electrical measurements with an in-depth structural analysis.**

Keywords-SiC MOSFET; Robustness; Failure analysis; Short-Circuit.

I. INTRODUCTION

The SiC powers Metal Oxide Semiconductor Transistors (MOSFET) are very attractive for high temperature application in automotive or aeronautics. The isolated gate and the low drain to source resistance in on-state (R_{DSon}) make them ideal for using the MOSFET as a switch for all power electronics applications where efficiency, compacity and robustness are needed.

First, SiC MOSFET had been realized since 1986 [1] even if oxide on SiC were studied before. Many challenges in design have been made in order to expect higher frequencies switching than with Silicon Insulated Gate Bipolar Transistor (IGBT) and with higher operating temperature [7]. Investigations of different types of structures with 4H-SiC or 6H-SiC polytypes have been proposed with different challenges [8] and the vertical power devices double implanted MOSFET (DIMOS) shows very good performances due to its higher bulk mobility [9].

In order to evaluate the robustness of this type of MOSFET, several tests are possible. High temperature gate-bias (HTGB) and high temperature reverse-bias (HTRB) tests are common qualification tests for discrete semiconductor devices [10] however the short-circuit test [4][6][11] is one of the most used to evaluate the ruggedness of power devices during conduction. By using this test on a commercial 1.2 kV SiC Power MOSFET, the thermal instability and the electro-thermal behaviour are investigated in [11] by the mean of a state-of-art IR thermographic set-up. Othman et al. [4][6] present a comparison study on performances, reliability and robustness between SiC MOSFET and JFET devices. The aim of the studies is to evaluate the abilities and effects of each technology on the conception of power converter for avionic applications and to analyse their capability to withstand with harsh electrical constrains. The purpose of this paper is to present preliminary experimental results that could help to understand impact of the physical failure on the performance of the SiC MOSFET. First, the DIMOS is characterised in short circuit test. Then, two failure mechanisms are put in light with microstructural analysis.

II. COMPONENT UNDER TEST

The device under test (DUT) is a 1.2 kV SiC-MOSFET with continuous drain current I_{Dmax} equal to 24 A at 25 °C. The DUT is a vertical DIMOS transistor and its structure for an elementary cell is presented in Figure 1.

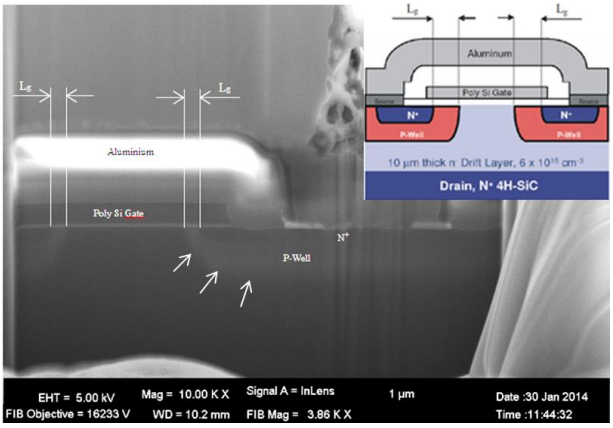


Figure 1. Structure investigation of DUT.

The Poly-Silicon Gate is isolated from the lateral channel with a thin oxide layer and the channel length is Lg. The Drain terminal is connected to the bottom of the device and is coupled to the channel through a drift layer that contributes to the on-state resistance and allows to apply large voltages between Drain and Source during off-state.

 ISBN: 978-1-61208-379-7

The other two terminals are fixed on the top of the device and the Source electrode, which is made of Aluminum, entirely covers the top face except on the Gate bonding wire pad.

Finally, the die is brazed on a Copper base plate that carries out the heat preventing from over self-heating during electrical conduction. The brazing, an alloy of Tin, Silver and Antinomy, is limited to a maximum operating temperature of 260°C.

As our main goal is to correlate the physical failure analysis of the DUT and the analysis of its cross section structure with the help of Scanning Electron Microscopy (SEM) after Focused Ion Beam (FIB) sample etching, the first step is to make sure that the depackaging process has no impact on the DUT electrical behavior. In order to verify this point, we have compared current-voltage (IV) characteristics before and after opening.

III. IVT CHARACTERIZATION

The static behavior of the transistor is studied over the recommended range of operating temperature according to the constructor datasheet (from -55 °C to +135 °C). Our study is restricted to the evolution of the drain current I_D depending on temperature variation, reflecting the behavior of the transistor.

To perform the current-voltage-temperature (IVT) measurements while providing thermal stabilization of the DUT, three experimental means have been implemented. For the negative temperatures, a cold forced air system is used; for positive and high temperatures, a Peltier module and a plate heating resistor are required. A thermocouple sensor, positioned as close to the base plate of the DUT as possible, controls its temperature. This setup provides a good accuracy and a satisfactory stability during measurements.

Pulsed mode is suitable to avoid any self-heating of the DUT during the static measurements. To satisfy this requirement, the pulse drain temporally includes the gate pulse. The pulse drain duration is fixed to 3 microseconds, with a duty cycle equal to 0.15%.

The output curves I_D (V_{DS}, T) in Figure 2 are plotted for a low gate voltage V_{GS} equal to 10V over a temperature range extending from -50 °C to 135 °C showing a steady increase in the drain current with the temperature increase.

The input curves I_D (V_{GS}, T) in Figure 3 are plotted for the same values of temperature, and for an output voltage V_{DS} equal to 18 V. The measurements suggest, as expected, a significant reduction in the threshold voltage V_{th} as a function of temperature.

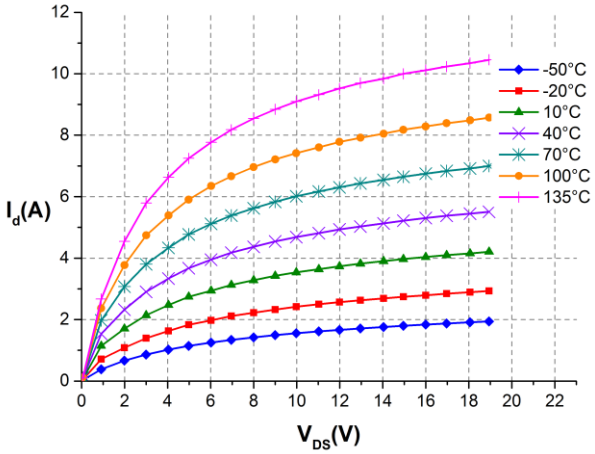


Figure 2. Output characteric I_D (V_{DS},T) for different temperatures with V_{GS} = 10V.

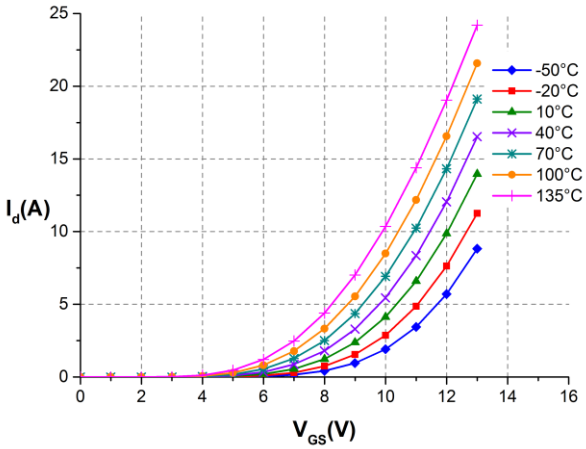


Figure 3. Input characteric I_D (V_{GS},T) for different temperatures with V_{DS} = 18V.

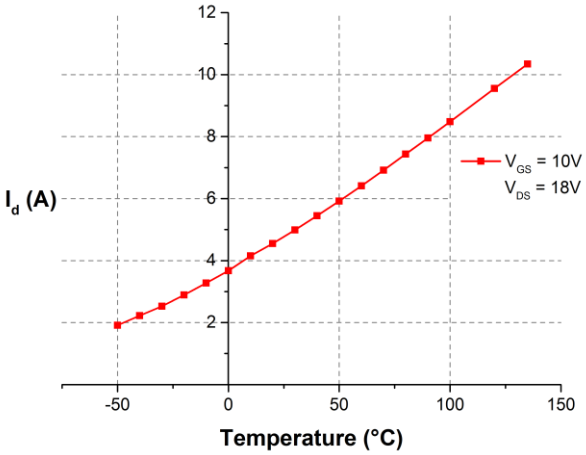


Figure 4. Drain current I_{Dsat} variation versus the temperature for V_{GS}=10V and V_{DS} =19V.

 ISBN: 978-1-61208-379-7

The variation of the drain current is given in Figure 4 in the area of near saturation of the output characteristic, for a gate voltage VGS equal to 10 V and for a drain voltage VDS equal to 19 V.

IV. ROBUSTNESS

Once the fresh electrical characterization of the devices achieved, a stress campaign is carried out under short circuit tests at room temperature, T = 25 °C. We apply a drain voltage of 600V without any output load. During the short circuit phase, the DUT is maintained in the on-state by applying a gate voltage of 20V and is then switched off by reverse biasing the gate [6]. Dedicated test bench schematic circuit is shown in Figure 5.

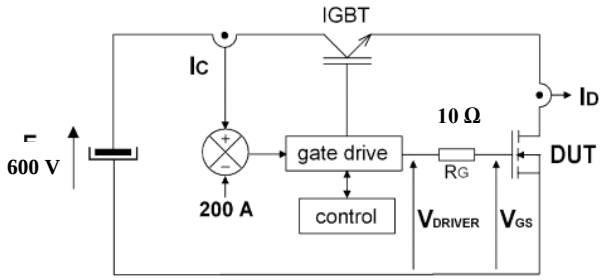


Figure 5. Test circuit for short circuit robustness analysis [6].

The first step consists in evaluating the energy responsible for the failure over long pulse duration. We applied the short circuit stress to a first component DUT1 for long pulse duration until complete failure. Figure 6 reports waveforms (V_{Driver}, V_{GS} and I_D) measured during the short circuit test. The corresponding dissipated energy before failure is about 0.96 J. Results indicate that the device was capable to sustain short circuit tests for a gate drive duration t_{SC} equal to 16 µs.

From these observations, we apply the same constraints to a second component DUT2 by applying successive short circuit tests with pulse duration t_{SC} increased by 2 µs at each test. In these conditions of experimentation, we try to determine the critical energy that is the maximum value of energy controllable by the DUT.

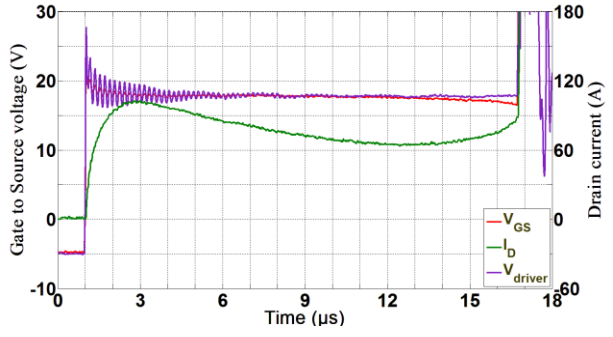


Figure 6. DUT1 : Robustness analysis under long pulse short circuit.

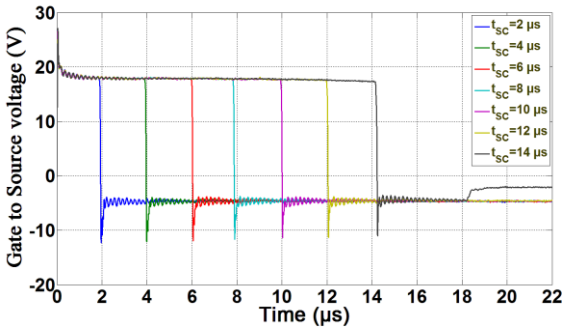


Figure 7. DUT2 : Evolution of gate to source voltage during short circuit tests showing a gate leakage current.

