Linking Aging Measurements of Health-Monitors and Specifications for Multi-Processor SoCs

Hans G. Kerkhoff, Jinbo Wan & Yong Zhao
Testable Design and Test of Integrated Systems (TDT Group)
University of Twente, Centre of Telematics and Information Technology (CTIT)
Enschede, the Netherlands
h.g.kerkhoff@utwente.nl

***Abstract*—A new generation of highly dependable multi-processor Systems-on-Chip for safety-critical applications under harsh environments with zero down-time is emerging. In this paper[1], the approach towards reaching this ultimate goal is explained. Crucial is this method is linking the measurement data of so-called (on-chip) health monitors during life time with the measurements of degrading key performance parameters of the cores involved. The focus will be here on processor cores, with delay as one of the most critical aging dependent parameters. An extensive (accelerated)-test program was set-up to evaluate the aging of both the health monitors as well as delay of an industrial reconfigurable processor core in harsh environments. The correlation between them will serve as the basis of real-time on-chip health-monitoring based prognostics for life-time prediction, enabling a zero down-time for safety-critical applications.**

Keywords—Dependable MP-SoC systems; prognostics; real-time on-chip health-monitoring; accelerated reliability testing; specification-based testing.

I. INTRODUCTION

The trend of downscaling and increased complexity of digital ICs [1] has enabled the implementation of many-processors (MP) in complex SoCs. Especially in the case of homogeneous MP-SoCs, this turns out to be an extremely nice feature in terms of reliability. However, the downside of downscaling and complexity is the increase in variability and decrease in reliability of the components [1, 2]. To counteract this loss, use can be made of multi-processors. Our generic approach for implementing high-dependability SoCs uses on-chip health monitor tests or measurements on processor cores during their operational life to evaluate their health and subsequent repair of (to be) faulty processors by remapping and rerouting (spare) correct cores using run-time mapping software [3, 4, 5]. Combining health monitors and prognostics in electronics as such has been suggested before [6].

This paper deals with enhancing (long-term) dependability of many-processor SoCs being used in high-level security and automotive applications[1]. In our case, aging faults (e.g. NBTI [7]) could degrade system performance under *harsh* environmental conditions, such as temperature and voltage stress. At some moment, this can finally result in a system failure due to permanent faults. In earlier work, e.g. paper [3], the processor cores were structurally tested via scan chains during their life time. This has the disadvantage that during those tests, the processor cores have to be isolated (non-operational) to be tested; furthermore, *reactive* repair via remapping is performed only *after* fault detection via testing. As a consequence, the mean down time can be considerable

The new generation of highly dependable multi-processor systems monitors the cores via on-chip health monitors (HM) during their operational life [4, 5], enabling a *proactive* repair and hence *zero* mean down time. In addition the latter approach also has shown to feature a lower dependability test power-dissipation [8], which is a big advantage especially in mobile applications. It is a good example of the new generation power-aware high-dependability systems.

Our prognostic approach for life-time prediction of cores uses on-chip health monitors (e.g. supply-voltage monitoring) per core in combination with advanced prediction algorithms in software to ensure a high dependability, using the same *repair* mechanisms as before [3]. However, this approach assumes that there is a close *correlation* between the on-chip HM measurements and key core specification parameters, like e.g. the maximum operating clock speed or the dynamic power current, as function of time (aging). In other words the final goal is that only the on-chip set of HMs will accurately *predict*, together with embedded life-time prediction software, when cores are expected to fail in time. In this advanced way, a timely replacement can be made. It is much more efficient than an automatic scheduled repair action determined at *design* time.

This paper is organized as follows. In section II, the basic principle is explained how to obtain the degree of correlation between a set of health monitors measurements and key specification parameters of processor cores during aging. Our target processor core is an (embedded) Xentium™ reconfigurable DSP core, implemented in 90nm CMOS technology by Recore Systems (Figure 1). The next section

[1] This research has been conducted as part of the Sensor Technology Applied in Reconfigurable systems for sustainable Security (STARS) project and the ENIAC "ELESIS" project (co)financed by the Netherlands Enterprise Agency (RVO).

describes the organization of the accelerated reliability tests that are carried out on both health monitors as well as the Xentium processor core. It includes conditions as well as the test set-up. In section IV, the health monitors chip from Ridgetop Europe in 90nm CMOS is introduced and typical measurement results are shown in the case of an NBTI monitor. Subsequently, the problems encountered in accessing and measuring the (embedded) Xentium processor key specifications are treated in section V. Finally, as an example, a link in terms of correlation is made between delay degradation via the NBTI HM and the expected delay degradation in the Xentium processor IP.

