Intelligent Sensing and On-Chip Learning for Silicon Lifecyle Management

1st Tim Strobel
Institute of Smart Sensors
University of Stuttgart
Stuttgart, Germany
0009-0001-7876-9915

2nd Sarah Rottacker
Advantest Europe GmbH
Boeblingen, Germany
0009-0004-3079-3061

3rd Roland Rösslhuber
Advantest Europe GmbH
Boeblingen, Germany
0000-0002-4610-2382

4th Jens Anders
Institute of Smart Sensors
University of Stuttgart
Stuttgart, Germany
0000-0002-2498-0352

***Index Terms*—Silicon Lifecycle Management, On-chip Sensing, On-chip Learning, On-chip Inference, PVT Monitoring, Voltage Parameter Optimization, Body Biasing**

I. Introduction

The design of advanced ASICs necessitates the abandonment of pessimistic safety margins. In order to ensure reliability across a range of process corners, temperature areas, and silicon aging states, there is a need for intelligent methods for testing and reliability. This work investigates the potential for optimizing a chip's performance throughout its lifecycle using on-chip sensors and edge computation as part of the Graduate School Intelligent Methods for Test and Reliability [1]. The efficacy of our methods is evaluated through their application to the graduate school's ASICs. Initially, the Time-to-digital converter and the Temperature-to-digital converter were selected as the primary use cases. These two ASICs possess multiple tuning parameters, which can be used to optimize the chip's performance. To this end, optimization techniques are employed to identify a parameter set that ensures reliable performance across different process corners, temperature ranges, and aging states.

II. Use Cases

The graduate school's ASICs are based on Global Foundries' 22-nm fully depleted silicon-on-insulator (FD-SOI) technology. A distinctive feature of FD-SOI technology is its capacity to use body biasing for performance enhancements. To this end, a tuning voltage V_{BB} can be applied to the body of a transistor, thereby affecting its operational characteristics, such as switching speed and leakage. Forward biasing (NMOS: $V_{BB} > V_S$, PMOS: $V_{BB} < V_S$, where V_S is the transistor's source voltage) increases the switching speed but also the leakage, while reverse biasing (NMOS: $V_{BB} < V_S$, PMOS: $V_{BB} > V_S$) lowers switching speed as well as leakage.

A. Time-to-Digital Converter

The Time-to-Digital Converter (TDC) measures the time difference between an event and the next rising clock edge. This measurement principle is also known as time tagging. To this end, the TDC uses delay lines as its basic building block. A delay line consists of a number of delay elements, such as inverters, which delay a given signal by a defined amount of time. By counting the number of delay elements, the signal edge propagates before the arrival of the next clock edge, a coarse estimate of the time tag can be obtained. A more precise measurement of the time tag is then obtained by passing the residue of the coarse conversion to a Vernier TDC, whose operating principle relies on the propagation difference of the signal through two delay lines with nominally different delays. For optimum performance and robustness against process, voltage and temperature (PVT) variations, the coarse delay line, the residue generation and the Vernier delay line have to be adjustable. We can control the performance of the TDC by adjusting the supply voltages and body bias voltages for PMOS and NMOS devices separately. In total, nine supply voltages and six body bias voltages define our parameter space for optimization.

B. Temperature-to-Digital Converter and Aging Sensor

The temperature-to-digital converter measures the temperature using the PTAT (proportional to absolute temperature) measurement principle. In addition to the temperature sensor, the same ASIC contains an aging sensor implemented as a ring oscillator. The temperature and aging sensors allow us to perform on-chip aging and temperature compensation. To compensate for the detected changes and to ensure reliable performance of the ASIC, several tuning parameters are available, including supply and body bias voltages.

III. Methods

In order to enhance the performance of the chip, we employ efficient computational methodologies that can be executed on the edge using a field-programmable gate array (FPGA) development board. The FPGA development board performs the computations and manages communication via the high-speed, low-voltage differential signaling (LVDS) interface with the TDC. Edge computation and high-speed communication reduce the latency of each optimization iteration and ensure data privacy. In a subsequent ASIC version, the hardware for

This work is funded by Advantest as part of the Graduate School Intelligent Methods for Test and Reliability (GS-IMTR) at the University of Stuttgart.

computation could be incorporated into the ASIC based on the developed FPGA prototypes.

A. Hyperdimensional Computing

Hyperdimensional Computing (HDC) [2] is a computational approach inspired by the human brain that utilizes elementary operations, such as the XOR and majority vote, to efficiently learn from limited data [3]. Consequently, HDC holds significant potential for edge and on-chip computations. For the specified optimization task, Reinforcement Learning (RL) is particularly interesting due to its ability to learn optimization policies through interactions. HDC-based Reinforcement Learning (RL) algorithms exhibit efficiency while maintaining high performance [4].

B. Surrogate Modeling

In the context of model-based optimization techniques, surrogate models serve as simplified versions of the underlying system (i.e., the ASIC) to direct the optimization algorithm towards the desired regions within the search space. These surrogate models can be defined probabilistically, which entails the calculation of an uncertainty measure for each data point. By selecting the subsequent parameter combination within the search space in a manner that identifies a region of elevated uncertainty, we can promote the exploration of the search space while concurrently executing the optimization process.

The objective is to gather data from laboratory measurements for the specified use cases outlined in Section II. These datasets will subsequently be employed for surrogate modeling. A strategy for data acquisition entails the selection of new sample points in a manner that maximizes the acquired information. In addition to laboratory measurements, simulation data collected during the ASIC's design phase can be utilized for surrogate modeling. This approach is particularly compelling because it does not necessitate a dedicated data collection step and the surrogate model can already describe regions that are more likely to contain an optimum for the given optimization problem.

IV. Outlook

This work is in its early stage. Future work includes the publication of measurement results as well as results from optimization experiments using edge computation. Of special interest is the comparison of different optimization techniques with respect to the speed, the performance of the found solution, and the computational effort.

References

[1] H. Amrouch, J. Anders, S. Becker, M. Betka, G. Bleher, P. Domanski, N. Elhamawy, T. Ertl, A. Gatzastras, P. Genssler, S. Hasler, M. Heinrich, A. van Hoorn, H. Jafarzadeh, I. Kallfass, F. Klemme, S. Koch, R. Küsters, A. Lalama, R. Latty, Y. Liao, N. Lylina, Z. N. Haghi, D. Pflüger, I. Polian, J. Rivoir, M. Sauer, D. Schwachhofer, S. Templin, C. Volmer, S. Wagner, D. Weiskopf, H.-J. Wunderlich, B. Yang, and M. Zimmermann, "Intelligent methods for test and reliability," in *2022 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2022, pp. 969–974.

[2] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cognitive computation*, vol. 1, pp. 139–159, 2009.

[3] H. Amrouch, M. Imani, X. Jiao, Y. Aloimonos, C. Fermuller, D. Yuan, D. Ma, H. E. Barkam, P. R. Genssler, and P. Sutor, "Brain-inspired hyperdimensional computing for ultra-efficient edge ai," in *2022 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, 2022, pp. 25–34.

[4] Y. Ni, D. Abraham, M. Issa, Y. Kim, P. Mercati, and M. Imani, "Efficient off-policy reinforcement learning via brain-inspired computing," in *Proceedings of the Great Lakes Symposium on VLSI 2023*, 2023, pp. 449–453.