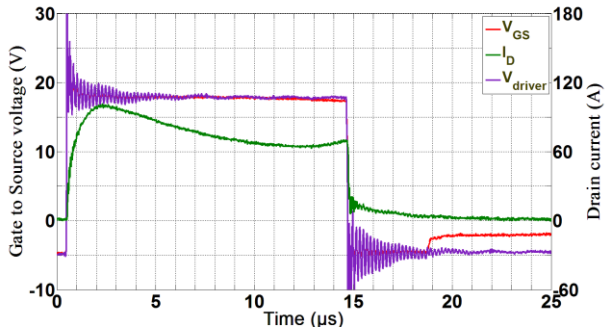


Figure 8. Failure of DUT2 during short circuit test, V_{DS} = 600V, t_{SC}=14µs, T=25 °C.

As illustrated, Figure 7 shows gate voltage waveforms measured during short circuit operations. The last short circuit stress of t_{SC} = 14µs applied to DUT2 causes the device failure. Figure 8 reports waveforms (V_{Driver}, V_{GS} and I_D) measured during the last short circuit test

V. DISCUSSION

A. Static I-V characteristics of the DUT

The drain current expression in linear operating region is given by (1), which describes the evolution of I_D (V_{GS}, V_{DS}) characteristic. For drain-source voltage (V_{DS}) values lower than V_{GS}-V_{th}, the drain current characteristic is linear, the behavior of the component is then like of a linear voltage-controlled inductance.

$$I_D = \frac{\mu_n W C_{ox}}{L} \cdot \left[(V_{GS} - V_{th}) V_{DS} - \frac{{V_{DS}}^2}{2} \right] \qquad (1)$$

Where W and L are respectively the gate width and the gate length, μ_n is the charge-carrier effective mobility and C_{ox} is the gate oxide capacitance per unit area.

The drain voltage corresponding to saturation mode is defined by the difference between the applied gate voltage and the threshold voltage that is, $V_{Dsat} = V_{GS}$-V_{th}. The V_{Dsat} value read from the output curve of Figure 2, for a temperature of 40 ° C, is equivalent to 10 - 2.9 (V), that to say 7.1V. Beyond this value, for increasing V_{DS}, the

 ISBN: 978-1-61208-379-7

component works in saturation mode and the current value maintains substantially the I_{Dsat} value. Then, the current value is independent from V_{DS} and its expression is given by (2). By experimental measurement, we found a decrease in Vth with temperature over a range from -50 °C to 135 °C. For a given value of V_{GS}, equation (2) predicts an increase of the saturation current I_{Dsat}, which is experimentally verified.

$$I_{Dsat} = \frac{\mu_n W C_{ox}}{L} . \left[\frac{(V_{GS} - V_{th})^2}{2} \right] \quad (2)$$

More rigorously, we should also take into account the evolution of the mobility μ_n which decreases with temperature [5]. V_{th} is also decreasing with temperature; the two variations (μ_n and V_{th}) have an opposite effect on I_{Dsat}. Threshold voltage variation is predominant at low gate voltage [2][3], which corresponds to our IVT characterizations as shown in Figures 2 and 3. The opposite behavior occurs for high gate voltage values, i.e., V_{GS} equal to 20V corresponding to robustness tests conditions presented in section IV.

B. Robustness

It is clearly seen during these robustness tests that the two devices had different failure mechanisms.

As shown in Figure 6, DUT1 fails after thermal runaway by the rapid increase in the drain current due to failure in short-circuit between drain and source. Similar failures were observed for SiC JFET devices for which increase of the temperature beyond the fusion limit of the device metallization was proposed in order to explain failures [4].

However, the transistor DUT2 first turns-off safely and short-circuits current falls down to zero as seen in Figure 8. But, few microseconds after the switch-off, a sudden short-circuit is observed between gate and source. So, gate failure occurred during transistor's off state after short-circuit but no failure occurred between drain and source electrodes. According to some authors, the gate oxide could be destroyed and the device became uncontrollable [6]. The time between the end of short-circuit and the gate failure may be due to temperature diffusion inside the device, which delay heating on the gate oxide

In Figures 6 and 8, we can clearly notice that a significant leakage current appears about 12 µs after the beginning of the short circuit, which seems to be responsible for the devices failures even if the failures seem to be different regarding the gate to source voltage evolution at failure. The decrease of the gate to source voltage during short circuit seems to show degradation between gate and source electrodes during this particular test [4].

C. Structural analysis of degradations

The stressed component DUT1 presenting GDS short-circuit degradation was unpacked and an example of a FIB cut (Focused Ion Beam) is given in Figure 9.

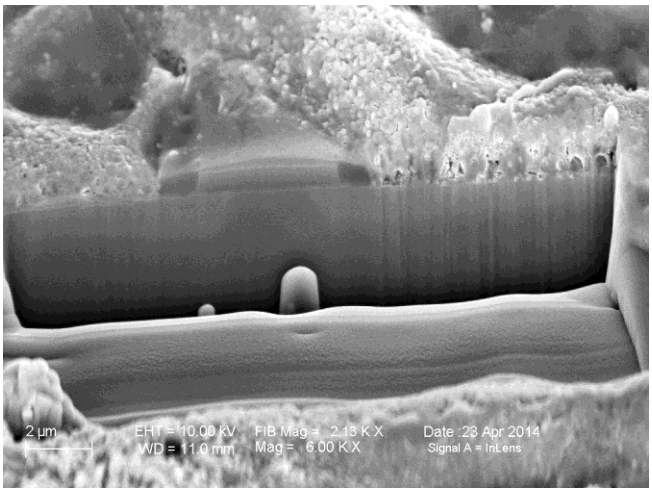


Figure 9. Material degradation analysis of DUT1.

Analyzing the figure, we could make the following observations:

- We are able to identify the different elements of the structure of the MOSFET,
- The aluminum layer corresponding to the source contact largely melted.

However the reduction of the oxide over the gate polysilicon may be due to the unpacking process; further tests will be performed.

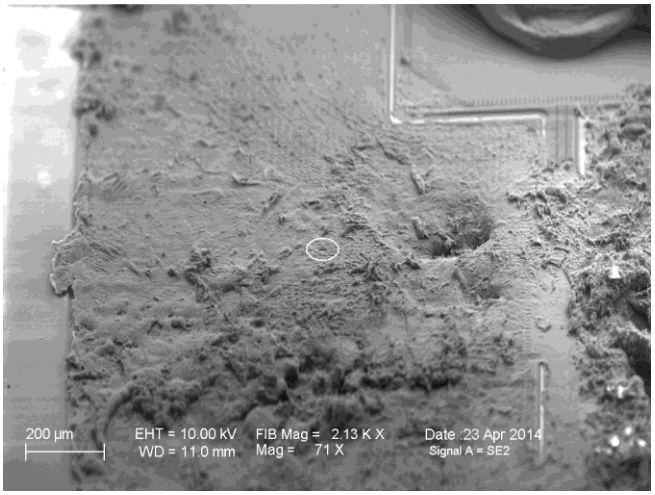


Figure 10. FIB cut location and surface degradation of DUT1.

Regarding the gate oxide, between channel and polysilicon gate, it does not appear to have been damaged by the stress but the FIB cut was performed in a randomly selected location and not where the surface degradations appeared to be the strongest Figure 10.

VI. CONCLUSION

A complete thermal IVT characterization has been performed and shows a behavior consistent with the manufacturer's data. Short circuit tests, conducted under extreme conditions, showed a destruction of the two devices

 ISBN: 978-1-61208-379-7

under test. However, different failure modes were observed. To better understand the failure mechanisms, current studies are carried out under softer short circuit test in order to highlight significant variations in electrical measurements. Defects generated in the device structure will be located and analyzed by SEM and FIB in the aim to correlate these observations with the electrical measurements.

REFERENCES

[1] Y. Kondo et al., "Experimental 3C-SiC MOSFET," Electron Device Letters, IEEE, vol. 7, no. 7, July 1986, pp. 404–406.

[2] P. H. Aaen, J. A. Pla, and J. Wood, "Modeling and characterization of RF and microwave power FETs," Cambridge University Press, Ed2007.

[3] Md. Hasanuzzaman, S. K. Islam, L. M. Tolbert, and M. T. Alam, "Temperature dependency of MOSFET device characteristics in 4H and 6H silicon carbide (SiC)," Solid-Sate Electronics, vol. 48, 2004, pp. 1877-1881.

[4] D. Othman et al., "Comparison study on performances and robustness between SiC MOSFET & JFET devices – Abilities for aeronautics application," Microelectronics Reliability, vol. 52, 2012, pp 1859-1864.

[5] B. J. Biliga, "Silicon Carbide Power Devices," North Carolina state university, USA, World Scientific Publishing, Ed 2005.

[6] D. Othman et al., "Robustness of 1.2 kV SiC MOSFET devices," Microelectronics Reliability, vol. 53, 2013, pp. 1735-1738.

[7] K. Matocha, "Challenges in SiC power MOSFET design," Solid-State Electronics, vol. 52, no. 10, October 2008, pp. 1631–1635.

[8] S. Dimitrijev and P. Jamet, "Advances in SiC power MOSFET technology," Microelectronics Reliability, vol. 43, no. 2, February 2003, pp. 225–233.

[9] V. R. Vathulya and M. H. White, "Characterization and performance comparison of the power DIMOS structure fabricated with a reduced thermal budget in 4H and 6H-SiC," Solid-State Electronics, vol. 44, no. 2, February 2000, pp. 309–315.

[10] L. Yang and A. Castellazzi, "High temperature gate-bias and reverse-bias tests on SiC MOSFETs," Microelectronics Reliability, vol. 53, no. 9–11, September–November 2013, pp. 1771-1773.

[11] M. Riccio, A. Castellazzi, G. De Falco and A. Irace, "Experimental analysis of electro-thermal instability in SiC Power MOSFETs," Microelectronics Reliability, vol. 53, no. 9–11, September–November 2013, pp. 1739-1744.

 ISBN: 978-1-61208-379-7

A Log-Tool Suite for Embedded Systems

Harald Schuster, Martin Horauer, Michael Kramer
University of Applied Sciences Technikum Wien
Vienna, Austria
e-mail: {schuster, horauer, kramer}@technikum-wien.at

Heinz Liebhart, Josef Büger
Kapsch TrafficCom AG
Vienna, Austria
e-mail: {heinz.liebhart, josef.bueger}@kapsch.net

***Abstract*—Logging is a common method to monitor the operation of a system and to identify failures of services and system components. When developing software for an embedded system there are stricter restrictions that result from the limited hardware resources. The performance of the hardware as well as the available memory is limited. Furthermore, embedded systems are often not accessible for a long period of time. This paper presents tools to improve existing logging methods without compromising existing development flows. Essentially, these tools on one hand transparently pre-process the source code in order to optimize logging instructions and add on the other hand a post-processing step when analyzing log-files. The presented tools were evaluated by integrating them to an industrial project where they show a significant improvement of the logging performance.**

Keywords*—*Logging; Embedded Systems; Memory Limitation; Log Analysis.

I. INTRODUCTION

Electronics paired with software is a key enabler for many modern products and systems. For example, it allows to add new features and functionalities, improve reliability, safety, environmental efficiency, or comfort. At the same time the complexity of these systems is steadily increasing mandating more efforts and new approaches for verification. In practice, various processes and approaches are in use — very common are static analyses, code reviews paired with testing and sometimes even formal methods. For embedded systems this problem is aggravated since usually different environmental conditions, limited controlability, and observability need to be taken into consideration. Thus, rigorous verification is a challenging task that is often limited by economical aspects.