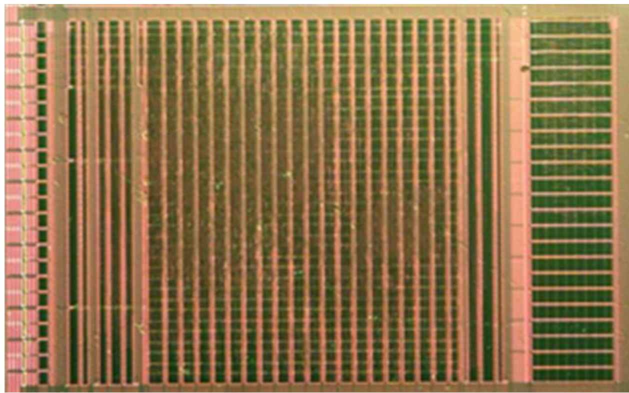


Figure 1. Photomicrograph of the embedded target Xentium™ processor IP, being part of an evaluation heterogeneous many-core SoC of Recore Systems.

II. Basic Principle

The basic idea to design highly dependable MP-SoCs has been presented most recently by us in reference [5]. It uses a number of on-chip health monitors around *each* processor core, as shown in Figure 2. The HM data is, for instance periodically, communicated to an embedded (ARM) processor, where it is being used as input to predict the life-time of that Xentium core using dedicated embedded software.

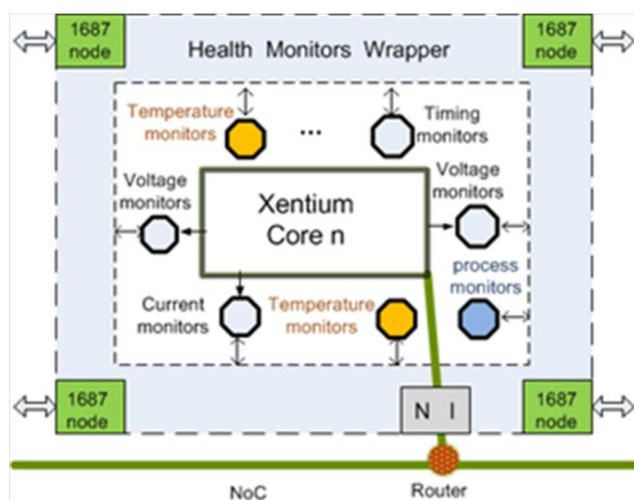


Figure 2. Symbolic drawing of a single Xentium™ core with health monitors (HM), monitors wrapper and network interface (NI). *Optionally* P1687 nodes, P1687 client/host network and TAP controller are available for measurement data communication [5].

There are several categories of health monitors. As life-time is usually defined as the time in which (crucial) performance parameters stay with their allowed boundaries, and these are normally also influenced by environmental conditions, measuring these conditions is vital. Temperature and supply voltage are examples of these: both directly influence for instance delay, and hence clock frequency, in digital circuits [9]. Beside the above environmental health monitors, *technology-related* health monitors, like ring oscillators (RO) are popular [5, 7]. Other examples of these truly non-invasive HMs are e.g. NBTI and HCI monitors. By non-invasive is meant that the integrity of the IP remains intact (worst-case, only IOs are observed which are not to be affected, e.g. by heavy loading). By regularly activate these monitors for measuring purposes, the degradation mechanisms under the *actual* stress conditions of the SoC are kept up-to-date and can be subsequently used for real-time prognostics of the core life time. An interesting example in the case of mechanical bearing degradation via vibration health monitors has been presented in [10].

