Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

PROCESS DIGITAL TWIN FOR SEMICONDUCTOR TEST EQUIPMENT USING DES AND ML

Zach Eyde[1,2], Robert Dodge[1], and Giulia Pedrielli[1]

[1]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA
[2]Intel Corporation, Chandler, AZ, USA

ABSTRACT

Implementing a digital twin (DT) in semiconductor manufacturing is a rising area of research. In this work, we address key challenges faced to a practical DT implementation in the context of Testing. Specifically, we consider Intel modules (Burn In, Structural Test and System-Level Test) and build an infrastructure for their twinning. According to our infrastructure a DT-agent contains several models whose accuracy and execution complexity are largely different and that can be adopted for providing insights offline as well as at runtime. We introduce the enabler for the Semiconductor Tester DT-agent which we constructed as a high-fidelity Discrete Event Model to support offline optimization and analytics tasks. Such model can be used to train a ML model that can, instead, be used at runtime. The team will investigate methodologies to auto-generate the DES from a base model and auto-repair the DES based on production data.

1 INTRODUCTION

With the huge demands for semiconductors to power automotive, HPC and AI applications, methods for improving facility output and efficiency as needed. The use of a Digital Twins (DT) can lead to this needed increase in productivity and equipment utilization in semiconductor fabrication facilities (Sivasubramanian et al. 2023). While semiconductor wafer fabrication has received important contributions, limited research has been conducted on DTs, particularly inline process DTs in semiconductor Assembly and Testing (A&T). Some key challenges in utilizing DT in the semiconductor A&T are: (i) large resource commitments to generate/sustain models, (ii) computational speed for real-time decision making, (iii) domain experience on process and equipment, (iv) and lack of access to data and/or physical twin for bi-directional data flow implementation. The research team, in collaboration with Intel, is producing a process DT framework to address the key challenges in semiconductor Testing equipment using a modular discrete event simulation coupled with graph machine learning (ML). The modular DES enables lower cost of development for different A&T equipment, while using graph ML methods, policies can be evaluated in real-time to optimize performance objectives. Additionally, methodologies for auto-generation, auto-simplification, and auto-repair of the DES aim to address the overall sustainability while allowing non-simulation and/or non-domain experts to utilize the process DT. For the proof-of-concept demonstration, the team focuses on the Burn-In equipment at Intel, however this approach would be applicable to all A&T equipment (System-Level Testing, Structured Testing, and Assembly)

2 MODULAR SEMICONDUCTOR TEST DES

To create a modular high-fidelity DES for use in semiconductor Testing, the team is leveraging the Python package Simpy (Simpy 2024). Using an object-oriented programming style, modular classes were developed to encompass the various aspects of semiconductor product and Testing equipment. This approach is not constrained to Simpy or any specific simulation software, using any commercial-off-the-shelf (COTS) simulation tool is applicable. Simpy was chosen due to team familiarity and the open-source

licensure of the package. Figure 1 depicts the high-level movements of trays and units through the Burn In equipment at Intel. At each junction of a movement, the current policies used by the test equipment is modeled, however, the team is developing a framework to add in user-defined policies into the simulator. In addition, a graphical-user interface is under development to enhance usability and deliver a DES that can be leveraged at Intel. Validation of the DES is ongoing, initial efforts have shown performance matching on key metrics such as Units-per-hour (UPH) and cycle time for fundamental use cases. Validation will continue to fine-tune the policy logic for more complex use-case such as unit sampling, multi-product operation, and machine failures. Additionally, the DES will be used to perform capacity analysis, design WIFs and other equipment optimizations of the manufacturing equipment.

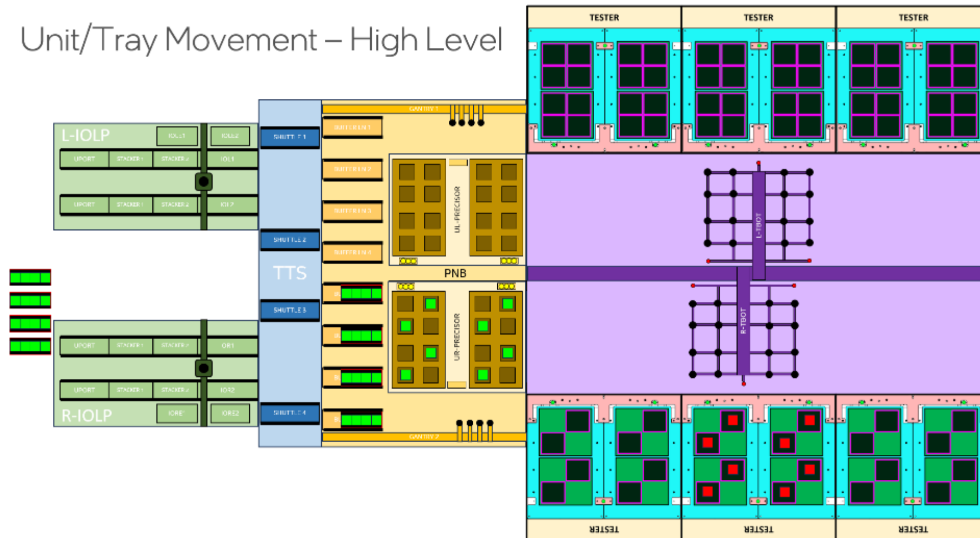

Figure 1: Illustration of the Intel Burn-In Module.

3 CONCLUSIONS AND FUTURE WORK

In our approach, the team will use ML as the backbone of the process DT. This approach is required to meet the decision-making speeds needed for an inline process DT. The DES output will be used to train the ML model to generate predictions and evaluate policies. To create the bi-directional data flows needed for a DT, the ML model is connected to various systems of the Testing equipment and Manufacturing Execution System (MES). Streaming data to the ML Model alone is not sufficient to enable real-time decision-making. The ML model needs to be connected to PLC and Tool Automation software to affect decisions. While this research will start with an ultra-high fidelity DES model of the equipment, the team is looking to prove that manufacturers can start with simpler models and "learn" over time by feeding data to the ML model. If this is validated, it will open process DT up to any engineer and equipment type with extensive simulation expertise, eliminating the barrier for most manufacturing industries.

REFERENCES

"SimPy – Discrete event simulation for Python," [Online] Available: https://simpy.readthedocs.io/en/latest/

Sivasubramanian, C. K., R. Dodge, A. Ramani, D. Bayba, M. Janakiram, E. Butcher, J. Gonzales, & G. Pedrielli. 2023. "DTFab: A Digital Twin based Approach for Optimal Reticle Management in Semiconductor Photolithography". *Journal of Systems Science and Systems Engineering* 32(3):320-351.