In the field of embedded systems, Embedded Linux gained momentum in recent years due to plummeting costs of available hardware resources, the maturity and the feature rich functionality of the open-source code-base. Here, as with desktop operating systems, logging is a common approach to identify problems of an operational system. Services and applications can write information, warnings and encountered errors to log-files in a chronological fashion. Analyzing these logs can help to identify problems and hint to their sources. Especially when dealing with long-running server applications logging is a common practice. For example, [1] observed that on average common server programs provide one logging instruction every 30 code lines and that 18% of all committed revisions are due to modifications of the respective logging messages. In contrast, to desktop and server systems, memory is still a limiting and costly factor for embedded systems. Hence, for such systems a trade-off between the time a log-file is accumulated, the number of log messages, their frequency and verbosity is important. On one hand, lowering these values is necessary when the amount of available memory is limited – this will typically reduce the expressiveness and usefulness of the logs. On the other hand, frequent and verbose log-messages will also impair on the available processing power — the latter being another limited factor.

The main contribution of this paper is to present some tools and a flow that improves the efficiency of existing logging practices. In particular, it optimizes the length, verbosity and performance of the logs without impairing the existing log facilities and requires only a marginal adaption of the development flow. We evaluate our approach using an industrial strength Embedded Linux device that is executing a communication stack used for road-pricing systems.

The remainder of this paper is structured as follows: Section II provides an overview of existing tools and approaches to improve the performance of logging on embedded systems. Then we present our modified tool flow before we provide some implementation details of our LogEnhancer and the LogAnalyzer tools. Afterwards, we describe how we evaluated our tools before we conclude the paper giving an outlook on future directions.

II. RELATED WORK

Logging is a common approach to identify problems in a computing solution [2][3][4]. When looking at Linux, for example, multiple different logging methods are available. In the majority of these cases, however, these methods are designed for standard desktop or server systems. Widespread in use are `syslog` [5] and their forks `rsyslog` and `syslog-ng`. Another common logging method is `Journal` which is included in the system and service manager `Systemd`. All these systems, however, are not optimized for embedded targets.

For these systems dedicated logging methods exist that improve several aspects when compared to the "standard" methods, cf. [6]. For example, Amontamavut et al. [7] describes a logging mechanism that is optimized for embedded systems with small memory. Their approach is to split the logging mechanism into two parts. One part is executed on the embedded device and sends reduced log messages via Ethernet to a server. The second part is the server and consists of different monitoring and debugging tools. One of these tools is a log analyzer to generate readable log messages. Overall, this approach reduces the log-file size and improves the log performance, however, it requires an active network connection to a remote host when the system is operational; a requirement that rules this approach out for our intended field of application.

Jeong et al. [8] describe a logging mechanism to improve the message throughput and to speed up the latency. They achieved their goals by avoiding the message transfer between

 ISBN: 978-1-61208-379-7

the user and the kernel space. The message is stored in a shared memory and forwarded to flash memory or a remote host by a message collector. The developer can use the normal logging API. This changes increases the message throughput but does not affect the log-file size.

III. Approach

The motivation of our work is to improve the logging functionality of an Embedded Linux device. In particular our approach intends to optimize memory usage and performance without impairing on the existing development process, in particular, how log messages are coded. To that end we first describe existing approaches and development flows used by our industrial partner before we detail our new concept and implementation.

A. Status Quo

We implemented and evaluated our approach on code provided by our industrial partner used for embedded devices in the field of road pricing around the globe. In fact, there is already a very large base of devices running in the field; for instance, most commercial road vehicles in the European Union are equipped with these devices. Hence, similar to automotive electronics even small savings have a significant impact on the costs thus simply expanding available resources (computing power, memory size) is critical. Here, next to elaborate testing, logging and off-line analysis of the log-files is a common approach to tackle runtime mis-behavior.

The target employs OpenWRT a Linux distribution for embedded devices as operating system. The application mostly uses C code along with CMake as build system. The employed logging method employs macros to emit log-messages and define statements to set the logging level. The logging levels are ERROR, WARNING, NORMAL, INFORMATION, DEBUG and TRACE. Furthermore, using logging masks it is possible to filter the kind of messages that are logged (e.g., communication related log-messages, layer 2 messages, layer 3 messages).

The source-files consist of roughly 150k lines (counted using the tool cloc) of source-code targeting 3 different targets; Table I lists the log messages found therein separated by their logging level. In total, the source-files hold 2259 log messages; thats on average about one log message every 67 lines of code. Compared to several open-source projects listed in [1] this code contains roughly 2-3 times less log messages. One reason therefore, might be that the code at hand targets an embedded platform, whereas the listed references are aimed at server applications.

TABLE I. LOG MESSAGE STATISTICS

LOG_ERR	LOG_WARN	LOG_INFO	LOG_DBG	LOG	LOG_HEX
517	262	340	53	1085	2

Logging is implemented as follows: The developer uses a logging macro to insert log messages into the source code. Furthermore, he has to define the logging level (e.g., LOG_ERROR) and the logging mask (e.g., LM_COM). The logging level is set by a macro – there exists one for every level. The logging mask is the first parameter of the macro followed by the message and the arguments that should be logged. The text string of the message is often put together using several ANSI/ISO C functions and finally output to *stdout*. In order to get a coherent logging style, the use of these macros is enforced by a strict development process using various approaches like code-reviews or static code analysis. Hence, our approach must comply with the same rules.

```
/* logging macro */
ret = 12;
LOG_ERR(LM_ALL, "layer2_init returned %d %s\n", ret,
     time_buffer);

/* logging macro resolved */
log_(LL_ERR, LM_ALL, "layer2_init returned %d %s\n",
     ret, time_buffer);

/* log message */
<E>LM_ALL: layer2_init returned 12 2014-02-16_13
    .52.14.012
```

Figure 1. The ‘present’ logging approach.

Figure 1 depicts an example of a typical logging macro and how it is resolved. The last line of this listing illustrates a log message entry produced by this code.

B. Requirements

The requirements for our approach are based on the above analysis of the source code, the physical limitations of the embedded target as well the existing development flow. The detected requirements are as follows:

(i) File Size: A significant reduction of the file-size of the log-files shall have top most priority. This will make room for more and more verbose logging instructions and longer logging intervals before the logs get rotated.

(ii) Performance: Improve the performance or at least keep the penalty on the performance as low as possible. Note that this requirement rules out computing intense (run-time) compression approaches.

(iii) Tool Flow: Keep modifications to the existing tool-flow and especially the way logs are generated as low as possible. The developers should not be affected by our approach requiring that the tools must be integrated in the standard build process in use.

(iv) Log Information: Enhance the log messages in a way so that the time and origin of the message is better traceable to locations in the source-code of the respective revision.

(v) Filtering: Allow for a better, more efficient filtering of relevant messages that led to erroneous log-entries. The aim here is to help in the investigation of causes that were leading to this message.

(vi) Revision Management: Typically many embedded devices are operational in the field using different revisions of the respective application. Thus proper bug tracking and revision management are essential in order to correlate the log-files with the correct revision of the source-code.

C. Design Concept

Our design approach first runs as a pre-step to the build process and is as follows:

 ISBN: 978-1-61208-379-7

```
/* replaced logging macro */

log_ts_(LL_ERR, LM_ALL, "1 %d\n", ret);

/* log message */
1 12 <ts>
```

Figure 2. The 'new' logging approach.

```
<E>LM_ALL: layer2_init returned 12 [l2_state_machine
     @ layer2.c(368)]
```

Figure 3. Reconstructed log message.

(1) Enumerate found logging-commands in the source-code and store the messages in a separate decoder-file.
(2) Replace the logging commands with the enumerator from step (1) along with the values of variables found in the logging command separated by a white-space.
(3) Build the program and deploy the binary to the target system and set the latter into operation.

When the log-files are analyzed the following steps are performed:

(4) Retrieve the log-files from the host; therefore the target must be connected to the respective workstation.
(5) Reconstruct the original log-messages using the logged enumerator and variables along with the decoder-file.

By logging only a simple number instead of an entire text string the logging becomes more efficient and the log-files get smaller. Furthermore, the code on the embedded target becomes slightly more efficient as well, since we offload the bulk of the work (putting the log-commands together) to a host computer.

Figure 2 illustrates our new logging approaches produced by our LogEnhancer tool; the first code line shows the log command as replaced by our tools. The last line again shows the log message written to the log-file; note that for readability we replaced the binary time-stamp with <ts> in this example.

Filename, and source-line of the log-command that produced a log-entry is stored along with the enumerator in the decoder-file. Figure 3 shows the log-message that was reconstructed by our LogAnalyzer tool. Note that this log message additionally contains the name of the function, the file-name and the source-code line where the log-entry originated.

IV. Tool Implementation

As described in Section III, the tool is split into two parts the LogEnhancer used prior to the build process and the LogAnalyzer used when log-messages get inspected and analyzed. Both tools were implemented using Python.

A. LogEnhancer

The detailed work flow of the LogEnhancer is illustrated in Figure 4. This process is transparent and hidden from the developer and works as follows:

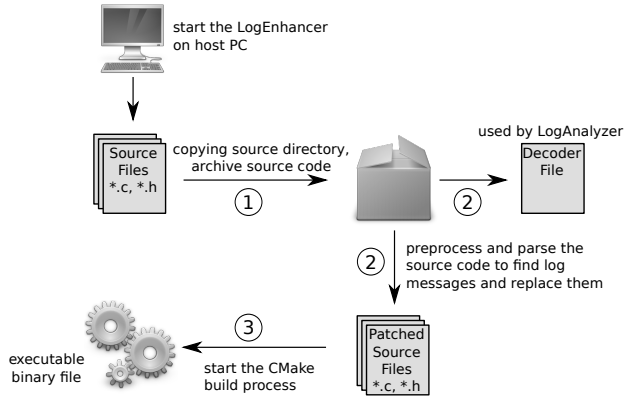


Figure 4. Workflow of the LogEnhancer.

1. The tool creates a copy of the source code in a temporary sub-directory and additionally creates an archive file for backup purposes. For proper identification purposes both the directory and file-name use date, time-stamp, and version information in their name.
2. Next, the code is pre-processed using the compiler infrastructure. Based on the pre-processed files, the C-code is parsed using the Python module PYCParser that supports almost the entire C99 language standard. The parser generates a structured tree representation of the source-code that allows convenient identification of all logging commands. The module provides a function to locate specific function calls. Furthermore, the function calls are represented as a node in the tree and also includes information about the exact position of the logging-command within the source-code. This location information and the logging message is written to a decoder-file. The file-name of the decoder-file follows the same naming convention as in step 1 in order to provide a convenient mapping to the respective source-files. Furthermore, the file-name is added as an ERROR log-command to the source-code that gets executed once when the system is put into operation. A small additional tweak of the script that is responsible for the rotation of the log-files is necessary in order to retain this information. Next, the logging-command is modified in a way so that a single enumerator is output along with the values of the variables (if there are any logged at the particular location), cf. Figure 2.
3. The modified source-files are built by invoking the standard build process — our industrial partner uses CMake therefore. As a result, we obtain the executable binary that can be deployed to the target system.

All these steps are transparent to the developer.

B. LogAnalyzer

The LogAnalyzer is run whenever a recorded log-file is analyzed in order to obtain readable log-files, cf. Figure 3. The involved steps are:

a. Scan the log-file for the file-name in order to map the correct decoder-file and source-tree to the log-file, cf.