Predictions just based on simulations during the *design* stage tend to be very inaccurate; one of the reasons being that the stress conditions like stress time and stress values are unknown in advance in practice. It is therefore essential in our case to calibrate/verify the life-time prediction algorithm and its coefficients, based on real on-chip HM measurement data as well as the (indirectly measured) core life time, to avoid at any cost, replacement *after* a Xentium failure. As many HMs are non-invasive care should be taken that a good correlation exists between the measured HM data and key specification parameters of the core. In this paper, two examples of important processor specification parameters are examined being the propagation delay and the I_{DDQ} / dynamic power-supply current. It has been shown by many authors e.g. [7], that a failure as result of increased delays in processors is a significant reason for a reduced life-time resulting from aging.

Hence in this paper the focus is on the investigation of the degree of correlations between real HM measurement data (temperature, power-supply voltage and process health monitors) and the delay degradation in the Xentium core as result of stressed aging. In addition also the determination of the life-time of the Xentium via accelerated reliability tests is tackled. It is obvious there should be a direct relation between the latter two.

III. Reliability Accelerated Test Plan

In order to provide measurement (and simulation) data with respect to the aging of integrated circuits (silicon, not packaging), a proper reliability test plan had to be devised [11]. Ideally, the conditions should be close to what can be expected in the mission profile of the application, which is automotive in our case. However, in order to reduce the acceptable measurement test time, *acceleration* is required in terms of stressing the temperature, power-supply as well as in our digital case, workload and clock frequency. Often,

standards are followed such as JEDEC [12, 13]. Our *primary* goal is to find *relationships* between health monitors and in our case delay (and I_{DDQ}) of the Xentium core. Later on, it is also very useful to find out at what time the core starts to actually fail as this is related to the life time. It was decided that two types of reliability tests should be carried out: first the High Temperature Operating (Bias) Life test, abbreviated by HTOL [12], sometimes referred to as the burn-in test. This is a well-known method to weed out infant mortality failures. The parameters used in HTOL are listed in TABLE I.

TABLE I. The used HTOL conditions

Number of devices	46
Stress temperature	125^0C
Stress power supply	2.5V
Stress clock frequency	220MHz
Duration	1000hrs (6 weeks)

The second test to be carried out is the Power Temperature Cycling test, abbreviated PTC [13], of which the parameters are provided in TABLE II. This test is usually referred to as shock temperature cycling, and it affects both the silicon as well as the package. This rather aggressive test could eventually lead to failure of the core; if this is the case, the life time can be estimated via known reliability calculations [14] and hence the prognostics model validated / calibrated.

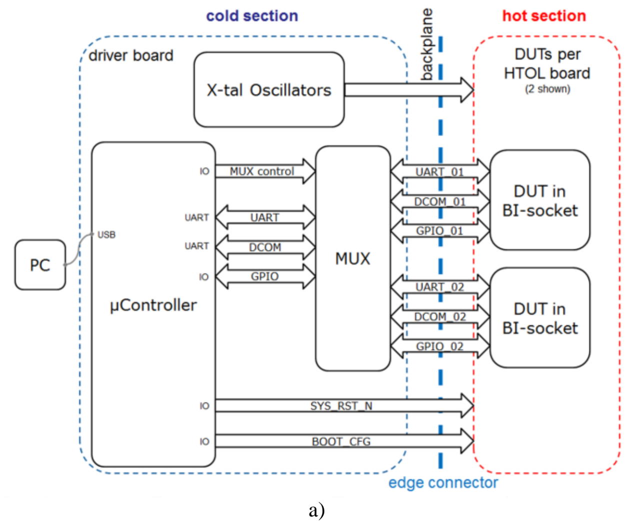
TABLE II. The used PTC conditions

Number of devices	46
Stress temperatures	-40^0C - $+150^0$C
Stress power supply	2.5V
Stress clock frequency	220MHz
Duration	1000 cycles (1000hrs, 6 weeks)

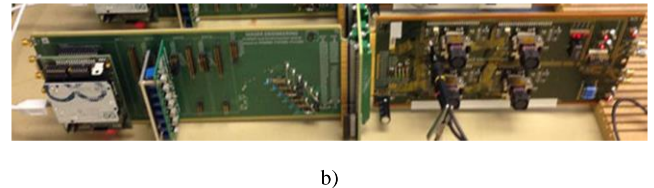
The HMs used in our case are three identical 90nm CMOS-based chips, as will be discussed later in section IV. This means that already much more than 50 devices per HM are available in total, because arrays of devices have been used per chip. Their special test board is a different, simpler design as compared to the SoC chip with embedded Xentium core. They are three 4-layer polyimide PCBs with strip lines which can withstand the high stresses. The test set-up uses cold and warm zones which are separated by 44-pins edge connectors. Because only one ProCheck [15] test system is available, special considerations for flexibility are required. This system has all resources in hardware and software to provide measurement monitor data on e.g. NBTI, TDDB, HCI, and EM etc. [15]. For ring oscillator delay tests, a separate frequency counter is used for evaluating the characteristic CMOS process delay.

