Experimental platform and digital twin for AI-driven materials optimization and discovery for microelectronics using atomic layer deposition

Anonymous Author(s)
Affiliation
Address
email

Abstract

Atomic layer deposition (ALD) is a thin film growth technique that is key for both microelectronics and energy applications. Its step-by-step nature and its integration into fully automated clusters with wafer handling systems make is an ideal tool for AI-driven optimization and discovery. In this work we describe an experimental setup and digital twin of an ALD reactor coupled with in-situ characterization techniques that we have developed as a platform for the development and validation of novel algorithms for self-driving labs. Preliminary results show that it is possible to achieve a 100-fold reduction in the time required to optimize new processes. Finally we share some of the lessons learned during the design and validation of our self-driven thin film growth tool.

1 Introduction

Atomic layer deposition (ALD) is a thin film growth technique that is key for applications such as microelectronics and energy technology.(1) A key aspect of ALD is that it allows the growth of materials with high reproducibility and a monolayer-level of control. It is also eminently combinatorial in nature: a growth is composed of a sequence of ALD cycles, each involving a pair of precursors to provide an Angstrom-thick layer of material. This exquisite level of control, together with the fact that it is often integrated into fully automated clusters in semiconductor industry, makes ALD a unique platform for the demonstration of self-driving approaches to materials design.

One of the potential approaches to implement AI-driven materials optimization and discovery is through the use of in-situ characterization techniques. Moving the characterization phase in-situ greatly accelerates the search and optimization processes, using the output of these techniques as a real time feedback to drive materials growth towards the desired goal.

On the other hand, ALD, like many other materials synthesis techniques, brings unique challenges that need to be overcome to develop successful self-driving approaches. First, it involves working with highly reactive chemicals: in addition to safety concerns, moving away from safe operating conditions can cause damage or lead to significant reactor down times. Second, growth is an irreversible process: overshooting optimal conditions often means having to start over. Algorithms developed for self-driving ALD tools must therefore contend with these issues. The development of experimental platforms and digital twins that can be used to develop novel algorithms and demonstrate self-driving capabilities is critical to address these challenges.

In this work, we describe our approach to build an AI-driven atomic layer deposition platform, which includes both an experimental capability as well as a series of digital twins at different levels of fidelity. For the experimental set up, we focus on the design considerations to augment an

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

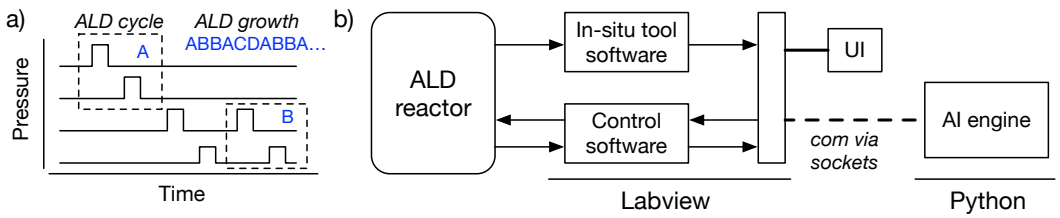

Figure 1: a) Anatomy of an ALD process, comprising a sequence of ALD cycles. During each cycle, substrates are exposed to two different gaseous precursors. b) Scheme of the experimental self-driving ALD tool.

existing experimental reactor with AI-driven capabilities. Finally, we provide some examples of algorithm development and describe our experimental demonstration of real time optimization in our experimental platform.

2 Background

The core idea behind atomic layer deposition is the use of self-limited surface chemistry to sequentially build a material with an almost monolayer level of control.(5) The fundamental unit in an ALD process is an ALD cycle: a combination of two molecules with complementary reactivity that are sequentially inserted into the reactor [Fig 1(a)]. Each precursor reacts with a finite number of reactive sites on the surface. Ideally, once those reactive sites are consumed, the reaction stops: the reaction is therefore self-limited. A second precursor is then inserted in the reactor to bring the surface back to its original state. The net result is an incorporation of a layer of material. ALD is therefore a type of chemical vapor deposition technique, but its self-limited nature makes it highly reproducible and capable to create homogeneous films in 3D and nanostructured substrates.

At the highest level a single ALD growth can be defined as a sequence of cycles $ABBBACDAAAB\ldots$ where each letter in the sequence is chosen from a finite set of processes present in the reactor and leads to the growth of a different material. If we consider a process comprising hundreds of ALD cycles, this provides a extremely large but well defined configuration space. Each cycle adds a certain amount of material. However, since growth is driven by surface reactions, the exact amount depends on the state of the surface. Consequently, the amount of material incorporated during A is different depending on the prior cycles: AA and BA can lead to different amounts of A during the second cycle. This memory can extend a number of cycles into the past.

Traditional approaches to process development usually only explore a few combinations within this large design space. The bulk of the works in the literature involve only one or two materials, and in the case of binary materials usually periodic sequences of the type: $n \times A + B$ are used. However, it has been observed that different distributions of an equal number of A and B cycles can lead to markedly different materials properties.(3)

In addition to this high level search space, each ALD cycle also requires to be optimized: an ALD cycle is characterized by four times (t_1, t_2, t_3, t_4), which correspond to the precursor dose time, precursor purge, co-reactant dose times, and co-reactant purge. For each unknown process there is an optimum set of times that ensures homogeneous growth across the whole reactor (and therefore across a whole wafer), while maximizing throughput.