 ISBN: 978-1-61208-379-7

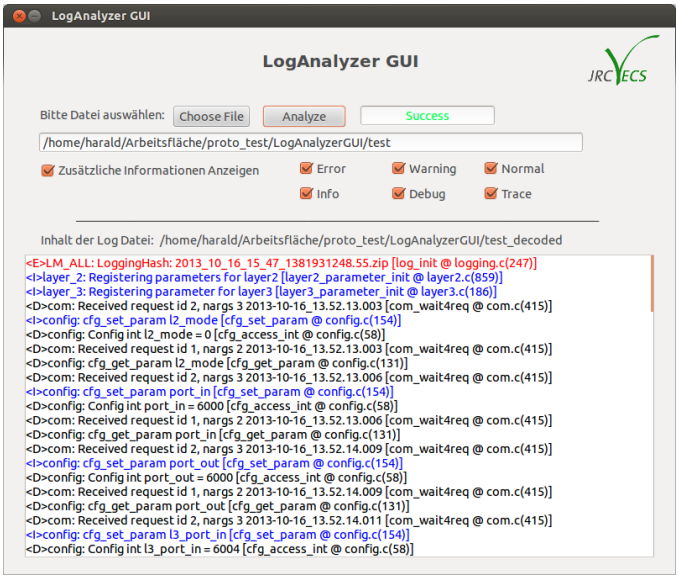


Figure 5. Graphical User Interface of the LogAnalyzer.

Section IV-A. In case no suitable files are available, a respective notification is output to the developer.

b. Next, the LogAnalyzer parses the log-file line by line and decodes the messages with the help of the decoder-file.

Using various options this tool allows to apply convenient filters to the log-file analysis. The LogAnalyzer tool itself is available both as a command-line tool and a program with a graphical user-interface (GUI), see Figure 5 for a screen-shot. The GUI variant, additionally supports a colored highlighting mode for a more convenient overview. The reconstructed output can be saved as a simple text-file.

V. Evaluation

In order to evaluate our approach, we applied the tool suite on a industrial embedded systems project in the field of road pricing systems. One part of the project is a communication stack that handles the communication between an on-board vehicle unit and an observer unit located at defined way-points along traffic routes.

A. Setup

In order to evaluate our logging approach, we chose the setup in a way so that we are able to control the frequency the communication is triggered in order to test the system in a reasonable time frame. To that end, the communication stack runs in a loop and produces frequent log-entries.

The primary focus of our evaluation was on the log-file size, however, we additionally monitored also the impact on the required computational resources. To that end we automated the tests using scripts and executed them multiple times for different time spans in order to avoid impacts due to the initialization process of the stack. The following presentation illustrates the results using test-runs spanning 10 and 30 minutes, respectively. As expected (and also evaluated by additional experiments) longer test-runs yield the same results.

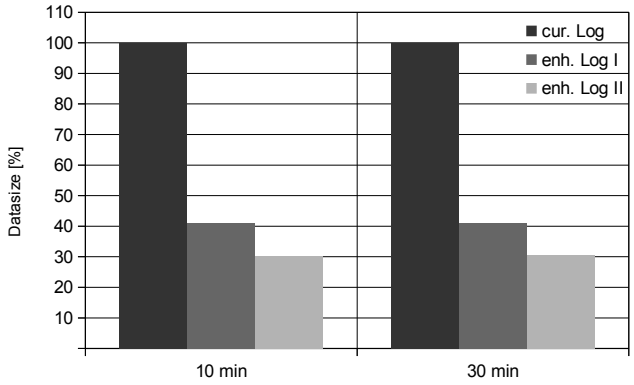


Figure 6. Log File Size with different Log Methods.

In addition we conducted the test-runs using different development stages of our implementation:

- **cur. Log:** These tests represent the reference runs where we executed the unaltered source-code using the standard logging method.
- **enh. Log I:** The first development stage of the tool suite replaces the entire text string of the log messages by a unique ID, thus only this ID along with a time-stamp is written to the log-file.
- **enh. Log II:** The second development stage includes an alteration of the time-stamp. Here the time-stamp is written in binary format without loss of information, whereas it is written in a readable format in both stages before.

B. Results

Figure 6 presents the results of two test-runs spanning 10 and 30 minutes, respectively, using the most verbose logging level. The run with the current log method (cur. Log) is the reference run and represents 100% log-file size. Using our tool suite at the first development stage (enh. Log I) we are able to reduce the log-file size by 59%. Performance wise we could monitor a very small almost negligible reduction of the required processing power. The second development stage (enh. Log II) reduces the file-size even further down to an average of 69%. In particular, the time-stamp is reduced from 23 bytes in readable format to 7 bytes in binary format. This reduction of the needed bytes has no significant effect on the somputing performance. A field test shows that the current logging method generates a log-file with a size of 48 MB within one day. By using the LogEnhancer in the second development stage the size of the log-file is reduced to 19 MB without any loss of information.

Summarizing, the results show a high potential to save memory space by replacing the text string with a unique number and by modifying the time-stamp. The solution has only a marginal impact regarding the improvement of the performance.

VI. Conclusion and Future Work

This paper presents an approach and some tools to improve the efficiency of logging in embedded systems with minimal

alterations on an existing design flow, i.e., how a designer adds logging commands to the source code and how the log-files get analyzed. The solution reduces the log-file sizes, slightly optimizes required computing resources, and even adds more verbosity to the log-messages like file, function and line-number where a logging message emanated without further ado. As a benefit logging times can be increased before they get rotated and more verbosity can be added, thus improving the chances to catch rare or unlikely problems encountered in the field.

The solution was designed for an industrial road pricing solution in mind, is however, applicable 'as is' for any other kind of Embedded Linux application. It can be improved even further in various different ways. For example, in practice many log-messages recur multiple times, thus instead of logging the message one could simply log the message once and how often they recur. Furthermore, a tool that guides the designer where to insert logging commands and what messages shall be logged could help to further improve the chances to catch problems. Additionally, tighter integration with test- and version-management systems would be advantageous.

ACKNOWLEDGMENT

This work has been supported by Kapsch TrafficCom AG and by the public funded R&D project Josef Ressel Center for Verification of Embedded Computing Systems (VECS) managed by the Christian Doppler Gesellschaft (CDG).

REFERENCES

[1] D. Yuan, S. Park, and Y. Zhou, "Characterizing Logging Practices in Open-Source Software", 1em 34th International Conference on Software Engineering (ICSE), June 2012, pp. 102–112.

[2] J. H. Andrews, "Testing using log file analysis: tools, methods, and issues, Proceedings of the 13^{th} IEEE International Conference on Automated Software Engineering, Oct. 1998, pp. 157–166.

[3] J. H. Andrews and Y. Zhang, "General test result checking with log file analysis", IEEE Transactions on Software Engineering, vol. 29, no. 7, July 2003, pp. 634–648.

[4] D. Yuan, J. Zheng, S. Park, Y. Zhou, and S. Savage, "Improving software diagnosability via log enhancement", ACM Transactions on Computer Systems (TOCS), April 2012, pp. 3–14.

[5] R. Gerhards, The Syslog Protocol, http://tools.ietf.org/html/rfc5424, 2009, [retrieved: Sept., 2014].

[6] Y. Shi, L. Renfa, L. Rui, and X. Yong, "Log analysis for embedded real-time operating system based on state machine", International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), Aug. 2011, pp. 1306–1309.

[7] P. Amontamavut, Y. Nakagawa, and E. Hayakawa, "Separated Linux Process Logging Mechanism for Embedded Systems", 18^{th} IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), Aug. 2012, pp. 411–414.

[8] J. Jeong, "High Performance Logging System for Embedded UNIX and GNU/Linux Application", 19^{th} IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), Aug. 2013, pp. 310–319.

 ISBN: 978-1-61208-379-7

Scan-shift Power Reduction Based on Scan Partitioning and $\bar{Q}$-D Connection

Tiebin Wu, Li Zhou and Hengzhu Liu
College of Computer, National University of Defense Technology
Changsha, China
e-mails: {tiebinwu@126.com, zhouli06@nudt.edu.cn, hengzhuliu@nudt.edu.cn}

***Abstract*—Excessive test power consumption is a great concern in modern VLSI testing. This paper presents an efficient scan-shift power reduction scheme based on scan chain partitioning and $\bar{Q}$-D connection. After partitioning the scan chains into several segments equally, selective $\bar{Q}$-D connection is introduced to reconfigure each segment, which only exploits the $\bar{Q}$ output port of the scan flip-flop and no additional hardware or routing overhead will be introduced. Experimental results show that the proposal can achieve 3.43% scan-shift power reduction on average with the help of selective $\bar{Q}$-D reconnection after scan partitioning. Furthermore, the proposed scan-shift power reduction technique can be acceptable for Built-In Self-Test (BIST) and non-BIST scan-based testing architecture without affecting test application time, test fault coverage, performance and routing overhead of the circuit under test.**

Keywords*—scan partition; $\bar{Q}$-D connection; low power test; scan-shift power.

I. INTRODUCTION

With the complexity and performance of very large scale integrated (VLSI) circuits growing, the power consumption density of advanced VLSI is rapidly increasing since the supply voltage cannot be reduced so much due to its noise margin [1]. Furthermore, power consumption in at-speed scan testing is significantly higher than that during normal functional operation, while modern VLSI testing aims to perform at-speed or even faster than at-speed testing to test the chip for high-quality screening [2][3]. Therefore, the increasing power consumption during testing has been a serious concern, which can result in voltage drop, yield loss, reliability problems and even heat damage of the Circuit Under Test (CUT) [4]–[6].

To address these problems, many testing power reduction methodologies have been proposed, such as X-filling approaches which explore the reassignment of don't care bits in test cubes to reduce switch activities [7]–[9], low power Test Pattern Generation (TPG) techniques that modify the architecture of TPG like Linear Feedback Shift Register (LFSR) to reduce the transitions of inputs of the CUT [10]–[12], test vector reordering methods through changing the order of the test vectors to reduce the number of transitions between two consecutive vectors [13][14], and power-aware Automatic Test Pattern Generation (ATPG) techniques [15][16], and so on.

For scan-based testing, scan-shift power is defined as the dynamic power consumption dissipated by serial shift operations during scan-in of test stimuli and scan-out of responses, while capture power is defined as the power consumption occurred in responses capturing mode [12][17]. Scan-shift power can be effectively reduced by scan cells reordering, which is one of the most attractive techniques to reduce scan-in and scan-out transitions by rearranging every scan cell with proper position [18][19]. However, this approach usually costs excessive high routing hardware and computation time according to the increasing number of scan cells. Selective reconfigurable inverters are inserted between scan cells to decrease the switching activities in the scan chains during shift operation in [20], which is claimed to be suitable for any scan architecture. Unfortunately, this technique not only requires large area and control cost, but also influences the performance of the CUT. Scan architecture modification techniques through modifying the scan architecture to reduce scan test power by inserting gates or partitioning scan chains into several segments [21]–[26]. Low power jump scan architecture is utilized by Chiu and Li [21] to reduce test power with penalty of speed performance degradation, while low power Illinois scan architecture is proposed by Chandra et al. [22] to reduce scan-in shift power but not suitable for decreasing the test power dissipated during scan-out shift operation.

Scan partitioning/segmentation is another scan architecture modification technique, which divides a given scan chain into several segments to cut down the shift process for full scan chain into a sequence of segment-wise shifts [23][24]. Scan partitioning scheme with scan freeze flip-flops and status registers is introduced by Kim et al. [23] to reduce the scan-shift power consumption. The test application time of the previous scheme will be raised since it requires additional test clock cycles to scan-in the configuration data stored in status registers. In [24], the given scan chain is partitioned into s segments, and only one segment is active during the scan shift operations, while all other segments are clock gated in hold mode to retain the scan test data. This approach is very efficient to reduce scan-shift power and can be applicable to Built-In Self-Test (BIST) and non-BIST schemes without affecting test application time, fault coverage, performance of the CUT and/or the scan cell routing cost.