The matter is different for the reliability conditions and measurements of the key parameters for the embedded Xentium core. As will be discussed later, using scan options turned out not to be feasible for a number of reasons, which seriously limits the measurement options. Hence it was decided to focus on functional capabilities only; the down side of this approach is the requirement for rather complex packages, boards and complex control. During the tests, one should get actual access to the embedded Xentium core in our SoC, which is not equipped with an IEEE 1500 infrastructure, being often seen in industrial chips. Furthermore, the Xentium cores have also to be stressed via the workload running on it; this in addition requires dedicated software.

The basic set-up is shown in Figure 3a, where the hot and cold zones can be distinguished, separated by the backplane edge connector. The driver board incorporates the crystal oscillators for the SoC, as well as a microcontroller for getting *access* as well as generating the *workload* for the Xentiums. It is connected via an USB to a PC on which dedicated software runs. An example of the driver board, edge connector and HTOL test board in practice is shown in Figure 3b.



a)



b)

Figure 3. a) Basic set-up of our Xentium (DUT) reliability test scheme
b) Example of cold driver board, edge connector and hot HTOL test board. Courtesy Maser Engineering

A number of measurements (e.g. temperature, supply-voltage) are carried out *during* the reliability tests using the driver board. For I_{DDQ} or dynamic current tests, the boards are removed from the oven and a dedicated interposer board is used for current measurements.

At a 1-week interval, more specific *external* measurements are carried out, like delay tests, using the R12 Recore Systems evaluation board, discussed in section V.

IV. THE USED HEALTH MONITORS (HM)

Key components in the new architecture of highly dependable multi-processor SoCs are health monitors [16]. In the Recore Systems 90nm CMOS Xentium design there was no possibility at that time to include local HMs. Hence, another 90nm CMOS chip was used [17], the 9SF test coupon from Ridgetop Group Inc. which includes a number of health monitors. For our purpose, three chips are being used. Each chip has 48 pins and measures 2*2mm, with 16 arrays of 64 DUTs; they include NMOS and PMOS transistors, as well as via and contact strings. Also two ring oscillators (RO) are incorporated, although they are not as advanced as our multi-sensor RO [5]. These technology-oriented HMs, together with the ProCheck test module, enable evaluation of NBTI, HCI, TDDM and EM aging effects [15]. The capabilities of the 9SF chips for local heating via on-chip poly-silicon resistor heating are not used, because of our own oven capabilities. The big advantage of this approach is the option of easy porting these HMs to our processing being used, hence enabling smooth integration later on with Xentiums. An example of measurement results in terms of cell propagation delay is shown in Figure 4, resulting from NBTI (V_{th} shift) aging. The set-up and layout of the 9SF NBTI monitor can be found in reference [17]. Figure 4 shows the propagation delay increases around 5.7% after a stress time of only 16 minutes under 125°C. The applied stress voltage is also shown.

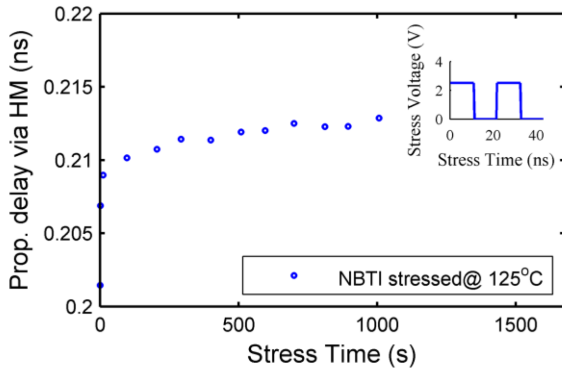


Figure 4: Example of the increased propagation delay of an inverter versus stress, based on NBTI V_{th} shift measurements. A pulse-wave stress signal has been used.

