AI-Driven Optimization Framework for Next-Generation Semiconductor Manufacturing: From Digital Twins to Self-Optimizing Process

Shivin Srivastava[a], **Jussi Keppo**[b], **Ek Hong Lim**, **Pan Jieming**[b], **Aaron Thean**[b]

[a] *BITS Pilani KK Birla Goa Campus, India*

[b] *IORA, National University of Singapore, Singapore*

1. Abstract

We propose a comprehensive artificial intelligence framework for optimizing semiconductor manufacturing processes, addressing the industry's pressing challenges of increasing complexity, yield variability, and resource efficiency. Our approach combines physics-informed neural networks with reinforcement learning to create accurate digital twins and enable self-optimizing process control across various manufacturing steps. The framework integrates real-time sensor data, physics-based constraints, and adaptive machine learning to optimize process parameters dynamically. We present a detailed implementation plan for Atomic Layer Deposition (ALD) as our initial case study that can be extended to other critical processes including etching, chemical vapor deposition, and lithography, with the potential to reduce overall manufacturing development cycles while significantly improving resource efficiency. This work presents a pathway toward transforming semiconductor manufacturing from experience-based optimization to AI-driven precision control.

2. Introduction

The semiconductor industry faces unprecedented challenges as device complexity increases and feature sizes decrease. Modern chip production requires precise control of thousands of manufacturing steps, with yields varying significantly among manufacturers. Traditional approaches relying on static recipes and manual optimization have created significant bottlenecks in the industry. Development cycles for new processes can stretch to 5-10 years, while yield rates remain suboptimal across multiple process steps. These traditional methods also result in inefficient resource utilization and high environmental impact. Additionally, manufacturers struggle with limited ability to adapt to real-time process variations, and face ongoing challenges in knowledge transfer and process standardization. These limitations collectively hinder the industry's ability to meet growing demand and technological advancement needs.

Our proposed framework addresses these industry-wide challenges through:

- Digital twin creation using physics-informed machine learning.
- Real-time process optimization via reinforcement learning.
- Adaptive control systems for multiple manufacturing processes.
- Knowledge transfer mechanisms between different process steps.
- Resource optimization across the manufacturing pipeline.

3. Proposed Methodology

Our framework consists of three interconnected components designed for broad semiconductor manufacturing optimization:

3.1 Universal Process Modeling through Scientific Machine Learning

We propose a generalized architecture for process modeling:

Physics-Informed Neural Networks (PINNs):

- Modular architecture adaptable to different manufacturing processes.
- Generic loss functions incorporating physical constraints:

$$L_{total} = L_{data} + \Sigma(\lambda_i L_{physics\ constraints_i}) + L_{process\ specific}$$

Extensible differential programming framework for parameter estimation:

$$\frac{\partial S}{\partial t} = F(S, P, t) + C(S, P)$$

where S represents process state variables, P process parameters, and C control inputs.

3.2 *Integration of Scientific Machine Learning (SciML) and Reinforcement Learning (RL)*

Our framework's key innovation is the novel integration of SciML and RL approaches (see Fig. 1). This combination enables both physics-aware prediction and optimal control while respecting physical constraints:

Key Integration Mechanisms:

1. Physics-Informed State Space
   - SciML models provide physically meaningful state representations.
   - Process variables mapped to physically interpretable latent space.
   - State transitions respect physical constraints.

   $s_{t+1} = f_{PINN}(s_t, a_t)$ where f_{PINN} ensures physical consistency

2. Constrained Action Selection
   a. SciML models define feasible action spaces.
   b. Physical constraints incorporated into policy network.
   c. Action validity checked against physics-based rules.

d. $a\_t = clip\left(\pi(s_t), a_{min(s_t)}, a_{max(s_t)}\right)$ where a_{min}, a_{max} are physics constraints