In this paper, we propose a scan-shift power reduction scheme based on scan partition and $\bar{Q}$-D connection. After evenly partitioned each scan chain into several segments as in [24], $\bar{Q}$-D connection is introduced to reconnect two consecutive scan cell for every segments, which will reduce the scan-shift power further. The proposed scheme requires no additional hardware with respect to [24] and keeps all advantages of [24], which is suitable for BIST and non-BIST test environments without penalty of test quality, performance degradation or additional hardware of the CUT.

The rest of this paper is organized as follows. Section II describes the scan-shift power metric and related works. Section III presents the proposed scan-shift power reduction scheme. The experimental results and comparison are shown in Section IV. Finally, Section V concludes this paper.

II. RELATED WORK

A. Scan-shift Power Metric

As mentioned above, scan-shift power is the dynamic power consumption dissipated during shift operations, which generally depends on the switching activity or transitions

 ISBN: 978-1-61208-379-7

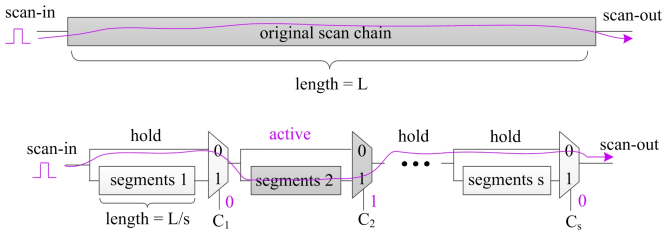


Figure 1. Shift operations of original scan chain and partitioned scan segments

occurred in the scan chain. Therefore, we utilize the widely used Weighted Transition Metric (WTM) [27][28] for the scan-shift power evaluation in this work. *Scan-in power* is the dynamic power consumption dissipated during scan-in of test stimuli, which can be calculated as (1), according to WTM [27][28].

$$WTM_{in} = \sum_{i=1}^{N} \left[\sum_{j=1}^{L-1} (t_{i,j} \oplus t_{i,j+1}) \times j \right] \quad (1)$$

where N is the number of test vectors, L is the length of scan chain and $t_{i,j}$ is the j^{th} bit of test vector t_i.

According to WTM [27][28], *scan-out power* can be calculated as (2), which is the dynamic power consumption dissipated during scan-in/out of test responses.

$$WTM_{out} = \sum_{i=1}^{N} \left[\sum_{j=1}^{L-1} (r_{i,j} \oplus r_{i,j+1}) \times (L-j) \right] \quad (2)$$

where N is the number of test vectors, L is the length of scan chain and $r_{i,j}$ is the j^{th} bit of test response r_i.

Therefore, the total number of weighted transitions of *scan-shift power* can be calculated as (3) according to WTM [27][28], since it is the sum of *scan-in power*, *scan-out power* and the total number of transitions between the MSB of the previous test response and the LSB of current test vector, which will propagate from the first scan cell to the last in the scan chains during scan-out operations.

$$WTM = WTM_{in} + WTM_{out} + \sum_{i=1}^{N-1} (t_{i+1,L} \oplus r_{i,1}) \times L \quad (3)$$

where $t_{i+1,L}$ is the LSB of test vector t_{i+1}, $r_{i,1}$ is the MSB of test response r_i, and r_i is the corresponding response of test vector t_i.

B. Scan Partition and Scan Hold

In [24], a low power scheme based on scan partition and hold is proposed, which is shown in Figure 1. After equally partitioning the scan chain into s segments, a multiplexer is utilized to connect two consecutive segments and each multiplexer is controlled by signal C_j ($1 \leq j \leq s$). During scan shift operations, only single signal of C_j is set to high, and the corresponding segment j is working in scan-in/out operations, while others are hold in bypass state. Note that, working in hold mode means that the scan cells of this segment will retain their scan test data unless the corresponding control signal C_k jumps to high. Therefore, the scan test data of every segment only require to scan-in/out through itself after scan chain partition and hold, while it need to shift through the full scan chain until it arrive at the appropriate position before scan partitioning. For example, if a transition exists between the last two scan cells of the scan chain ($Length = L$) for a test vector, this transition will propagate from the first scan cell to the $L-1$ scan cell before utilizing the low power scheme in [24], where it will cause $L-1$ scan cell transitions during scan-in operations. After partitioning the scan chain as shown in Figure 1, this transition will only propagate through the last segment, where only $L/s-1$ scan cell transitions will occur. Hence, the scheme in [24] can reduce the scan-shift power significantly without affecting test application time, fault coverage, performance and/or the scan cell routing cost of the CUT. Furthermore, this low power scheme can be easily extended to multiple scan chains test architecture.

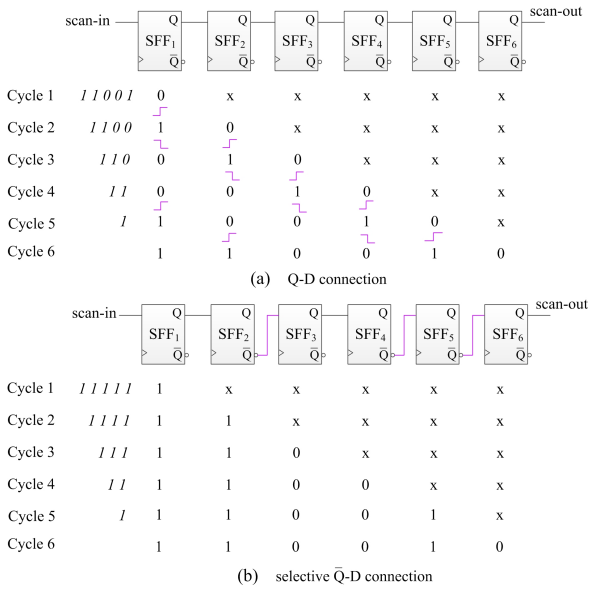


Figure 2. Example of selective Q̄-D connection [25]

C. Q̄-D Connection

Without introducing additional hardware and routing cost, selectively replacing the Q-D connection with Q̄-D connection between two consecutive scan cells in the scan chain can also reduce the switching activity during scan shift operations effectively [25][26]. For example, assume the scan chain include 6 scan cells and test vector t=110010 be applied to the chain. If only the Q-D connection is utilized to chain all scan cells, 11 transitions will occur in order to apply the test vector t, as shown in Figure 2(a). However, if we selectively configure the sub-chain of scan cells SFF2-SFF3, SFF4-SFF5 and SFF5-SFF6 with Q̄-D connection, the reconfigured pattern t'=111111 will be shifted into the scan chain instead to actually apply the original test vector t to the CUT, where no transition will occur during scan-in shift operations, as shown in Figure 2(b). It is obvious that, the scan-shift power can be reduced similarly without requiring any additional logic or routing cost, and it can be simply extended to multiple scan chains with multiple test patterns.

III. Proposed Scheme

In order to further reduce the scan-shift power, we apply the Q̄-D connection technique as in [25][26] into the segments

 ISBN: 978-1-61208-379-7

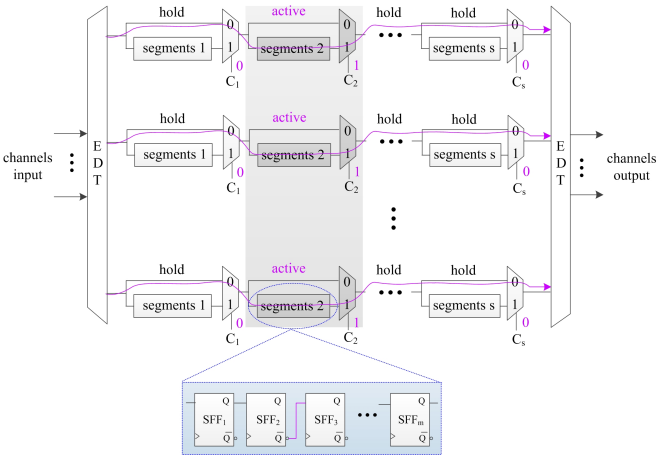


Figure 3. Proposed low power scan architecture

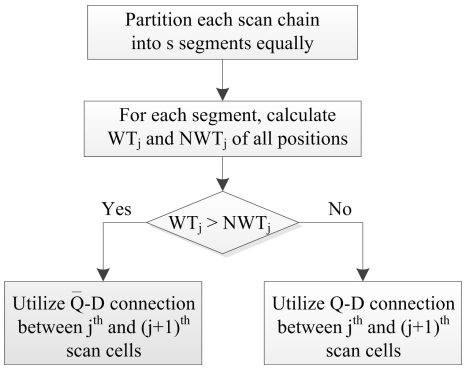


Figure 4. Selective Q̄-D reconnection flow

after partitioning the scan chains as in [24]. The proposed low power scan test architecture is shown in Figure 3. Firstly, every original scan chain is equally partitioned into s segments, where each segment has L/s scan cells and L is the length of scan chain. The scan-in and scan-out ports of each segment are connected to a multiplexer, and a signal C_i is utilized to control it. If and only if single C_i is set to high, each i^{th} segment for every scan chain is active for scan-in stimuli and scan-out responses, and others are hold to retain their scan test data and bypass the scan data of the corresponding active segment during scan shift operations as in [24]. Unlike in [24], where all scan cells are chained in Q-D connection, we selectively introduce Q̄-D connection as shown in the broken blue block in Figure 3 to reconfigure each segment after scan partitioning to further reduce the scan-shift power. It is obvious that, no additional hardware or routing overhead will be introduced to apply the selective Q̄-D connection technique after scan partitioning as shown in Figure 3.

Then, we will describe the procedure how to select Q-D or Q̄-D connection to chain each two consecutive scan cells for all segments in detail. The proposed selective Q̄-D connection flow is shown in Figure 4. First of all, we will show the calculation of WT_j/NWT_j , which denotes the the total number of weighted transitions/nontransitions between the j^{th} and $(j+1)^{th}$ scan cells in a given segment after scan partitioning for scan-in/out all test stimuli and responses. Assume each segment has $L_s = L/s$ scan cells, according to the scan-shift power metric WTM [27][28] described above, WT_j and NWT_j $(1 \le j \le L_s - 1)$ can be calculated as (4) and (5), respectively.

$$WT_j = j \cdot \sum_{i=1}^{L_s}(t_{i,j} \oplus t_{i,j+1}) + (L_s - j) \cdot \sum_{i=1}^{L_s}(r_{i,j} \oplus r_{i,j+1}) \quad (4)$$

$$NWT_j = j \cdot \sum_{i=1}^{L_s}(t_{i,j} \odot t_{i,j+1}) + (L_s - j) \cdot \sum_{i=1}^{L_s}(r_{i,j} \odot r_{i,j+1}) \quad (5)$$

where $\sum_{i=1}^{L_s}(t_{i,j} \oplus t_{i,j+1})$ / $\sum_{i=1}^{L_s}(t_{i,j} \odot t_{i,j+1})$ denotes the total number of transitions/nontransitions between the j^{th} and $(j+1)^{th}$ scan cells after scan-in all test vectors, while $\sum_{i=1}^{L_s}(r_{i,j} \oplus r_{i,j+1})$ / $\sum_{i=1}^{L_s}(r_{i,j} \odot r_{i,j+1})$ denotes the total number of transitions/nontransitions after scan-out all test responses.

As shown in Figure 4, after calculating all WT_j and NWT_j of each segment for all positions between every two consecutive scan cells, we can selective Q̄-D connection to connect the j^{th} and $(j+1)^{th}$ scan cells if WT_j is greater than NWT_j like in [27], then the reconfigured WT_j will be lower than NWT_j and the scan-shift power will decrease. Otherwise, the Q-D connection will be utilized to chain these two scan cells. It should be noted that the corresponding actual test vectors and expected responses would be changed exactly according to the selective Q̄-D connection modification in segments.