All the periodic measurement data of several HMs (e.g. NBTI, HCI and ROs) can be used later on to determine the link of their behavior with that of the Xentium key specification parameters.

V. MEASURING PERFORMANCE PARAMETERS OF THE DEEPLY EMBEDDED XENTIUM

The life-time of the Xentium processor is determined by the moment it fails to perform its specifications. As previously stated, the major cause of failure is often intolerable clock speed degradation during aging. Hence, at regular intervals (1 week), the multi-core SoCs are being evaluated *externally* using its specially designed (Recore Systems) evaluation board, as shown in Figure 5.

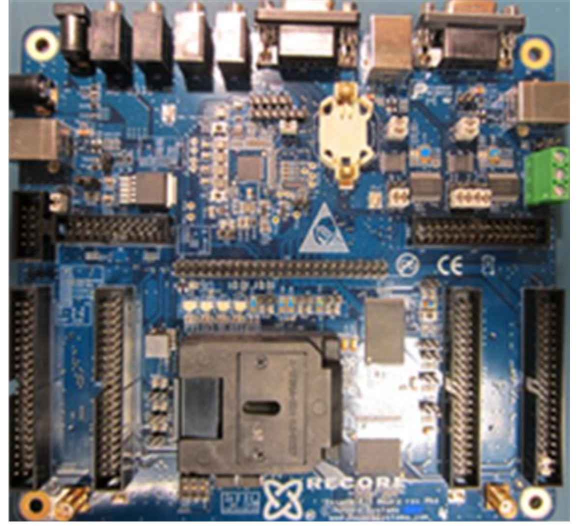


Figure 5. The R12 PCB for carrying out functional measurements of the evaluation heterogeneous multi-processor SoC of Recore Systems based on a Xentium IP. Location of SoC: bottom, middle. Courtesy Recore Systems.

Important measurements are the verification of the correct operation of the Xentium workload at the specified clock frequency. By lowering the power-supply voltage and increasing the clock frequency until the operation fails, data is obtained on the clock-frequency (resulting from delay) reduction.

At this moment, several Cadence simulations have been carried out showing the increase in the delay of the most critical path (fresh, non-aged delay 6 ns) of the Xentium processor due to aging. The stress profile has been taken the same as in Figure 4. The critical path is actually related to the multiplier in the Xentium [18]. The expected increased propagation delay times versus stress time is shown in Figure 6. Stress temperature, stress time, and voltage stress profile are indicated.

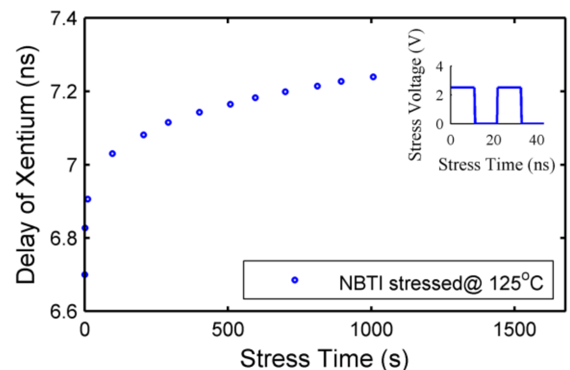


Figure 6: The increased delay (decrease in operating frequency) of the most critical path in the Xentium processor core obtained from simulation. This result will cause speed failure after some time (reduced dependability).

Actual *external* measurements have at this moment not yet been carried out using the R12 board and should confirm this simulated data.

There are however two complicating factors which should be further explained. First, the Xentium processor core is not available as a stand-alone core, but it is deeply embedded in a recent evaluation heterogeneous many-processor SoC of Recore Systems. It is basically an ARM-based design, employing an AMBA bus for data communication and incorporates several types of processor cores of which the Xentium is just one example. Data communication is taking place via an UART.

Second, via scan testing, relatively simple structural tests could be carried out on the Xentium, but also tests for determining delays, such as *launch-on-capture* transition testing [19] can be carried out. The advantages of this approach are that only a small number of pins have to be used for testing; we have even investigated the potential use of scan chains for *emulating* a large workload on the Xentium during the stress tests. Although the design allows for a 30% toggle rate for maximum power dissipation, it was concluded that this can probably not be achieved via scan chain activation.