3. Physics-Guided Reward Design
    - Reward includes physical correctness terms.
    - SciML predictions guide reward estimation.
    - Multi-objective optimization balancing physics and performance

$$R = R_{performance} + \lambda^1 R_{physics} + \lambda^2 R_{constraints}$$

4. Hybrid Value Function
    - Value estimates incorporate physical feasibility.
    - SciML predictions enhance future state value estimation.
    - Physics-based regularization of value network

$$V(s) = V_{RL(s)} + \alpha V_{physics(s)}$$

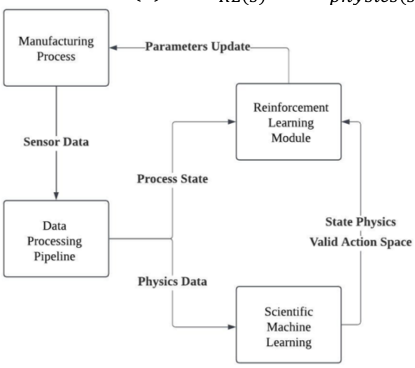

Fig. 1 Workflow

3.3 *Initial Case Study: ALD Implementation Plan*

Our initial validation will focus on ALD process optimization, targeting specific components essential for precise atomic-scale deposition. The framework will integrate existing ALD sensor systems to provide real-time monitoring and feedback of the deposition process. We will implement comprehensive reaction kinetics modeling to understand and predict the complex surface chemistry involved. The system will incorporate precursor timing optimization to ensure optimal layer-by-layer growth, while advanced film quality prediction models will enable real-time assessment and adjustment of the deposition parameters. These components work together to create a holistic approach to ALD process control and optimization. After capturing sufficient data, we also propose to use Topological Data Analysis and Causal analysis tools to uncover hidden patterns and correlations [6].

3. *Planned Implementation*

The initial system will be developed and deployed at the NUS NanoFab. It will incorporate multi-source sensor fusion systems feeding into a real-time processing pipeline, supported by automated feature extraction and cross-process correlation analysis.

4. *Expected Outcomes*

Based on literature review and preliminary simulations, we anticipate:

ALD Process Targets:

- Film thickness prediction accuracy: >90%
- Potential yield improvement: 20-30%
- Expected resource reduction: 25-30%

5. Discussion

Our proposed framework offers several potential benefits:

5.1 Technical Innovations:

- Unified approach to process optimization
- Novel integration of SciML with manufacturing processes
- Scalable architecture for fab-wide implementation
- Cross-process learning and adaptation capabilities

5.2 Expected Industry Impact:

- Projected cost savings of $100M+ annually per fab
- Potential yield improvements of 20-30%
- Development cycle reduction of 40-60%
- Significant reduction in resource consumption

5.3 Research Roadmap:

- Initial validation through ALD implementation
- Planned extension to additional manufacturing processes
- Future integration with supply chain optimization
- Development of industry-wide knowledge bases

6. Conclusion

Our proposed framework represents a significant step toward AI-driven semiconductor manufacturing. The theoretical foundation and initial simulation results suggest broad applicability across the industry. This research aims to demonstrate how AI can transform semiconductor manufacturing, enabling faster development cycles, improved yields, and more

AI-Driven Optimization Framework for Next-Generation Semiconductor Manufacturing: From Digital Twins to Self-Optimizing Process

Shivin Srivastava[a], **Jussi Keppo**[b], **Ek Hong Lim**, **Pan Jieming**[b], **Aaron Thean**[b]

[a] *BITS Pilani KK Birla Goa Campus, India*

[b] *IORA, National University of Singapore, Singapore*

efficient resource utilization.

References

[1] Kanarik, K.J., et al. (2023). "Human-machine collaboration for improving semiconductor process development." Nature, 616, 707-711.
[2] Yanguas-Gil, A., et al. (2023). "Experimental platform and digital twin for AI-driven materials optimization and discovery for microelectronics using atomic layer deposition."
[3] Lee, et al. (2023). "Machine learning driven channel thickness optimization in dual-layer oxide thin-film transistors for advanced electrical performance." Advanced Science, 10(33), 2303589.
[4] Paulson, N., et al. (2023). "Intelligent agents for the optimization of atomic layer deposition." ACS Applied Materials & Interfaces, 15(22), 25789-25801.
[5] Liu, X., et al. (2023). "Machine Learning Applications in Semiconductor Manufacturing: A Comprehensive Review." IEEE Transactions on Semiconductor Manufacturing, 36(2), 145-162.
[6] Giri, Janhavi, and Attila Lengyel. "Explainable Machine Learning Approach to Yield and Quality Improvements Using Deep Topological Data Analytics." International Electronic Packaging Technical Conference and Exhibition. Vol. 87516. American Society of Mechanical Engineers, 2023.

Appendix