TABLE I. TEST INFORMATION OF THE EXPERIMENTAL BENCHMARK CIRCUITS

Circuits	Inputs	Outputs	FFs	Gates	Faults	Tests	FC%
s5378	35	49	179	2779	4603	257	99.13
s9234	36	39	211	5597	6927	371	93.48
s13207	62	152	638	7977	9815	467	98.46
s15850	77	150	534	9775	11725	430	96.68
s35932	35	320	1728	16353	39094	65	89.81
s38417	28	106	1636	22257	31180	885	99.47
s38584	38	304	1426	19405	36303	672	95.85
b14	32	54	245	9821	22802	934	99.22
b15	36	70	449	8437	21988	627	96.29

IV. EXPERIMENTAL RESULTS

To validate the efficiency of the proposed scan-shift test power reduction scheme, experiments on several comprehensive large full scanned ISCAS'89 [29] and ITC'99 [30] benchmark circuits have been performed. The proposed test power reduction algorithm for simulation was implemented in MATLAB language, and the test patterns and corresponding expected responses utilized for experiments were generated by ATALANTA [31] (ATPG program developed at the Virginia Polytechnic Institute and State University) with X-bits random filling.

The test information of the experimental benchmark circuits is shown in Table I. The first column lists the names of the experimental benchmark circuits. Columns *Inputs*, *Outputs*, *FFs* and *Gates* show the numbers of inputs, outputs, scan flip-flops and gates of each circuit, respectively. Columns *Faults*, *Tests* and *FC%* present the number of collapsed stuck-at faults,

TABLE II. EXPERIMENTAL RESULTS OF THE SCAN-SHIFT POWER REDUCTION

Circuits	Chains	Segments=1			Segments=4				Segments=10			
		WTM_org	Proposed	Red%	WTM[24]	Proposed	Red1%	Red%	WTM[24]	Proposed	Red1%	Red%
s5378	2	1943787	1830353	5.84	475445	446129	77.05	6.17	189055	179094	90.79	5.27
	4	957020	894791	6.50	236783	222825	76.72	5.89	97569	92744	90.31	4.95
	8	472727	444311	6.01	119366	113788	75.93	4.67	51077	49093	89.61	3.88
s9234	2	4156921	3920896	5.68	1030628	976505	76.51	5.25	415022	394734	90.50	4.89
	4	2078919	1963258	5.56	517131	490474	76.41	5.15	208462	198392	90.46	4.83
	8	1031100	977132	5.23	258703	246105	76.13	4.87	106534	103102	90.00	3.22
s13207	2	47429139	45694727	3.66	11830483	11416182	75.93	3.50	4744012	4565319	90.37	3.77
	4	23687048	22857670	3.50	5919628	5701246	75.93	3.69	2371718	2289642	90.33	3.46
	8	11828258	11417421	3.47	2955846	2857563	75.84	3.33	1184348	1147993	90.29	3.07
s15850	2	29954492	28916310	3.47	7508937	7231576	75.86	3.69	2999171	2912128	90.28	2.90
	4	14993840	14474703	3.46	3764599	3638574	75.73	3.35	1504874	1460581	90.26	2.94
	8	7508809	7233919	3.66	1883954	1818053	75.79	3.50	757582	733894	90.23	3.13
s35932	2	44583480	43197336	3.11	11136528	10800522	75.77	3.02	4466048	4319577	90.31	3.28
	4	22283208	21598096	3.07	5575176	5396458	75.78	3.21	2235874	2170487	90.26	2.92
	8	11136528	10800522	3.02	2785266	2698430	75.77	3.12	1118348	1083776	90.27	3.09
s38417	2	560130270	545065774	2.69	140001136	136904852	75.56	2.21	55957232	54708300	90.23	2.23
	4	279915781	272803452	2.54	69730362	67882401	75.75	2.65	28009380	27315453	90.24	2.48
	8	139987369	136913057	2.20	34888702	34072528	75.66	2.34	14015059	13706833	90.21	2.20
s38584	2	341905860	328755899	3.85	85430686	82361649	75.91	3.59	34207213	32932619	90.37	3.73
	4	171103848	164558597	3.83	42764732	41220279	75.91	3.61	17112858	16484005	90.37	3.67
	8	85433488	82374315	3.58	21379404	20627752	75.86	3.52	8554334	8255648	90.34	3.49
b14	2	12764758	12490791	2.15	3213032	3152867	75.30	1.87	1303873	1274219	90.02	2.27
	4	6503977	6359161	2.23	1621931	1583124	75.66	2.39	654453	640416	90.15	2.14
	8	3211276	3153979	1.78	813013	798713	75.13	1.76	331315	326930	89.82	1.32
b15	2	30278294	29451219	2.73	7615643	7426652	75.47	2.48	3051162	2975560	90.17	2.48
	4	15221233	14811558	2.69	3806116	3708708	75.63	2.56	1530697	1491193	90.20	2.58
	8	7615643	7426652	2.48	1902929	1862893	75.54	2.10	772079	753689	90.10	2.38
Avg.				3.63			75.87	3.46			90.24	3.21

test patterns and the fault coverage obtained from ATALANTA ATPG tool for each benchmark circuit.

Table II illustrates the experimental results. Column *Chains* lists the number of scan chains of each circuit used for simulation, which ranges from 2 to 8. Columns 3-5, 6-9 and 10-13 present the experimental results of each circuit after each scan chain partitioned into 1, 4 and 10 segments, respectively. The total original WTM calculated by (3) for each original scan circuit under different number of scan chains is shown in column *WTM_org*, and the WTM after scan chain partitioned as in [24] is shown in column *WTM*[24], while the WTM of the proposed scheme is listed in column *Proposed*. To clearly illustrate the efficiency of the proposal, the scan-shift power reduction percentage of the proposal with respect to the original scan circuit and scan partition in [24] are presented in columns *Red1%* and *Red%*, respectively. The last row of TABLE II gives the average percentage of scan-shift power reduction under different scan partition.

As shown in Table II, compared with the original scan CUT, the proposed scheme can obtain scan-shift power reduction about 75.87% and 90.24% on average under partitioned segments number in 4 and 10 respectively, which is a little more than $\frac{s-1}{s} \times 100\%$ (s is the number of partitioned segments). While compared with the scan partition in [24], with the help of the selective $\bar{Q}$-D connection technique, the proposed scheme can achieve 3.61%, 3.46% and 3.21% scan-shift power reduction on average for different circuit with different scan chains under scan partitioned segments number in 1, 4 and 10, respectively. The average scan-shift power reduction under different partitioning is about 3.43%. Furthermore, the proposal need no additional hardware or routing overhead over [24] to reduce the test power further. Therefore, the proposed scheme is a efficient scan-shift power reduction technique. It should be noted that, a careful trade-off between hardware overhead and scan-shift test power reduction should be made before determining the number of partitioned segments for each scan chain, since more segments means more hardware overhead and higher power reduction together.

V. CONCLUSION

Scan partitioning is an attractive technique to reduce the scan-shift test power. In this paper, the selective $\bar{Q}$-D connection technique was introduced after scan partition to further reduce the scan-shift power. Since $\bar{Q}$-D connection only exploit the $\bar{Q}$ output of the scan flip-flops instead Q to scan the test data to the next scan flip-flop, it won't require any additional test hardware or routing overhead. The experimental results indicate that the proposal can achieve more about 3.43% scan-shift power reduction on average after introducing selective $\bar{Q}$-D connection technique to scan chain partitioned segments. Furthermore, the proposed scheme keeps

 ISBN: 978-1-61208-379-7

all advantages of [24], which is suitable for BIST and non-BIST test environments without penalty of test quality or performance degradation of the CUT.

ACKNOWLEDGMENT

This work is supported by Hunan Provincial Innovation Foundation For Postgraduate (NO. CX2012B031) and the National Natural Science Foundation of China (Grant No. 61176030).

REFERENCES

[1] J. Rabaey, "Low Power Design Essentials (Integrated Circuits and Systems)," Springer, 2009.

[2] Y. Zorian, "A Distributed BIST Control Scheme for Complex VLSI Devices," in Proc. of IEEE VLSI Test Symposium, 1993, pp. 4-9.

[3] P. Girard, "Survey of Low-Power Testing of VLSI Circuits," IEEE Design & Test of Computers, vol. 19, no. 3, 2002, pp. 82-92.

[4] P. Basker and A. Arumurugan, "Survey of Low Power Testing of VLSI Circuits," in Proc. of International Conference on Computer Communication and Informatics, 2012, pp. 1-7.

[5] G. S. Kumar and K. Paramasivam, "Test Power Minimization of VLSI Circuits: A Survey," in Proc. of International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-6.

[6] S. Ravi, "Power-aware Test: Challenges and Solutions," in Proc. of International Test Conference, Lecture 2.2, 2007, pp. 1-10.

[7] X. Wen et al., "A Highly-Guided X-Filling Method for Effective Low-Capture -Power Scan Test Generation," in Proc. of IEEE International Conference on Computer Design, 2006, pp. 251-258.

[8] S. Remersaro et al., "Preferred fill: A scalable method to reduce capture power for scan based designs," in Proc. of International Test Conference, 2006, pp. 1-10.

[9] J. Li, Q. Xu, Y. Hu, and X. Li, "iFill: An impact-oriented X-filling method for shift- and capture-power reduction in at-speed scan-based testing," in Proc. of Design, Automation and Test in Europe Conference and Exhibition, 2008, pp. 1184-1189.

[10] S. Wang and S. K. Gupta, "LT-RTPG: A new test-per-scan switching activity," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 25, no. 8, 2006, pp. 1565-1574.

[11] X. Lin and J. Rajski, "Adaptive Low Shift Power Test Pattern Generator for Logic BIST," in Proc. of Asian Test Symposium, 2010, pp. 355-360.

[12] Y. Sato, S. Wang, T. Kato, K. Miyase, and S. Kajihara, "Low Power BIST for Scan-shift and Capture Power," in Proc. of Asian Test Symposium, 2012, pp. 173-178.

[13] P. Girard, L. Guiller, C. Landrault, and S. Pravossoudovitch, "A Test Vector Ordering Technique for Switching Activity Reduction During Test Application," in Proc. of IEEE Great Lakes Symp. on VLSI, 1999, pp. 24-27.

[14] J. Tudu, E. Larsson, V. Singh, and V. Agrawal, "On Minimization of Peak Power for Scan Circuit during Test," in Proc. of IEEE European Test Symposium, 2009, pp. 25-30.

[15] S. Wang and S. K. Gupta, "ATPG for heat dissipation minimization during test application," IEEE Trans. Comput., vol. 47, no. 2, 1998, pp. 256-262.

[16] N. Ahmed, M. Tehranipoor, and V. Jayaram, "Supply voltage noise aware ATPG for transition delay faults," in Proc. of IEEE VLSI Test Symposium, 2007, pp. 179-186.

[17] S. M. Saeed and O. Sinanoglu, "Expedited Response Compaction for Scan Power Reduction," in Proc. of IEEE VLSI Test Symposium, 2011, pp. 40-45.

[18] S. Ghosh, S. Basu, and N. Touba, "Joint Minimization of Power and Area in Scan Testing by Scan Cell Reordering," in Proc. of IEEE Computer Society Annual Symposium on VLSI, 2003, pp. 246-249.

[19] N. Badereddine et al., "Scan Cell Reordering for Peak Power Reduction During Test Cycles," in Proc. of IFIP International Federation for Information Processing, 2007, pp. 267-281.