Moreover, (scan) test access via IEEE 1500 and/or the JTAG port is not available, and the Xentium scan test has actually to take place via multiplexed GPIO pins; this is not unusual in industrial SoCs. However, the latter also reduces the allowed scan-test speed. Because of all these issues, it was decided to abstain from using scan-chains and scan-based tests.

As a consequence of the above, functional evaluation of the specifications of the embedded Xentium processor core remained. The heterogeneous many-processor SoC has 233 balls BGA which indicates the interconnection complexity in that case. A single test board houses 3 SoCs, 8 boards 24, and hence two 1000hrs reliability tests are performed sequentially. As the embedded Xentium requires several start configuration files to accesses the Xentium, as well as a special program to run on the Xentium guaranteeing a very high workload (200mW), the driver board incorporates a micro-controller and associated PC (Figure 3) for performing these operations.

Many other tests are carried out with regard to the Xentium and the health monitors, but these will not be discussed in this paper.

VI. Linking HM measurements and Xentium Performance parameters

As previously stated, the concept for high dependability SoCs we apply, is using (multiple) on-chip health monitors to accurately predict the life-time of our processor, and subsequently replacing (just in time!) a faulty core with a fault-free core [5]. The concept of using relatively simple measurements to predict key specification parameters, like dependability, to some extent resembles the well-known approach of "alternate testing" [20].

Without getting too deep in this approach in this paper, a basic requirement is that in our case the health monitoring data should sufficiently correlate with the key specifications of importance with respect to dependability/reliability. In this section, the previous NBTI delay-related data (Figure 4) is linked to the Xentium delay data (Figure 6), both under the same stress regime of voltage, temperature and stress time. This is depicted in Figure 7.
Horizontal or vertical lines show that there is no correlation at all between the two; anything in between will indicate some degree of correlation. How much correlation is required is often an object of discussion.

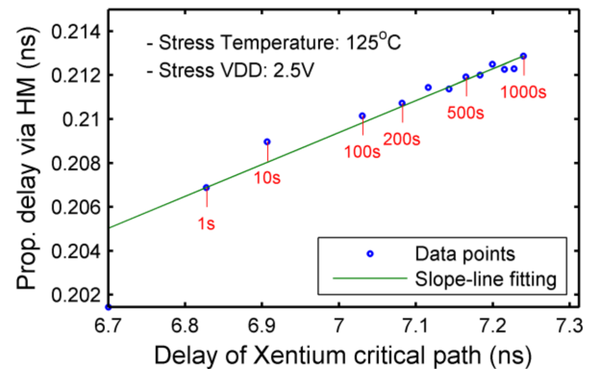


Figure 7. Delay obtained via NBTI health monitor measurements and delay (simulated) of the Xentium processor core under the same aging regime. Data points are in relation with the stress times (0 -1000s).

As can be seen from Figure 7, this correlation exists, and hence the concepts used in alternate testing, like deriving *mapping functions*, can be applied in principle. It can be further improved if multiple health sensors are incorporated which are showing correlations with the Xentium delay. The same principle also holds for other key parameters.

The advantage of our proposed reliability test program approach here, is that beside the information from the health monitors and Xentium parameters, also actual reliability measurements/calculations can be used to *calibrate* the life-time via really occurred failures (via PVT) [14]. This is rather unique, and should significantly improve the life-time prediction accuracy.

VII. Conclusions

In order to enable the implementation of high dependability SoCs with zero down-time in safety-critical applications under harsh environments, a method has been developed to establish a link between a specific set of health monitors and key performance parameters of processor cores.

It involves accelerated tests of both the health monitors and processor core after final tests. Both extensive measurement results are subsequently used to establish a (possible) correlation between them. These multiple correlations are further used for the embedded prognostics software to determine the life-time with unsurpassed accuracy. In addition, the accelerated tests on the Xentium also provide useful feedback on the accuracy of the prediction. Our approach enables the design of a new generation of very high dependability many-processors SoCs in safety-critical applications.

VIII. Acknowledgements

The authors would like to acknowledge the significant support of Hans Manhaeve of Ridgetop Europe with regard to the health monitors and the ProCheck test system. Eelke Strooisma of Recore Systems is acknowledged for the cooperation on the evaluation SoC of Recore Systems on the basis of Xentium DSPs, and Tijs Lammertink of Maser Engineering for the contributions in the reliability test approach program as well as the HTOL/PTC test-boards implementation.

References

[1] International Technology Roadmap for Semiconductors (ITRS) 2011 "Reliability & Manufacturability," (http://www.itrs.net), 2011.

[2] Y. Cao, P. Bose, and J. Tschanz, "Reliability challenges in nano-CMOS Design," IEEE Design & Test of Computers, pp. 6-7, 2009.

[3] X. Zhang and H.G. Kerkhoff, "A dependability solution for homogeneous MP-SoCs," in 17th IEEE Pacific Rim Dependability Conference (PRDC), USA, pp. 53-62, 2011.

[4] H.G. Kerkhoff and Y. Zhao, "The design of dependable flexible multi-sensory System-on-Chips for security applications," 15th IEEE Symposium on Design and Diagnostics of Electronic Circuits and Systems (DDECS), Tallinn, Estonia. pp. 133-138, 2012.

[5] Y. Zhao and H.G. Kerkhoff, "An Embedded Health-Monitoring Infra-structure for a Reliable Multi-core Processor," Proc. on Manufacturable and Dependable Multicore Architectures at Nanoscale (MEDIAN/ETS)Workshop, ISBN 978-2-11129175-1, Avignon, France, pp. 31-34, May 2013.

[6] N. M. Vichare and M.G. Pecht, "Prognostics and health management of electronics," Components and Packaging Technologies, IEEE Transactions on, vol. 29, pp. 222-229, 2006.

[7] K. Tae-Hyoung, et al., "Silicon Odometer: An On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circuits," IEEE Journal of Solid-State Circuits, vol. 43, pp. 874-880, 2008.

[8] Y. Zhao, X. Zhang and H. G. Kerkhoff, “Power-Dissipation Comparison of Two Dependability Approaches for Multi-Processor Systems”, in International Conference on Design and Technology of Integrated Systems in the Nanoscale Era (DTIS), Abu Dhabi, ISBN: 978-1-4673-6040-1/13, pp. 56 – 61, 2013.

[9] A.L. Shimi, “Kryotech SuperG Athlon 1000MHz”, AnanTech, December 1999.

[10] A. Elwany and N. Gebreel, “Real-time Estimation of Mean Remaining Life Using Sensor-Based Degradation Mechanisms”, Journal of Manufacturing Science and Engineering, ASME, vol. 13, pp. 051005-1-7, 2009.

[11] Y. Zhao, E. Strooisma, T. Lammertink and H.G. Kerkhoff, "Xentium Health Test Plan", Technical Report D3.4.c, University of Twente, the Netherlands, March 2014.

[12] JEDEC standard JESD22-A108D, http://www.jedec.org/standards-documents/, November 2010.

[13] JEDEC standard JESD22-A105C, http://www.jedec.org/standards-documents/, January 2011.

[14] M.G. Pecht and F.R. Nash, "Predicting the Reliability of Electronic Equipment,"Proceedings of IEEE, vol. 82, no. 7, pp. 992-1004, July 1994.

[15] E. Mikkola, "ProCheck™, A Comprehensive Fabrication Process Mismatch and Reliability Characterization Tool,” White paper, Ridge top Group Inc., 2013.

[16] J. Keane, et al., "On-chip reliability monitors for measuring circuit degradation," Microelectronics Reliability, vol. 50, pp. 1039-1053, 2010.

[17] http://www.ridgetopgroup.com/img/NBTI_layout.jpg

[18] http://www.recoresystems.com/products/xentium-vliw-dsp-ip/

[19] K-S Kim, S. Mitra and P.G. Ryan, "Delay Defect Characteristics and Testing Strategies, " IEEE Design and Test of Computers, pp. 8 – p16, September 2003.

[20] H. Goyal, A. Chatterjee and M. Purtell, "Alternate Test Methodology for High Speed A/D Converter Testing on Low Cost Test, "IEEE 14th Asian Test Symposium (ATS ’05), 1081-7735/05, pp. 1-4, 2005.