[20] S. Almukhaizim, E. AlQuraishi, and O. Sinanoglu, "Test power reduction via deterministic alignment of stimulus and response bits," in Proc. of 12th Latin American Test Workshop, 2011, pp. 1-6, 27-30.

[21] M. H. Chiu and J. C. M. Li, "Jump Scan: A DFT Technique for Low Power Testing," in Proc. of IEEE VLSI Test Symposium, 2005, pp. 277-282.

[22] A. Chandra, F. Ng, and R. Kapur, "Low Power Illinois Scan Architecture for Simultaneous Power and Test Data Volume Reduction," in Proc. of Design Automation and Test in Europe Conference, 2008, pp. 462-467.

[23] H. S. Kim, C. G. Kim, and S. Kang, "A New Scan Partition Scheme for Low-Power Embedded Systems," ETRI Journal, vol. 30, no. 3, 2008, pp. 412-420.

[24] E. Arvaniti and Y. Tsiatouhas, "Low Power Scan by Partitioning and Scan Hold," in Proc. of IEEE International Symposium on Design and Diagnostics of Electronic Circuits & Systems, 2012, pp. 262-265.

[25] S. Gupta, T. Vaish, and S. Chattopadhyay, "Flip-flop chaining architecture for power-efficient scan during test application," in Proc. of Asia Test Symposium, 2005, pp. 410-413.

[26] L. F. Chen and A. J. Cui, "A Power-Efficient Scan Tree Design by Exploring the Q'-D Connection," in Proc. of IEEE International Symposium on Circuits and Systems, 2013, pp. 1018-1021.

[27] R. Sankaralingam, R. R. Oruganti, and N. A. Touba, "Static Compaction Techniques to Control Scan Vector Power Dissipation," in Proc. of IEEE VLSI Test Symposium, 2000, pp. 35-40.

[28] A. Chandra and K. Chakrabarty, "Low-Power Scan Testing and Test Data Compression for System-on-Chip," IEEE Tran. on CAD of Integrated Circuits and Systems, vol. 21, no. 5, 2002, pp. 597-604.

[29] F. Brglez, D. Bryan, and K. Kozminski, "Combinational profiles of sequential benchmark circuits," in Proc. of IEEE International Symposium on Circuits and Systems, 1989, pp. 1929-1934.

[30] S. Davidson, "ITC'99 Benchmark Circuits-Preliminary Results," in Proc. of International Test Conference, 1999, pp. 1125.

[31] H. K. Lee and D. S.Ha, "On the generation of test patterns for combinational circuits," Dept Elect Eng, Virginia Polytechnic Inst State Univ, Blacksburg, Virginia, Tech Rep, 1993, pp. 12-93.

 ISBN: 978-1-61208-379-7

A Study of Transparent On-chip Instruction Cache for NV Microcontrollers

Dahoo Kim, Itaru Hida, Eric S. Fukuda, Tetsuya Asai and Masato Motomura

Graduate School of Information Science and Technology
Hokkaido University
Sapporo, Hokkaido, Japan
Email: { kim@lalsie., hida@lalsie., fukuda@lalsie., asai@, motomura@ } ist.hokudai.ac.jp

***Abstract*—Demands for low energy microcontrollers have been increasing in recent years. Since most microcontrollers achieve user-programmability by integrating non-volatile (NV) memories, such as flash memories for storing their programs, the large power consumption required in accessing an NV memory has become a major problem. This problem becomes even critical when lowering the power-supply voltage of NV microcontrollers to achieve power and energy reduction. In this paper, we try to solve this problem by introducing an instruction cache and thus reducing NV memory access frequency. Unlike general-purpose microprocessors, it is important for microcontrollers used for real time applications in embedded systems that the program execution time can be calculated accurately prior to its execution. Therefore, we introduce a "transparent" instruction cache, which does not change the existing NV microcontroller's cycle-level execution time, for reducing power and energy consumption, but not for improving the processing speed. We have conducted detailed microarchitecture design based on a major industrial microcontroller architecture, and studied, as a preliminary evaluaion, hit rates of several instruction cache configurations.**

Keywords–embedded system; micro-controller; instruction cache; non-volatile; low power design.

I. INTRODUCTION

In recent years, sensor networks have been widely studied as a fundamental technology to realize the "Smart Society" [1]. In order to implement sensor networks in various application fields, sensor nodes which can operate for a long time with a small energy source are required. Therefore, it is necessary to reduce power consumption of the microcontroller operating a sensor node.

Meanwhile, NV microcontrollers (microcontrollers integrated with non-volatile memories) are widely used due to its convenience to develop embedded system's software. However, the power consumption of the non-volatile memory dominates the total power consumption of the microcontroller [2]. Furthermore, it is hard to reduce the power consumption of non-volatile memories in microcontrollers. Focusing on this point, the purpose of our work is to reduce power and energy consumption by introducing an instruction cache to the microcontroller, reducing access to the non-volatile memory.

Besides, as shown in Table I, the traditional cache architecture research has aimed at improving the performance of the microprocessor by introducing the high-speed cache memory (SRAM) between the main memory and the datapath by reducing the memory access time [3]. On the other hand, in the case of NV microcontroller used for real time applications in embedded systems, it is important that the program execution time can be calculated in advance. Furthermore, the change of execution time due to cache misses should be avoided since such a change can cause problems to the system such as real time applications. Thus, it is necessary to introduce an instruction cache, which does not cause cache miss penalties, while leaving the speed of the memory access at cache hits unchanged.

TABLE I. CONCEPT OF THIS WORK

Traditional Cache Architecture Research	**This work**
Micro Processor: Datapath, I-Cache (SRAM), D-Cache (SRAM), Main Memory (SRAM, DRAM)	NV Micro Controller for Embeded System: Datapath, IM (NVM), DM (SRAM); Datapath, I-Cache (SRAM), IM (NVM), DM (SRAM)
High Performance	Low power & Low Energy Transparent (i.e., doesn't change cycle-level timing)

Therefore, in this paper, we aim at lowering power and energy consumption rather than improving the performance by introducing a transparent instruction cache, which does not change cycle level timing of existing NV microcontrollers.

The rest of this paper is organized as follows: Section 2 describes the features of our research. Section 3 describes the architecture of the proposed instruction cache employed to the NV microcontroller. Section 4 discusses preliminary evaluation of the proposed instruction cache. Section 5 summarizes future works.

II. FEATURES OF OUR RESEARCH

The Features of our Research are as follows:

1. **Examination of instruction cache suitable for microcontroller deployments** - As described in Section 1, in the case of a microcontroller, it is important to prolong the battery run time than to reduce the processing time. For this reason, the instruction cache that we propose in this paper is intended to reduce the power consumption rather than to improve the processing speed.

2. **Evaluation based on a realistic microcontroller architecture** - We evaluated the power and energy consumption of our system built on a base microcontroller, Renesas Electronics Corporation's 78K0R. Since not all of 78K0R's specifications

 ISBN: 978-1-61208-379-7

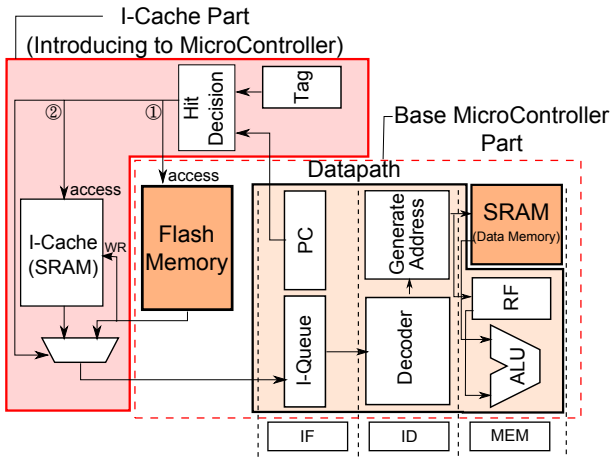


Figure 1. Concept of 1-word-per-line instruction cache architecture

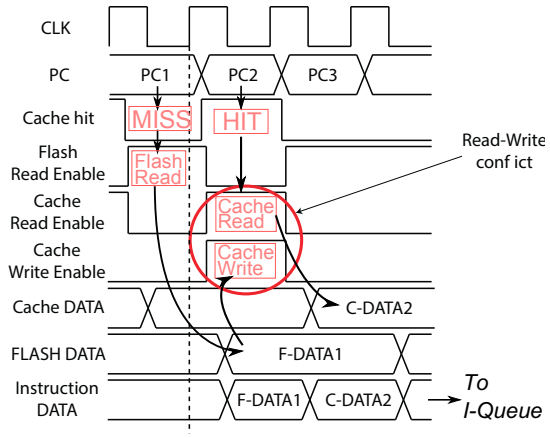


Figure 2. Instruction memory access timing in the case of cache hit immediately after cache miss

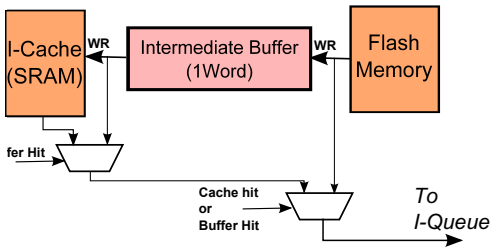


Figure 3. Intermediate buffer insertion method for 1-word-per-line instruction cache architecture

are open, we implemented the base microcontroller using only the publicly available information [5] [6] [7].

III. ARCHITECTURE

A. Base Microcontroller

Our microcontroller is based on 78K0R, which is widely used in various industrial fields [4], whose block diagram is shown as a part of Figure 1. 78K0R has a flash memory as its NV instruction memory and an SRAM as its data memory [5].

This architecture has a pipeline structure of three-stages (IF stage, ID stage, MEM stage) [6]. In the IF stage, the microcontroller provides an address from the program counter (PC) to the instruction memory (flash memory) and fetches an instruction sequence from the instruction memory. This instruction sequence is stored in an instruction queue (I-Queue). In the ID stage, the microcontroller decodes the instruction that has been fetched in the IF stage and extracts data memory's (SRAM) and RF's (Register File) addresses to be accessed. In the MEM stage, the microcontroller retrieves the data from the data memory and executes the instruction.

Also, the base microcontroller adopts the CISC architecture [6]. The number of the instructions of the base microcontroller is 915, and the instruction length is 1 byte to 5 bytes. Base microcontroller has four-byte (1 word) instruction queue (I-Queue) that contains an instruction sequence fetched from the instruction memory. Therefore, if a valid instruction exists in the I-Queue, there is no need to access the flash memory.

B. 1-word-per-line instruction cache

The base microcontroller fetches one word from the instruction sequence in the flash memory in one cycle [6]. Thus, as the first step to introduce the instruction cache to the base microcontroller, we designed a 1-word-per-line instruction cache architecture, as shown in Figure 1. We assumed that the operation timing of the instruction cache to be the same as the timing of the base microcontroller's access to the flash memory: The instruction sequence is read from flash memory in the subsequent cycle of the access to the flash memory.

Actual operation of the instruction cache is as follows: (1) In the case of a cache miss, the microcontroller accesses the flash memory and fetches 1 word from the instruction sequence. (2) In the case of a cache hit, the microcontroller accesses the instruction cache (I-Cache) and fetches 1 word from the instruction sequence in the instruction cache as well as the flash memory. Also, in the case of a cache miss, the instruction sequence which is read from the flash memory is written to the instruction cache. Since the bit widths of the instruction cache and the flash memory are the same, there is no penalty for writing to the instruction cache.

Since the instruction cache we propose does not allow cache miss penalty, there is a problem that a read access and a write access to the instruction cache can occur coincidentally. For example, as shown in Figure 2, when a cache miss occurs at PC1, microcontroller reads the flash memory. Then, in the next cycle, the instruction sequence is read from the flash memory. Thus, the instruction sequence for the cache miss (F-DATA1) must be written to the instruction cache in this cycle (the next cycle of the cache miss). In the cycle immediately after the cache miss, when a cache hit occurs at PC2, an instruction is fetched from the instruction cache. In this case, the collision of a read access and a write access to the instruction cache occurs.

For solving this problem, we designed **intermediate buffer insertion method** for the transparent instruction cache. This is a method to delay the write access to the cache by inserting an intermediate buffer to which the instruction sequence in the cache miss is written between the instruction cache and the flash memory. Figure 3 shows its architecture. In this method, the timing of writing the instruction sequence to the instruction cache is a cycle that does not access the instruction cache. If a valid instruction is still in the I-Queue of the datapath as described in the base microcontroller architecture, and if a

 ISBN: 978-1-61208-379-7

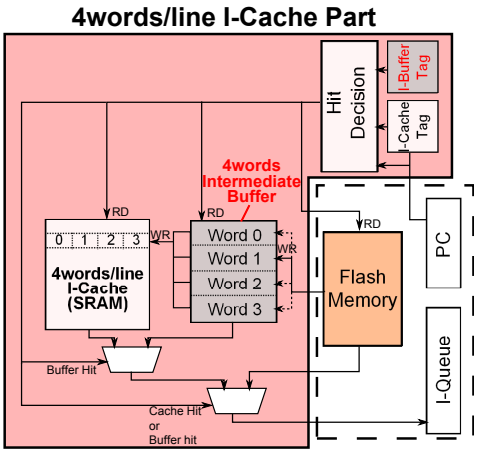


Figure 4. 4-word-per-line instruction cache architecture

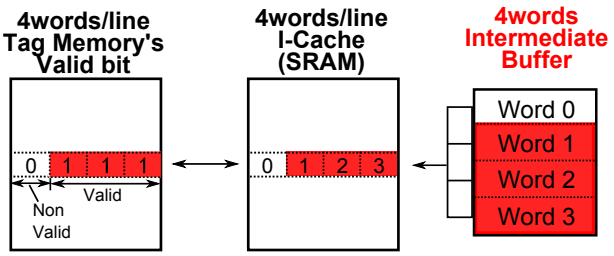


Figure 5. Writing operation of 4-word-per-line instruction cache

cache miss occurred, the microcontroller does not have access to the instruction cache.

Particularly, when a cache miss occurred, the instruction sequence that has been present in the intermediate buffer is written to the instruction cache and the instruction sequence that has been read from the flash memory is newly written to the intermediate buffer. Therefore, the intermediate buffer will not overflow with the buffer with only 1 word. Also, the *intermediate buffer insertion method* determines a cache hit in the intermediate buffer, and enables the instruction sequence to be read from the intermediate buffer.

C. 4-word-per-line I-Cache

As the second step to integrate the instruction cache to the base microcontroller, we designed 4-word-per-line instruction cache architecture that can take advantage of spatial locality.

The flash memory (NV memory) of the base microcontroller (existing NV microcontroller) reads one word from the instruction sequence in one cycle. Therefore, in order to implement the 4-word-per-line instruction architecture, a buffer for storing four words is required. We have implemented 4-word-per-line instruction cache by extending the number of words of the intermediate buffer, as shown in Figure 4. The instruction sequence that is read from the flash memory in one cycle is stored in the intermediate buffer one word per cycle. When all the four entries of the intermediate buffer are filled, the contents are sent to the 4-word-per-line instruction cache memory.

For example, when a miss occurs in the first cycle, the instruction sequence that is read from the flash memory is stored to one of the entries of the intermediate buffer by referring to the lower 2 bits of the instruction address (held in the PC). In the next cycle, if the flash memory is not accessed, in other words, if there are cache hits, buffer hits or I-Queue hits, the next instruction sequence is read from the flash memory by incrementing the PC and is stored to the next entry of the intermediate buffer. In this way, when four words are collected in the intermediate buffer, four words of the instruction sequence will be written to the instruction cache memory.

However, when a miss occurs before collecting the four words in the intermediate buffer, it is necessary to store the instruction sequence that has been read from the flash memory to the intermediate buffer. In this case, the existing instruction sequences in the intermediate buffer is written to the instruction cache, and the instruction sequence that has been read from the flash memory on a miss is stored to the intermediate buffer. For example, as shown in Figure 5, if a miss occurs when only three words of instruction sequence are stored in the intermediate buffer, the three words are written to the instruction cache, invalidating the 0-th word. In this paper, we extend the 4-word-per-line tag memory's valid bits to 4 bits, and make them indicate the valid word of each line.

However, when writing the line which has invalid words to the instruction cache, such as the case shown in Figure 5, it is expected that there is a case of a low hit rate, due to the deletion of the valid word that existed in one line of the instruction cache. We will evaluate our system in this regard.

D. Associativity

As the third step to integrate the instruction cache to base microcontroller, we have increased the associativity to 2 and 4 from 1. It is expected that the power consumption consumed by the control unit of the tag memory is increased because the control unit of the tag memory becomes complicated by increasing the associativity. Yet, there is a possibility that the hit rate will rises. We will evaluate our system in this regard as well. Also, we adopt a pseudo least recently used (LRU) replacement algorithm.

IV. Preliminary Evaluation

As the first step to evaluate the effect of proposed instruction cache on reducing power and energy consumption, we evaluated a hit rate on a few benchmark programs. It is expected that the effect increases with an increase in hit rate because of decreased access to a flash memory. Benchmark programs and each program's size are shown in Table II.

First, the hit rates of 1-word-per-line instruction cache for each benchmark programs are shown in Figure 6. Because the program sizes are small, they, from bubble sort program to factorial program, showed high hit rates for every cache size.

TABLE II. BENCHMARK PROGRAM AND PROGRAM'S SIZE

Evaluation program	size[Byte]
bubble sort	614
Celsius to Fahrenheit	585
Checksum	535
Copy verify	602
Factorial	606
EEMBC Coremark	11701

 ISBN: 978-1-61208-379-7

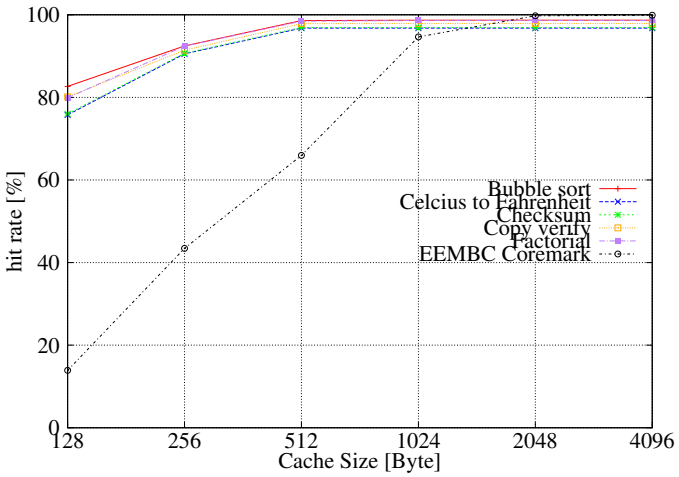


Figure 6. Cache hit rate by cache size for each programs (1-word-per-line)

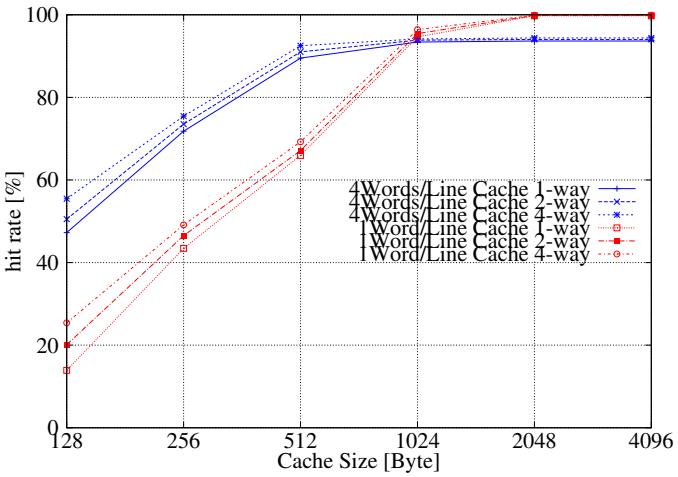


Figure 7. Cache hit rate by cache size (Evaluation program : EEMBC Coremark)

Hereafter, we will evaluate more detailed power consumption using EEMBC Coremark program.

In EEMBC Coremark program, the hit rate of each cache size is shown in Figure 7. With a large sized instruction cache (1 Kbytes or more), 1-word-per-line instruction cache has higher hit rate than 4-word-per-line instruction cache. This is because there is also a case that a valid data which was in the instruction cache in 4-word-per-line instruction cache has been discarded as described in 4-word-per-line instruction cache architecture. In the case of instruction cache of less than 1 Kbytes, 4-word-per-line instruction cache has a higher hit rate. This is because 4 word-per-line instruction cache can take advantage of spatial locality, even if the valid data has been discarded. Also, if the associativity is high, hit rate was high in both methods. Especially, when the cache size was small, its effect became significant.

Based on the above results, we will evaluate a more detailed power and energy consumption.

V. Conclusion and Future Work

In this paper, we proposed the transparent instruction cache architecture for reducing power and energy consumption of a practical NV microcontroller architecture. Unlike traditional cache architecture researches, we intended to reduce power and energy consumption rather than to improve processing speed. This is because, in the case of NV microcontroller used for real time application in embedded systems, it is important to prolong the battery run time and not to change the existing NV microcontroller's cycle-level execution time.

Our future work is to evaluate the effect of proposed instruction cache architecture by estimating more detailed power and energy consumption of this architecture. In order to estimate a more detailed power and energy consumption, we will develop an RTL description of the base microcontroller, and will integrate the proposed instruction cache architecture to the RTL description. We can generate back-annotated netlists after logic synthesis and placement/routing, then, using which power and energy reduction of the proposed architectures can be evaluated fairly accurately.

References

[1] V. C. Gungor, Bin Lu, and G. P. Hancke, "Opportunities and Challenges of Wireless Sensor Networks in Smart Grid", Industrial Electronics, IEEE Transactions on, pp. 3557-3564 , October, 2010.

[2] H. G. Lee, and N. Chang, "Energy-aware memory allocation in heterogeneous non-volatile memory systems", in Proc. of the 2003 international symposium on Low power electronics and design, pp. 420-423 , August, 2003.

[3] J. R. Goodman, "Using cache memory to reduce processor-memory traffic", in Proc. of the 10th annual international symposium on Computer architecture, pp. 124-131 , June, 1983.

[4] K. Oba, K. Kawai, R. Matsushita, K. Ishihara, and K. Eto, "Development of 16-bit All Flash Microcomputers " 78K0R/Kx3-L "Featuring Ultralow Power Consumption", Nec Technical Journal, 4(1), 35: , March, 2009.

[5] Renesas Electronics Corporation, "Renesas Demonstration Kit for RL78G14 User's Manual", http://documentation.renesas.com/doc/products/tools/r20ut2534eu0200_yrdkrl78g14_um.pdf , October, 2013.

[6] Renesas Electronics Corporation, "RL78 Family User's Manual: Software", http://documentation.renesas.com/doc/products/mpumcu/doc/rl78/r01us0015ej0210_rl78.pdf, January, 2014.

[7] Renesas Electronics Corporation, "78K0R/Hx3 User's Manual: Hardware", http://documentation.renesas.com/doc/products/mpumcu/doc/78k/r01uh0260jj0300_78k0rhx3.pdf, September, 2013.

 ISBN: 978-1-61208-379-7