There are a number of techniques that have been used to monitor this process in-situ: examples include quartz crystal microbalance (QCM), spectroscopic ellipsometry, electrical measurements, infrared techniques, and various synchrotron radiation techniques.(2; 4) The use of in-situ techniques provides a much faster loop during both optimization and search that can complement approaches involving wafer transfer across chambers in a cluster tool.

3 Methodology

3.1 Implementation of a self-driven ALD tool

We implemented our AI-driven platform in an existing ALD reactor comprising up to 10 different precursor channels and capable of being integrated with in-situ QCM, ellipsometry, and electrical measurements. Two other ALD reactors sharing the same underlying control hardware and software allow us to expand the range of in-situ techniques to include FTIR and synchrotron radiation techniques such as X-ray absorption fine structure and various X-ray scattering techniques.

Due to the need to control dose and purge times of different precursors, conventional ALD reactors come with the control hardware required to drive ALD processes. Likewise, many off-the shelf in-situ techniques come with the required read-out capabilities. Consequently, the approach that we have followed here aims at reusing as much of these existing components as possible.

The high level diagram of the experimental setup is shown in Figure 1(b). An experimental front-end component takes care of controlling growth and triggering and receiving the output of the in-situ techniques. This experimental component then communicates with an AI back-end that acts as an oracle, providing a new set of growth conditions. In our case, the existing front-end component is implemented in Labview. As part of the work to build a self-driving tool we modified it so that it could establish connections and communicate with the AI back-end. The back-end oracle was implemented as an internal local server that communicates with Labview via sockets. It was implemented in Python. Having the tool as a client of the AI back-end lets us reuse this component in different reactors and with digital twins. This flexibility comes at the cost of having to adapt AI algorithms to a situation where they have to wait to receive new inputs rather than requesting them on demand.

3.2 Digital twins for ALD reactors

One of the challenges of ALD is that the exact behavior of the growth process depends on a series of experimental parameters that characterize each ALD precursor and its interaction with the surface, such as the reaction probability or the precursor pressure inside the reactor. These parameters are not known for the vast majority of processes. Consequently, algorithms have to be able to perform under a broad range of conditions. This motivates the development of digital twins to serve as a platform for the evaluation and development of robust algorithms.

Accurate modeling of an ALD process requires computational fluid dynamics simulations of the reactive transport of the different precursors. However, when in-situ techniques are focused on a single point of the reactor, it is possible to replace these computationally intensive models with a surrogate model that is able to capture the key nuances of an ALD process at a much lower computational expense. Here we have leveraged the existence of both types of models in the literature, including open source codes for 3D simulations (i.e. see Yanguas-Gil and Elam(6) and references wherein), to create an algorithm development testbed. These tools implement the key inputs defining an ALD process, and provide outputs corresponding to various in-situ techniques, including the presence of experimental noise. In particular, we have implemented models for in-situ QCM, mass spectrometry, and spectroscopic ellipsometry.

4 Results

As a first step we have focused on demonstrating AI-driven optimization of a new or unknown ALD process using the capabilities we have developed. As mentioned in Section 2, each ALD cycle can be characterized by four set of times (t_1, t_2, t_3, t_4), corresponding to the dose and purge times for the two precursors involved in the growth of a specific material. These values are critical to ensure properties such as wafer scale homogeneity while maximizing throughput. Also, too short purge times can lead to uncontrolled growth in the reactor and, potentially, a catastrophic failure requiring extended downtime. We therefore chose this problem as an exemplar problem to demonstrate the self-driving capabilities of our tool. For this study, we used QCM as the in-situ tool to provide feedback for the optimization process.

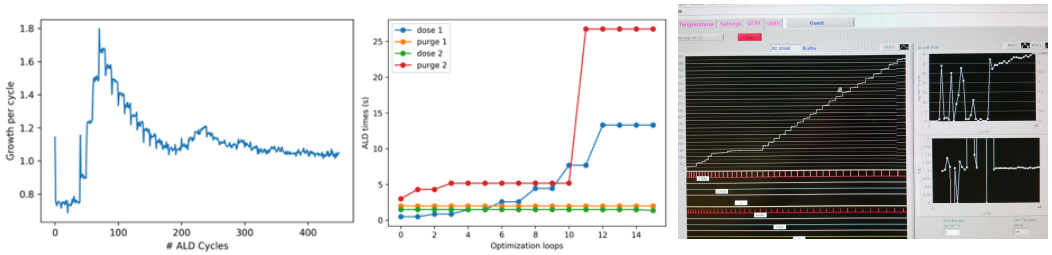

Figure 2: Left: evolution of film thickness during an optimization cycle using the digital twin. Center: characteristic times of the ALD cycle. Right: Snapshot of self-driving ALD reactor implementing the expert system optimization developed in this work.

4.1 Development of algorithms for self-driving ALD tools

First, we explored and fine-tuned two types of optimization algorithms using a digital twin of an ALD reactor. In this case, since the in-situ technique gets information from a single point of the reactor, we used a digital twin based on faster surrogate models. We focused on a black box Bayesian optimization approach and and an expert system that leverages the underlying physics to streamline the search of optimal conditions. An example of an optimization cycle is shown in Figure 2.

4.2 Experimental demonstration of self-driving ALD capabilities

Once we fine-tuned the optimization algorithm in the surrogate model we ported it to our experimental set up to validate its use using a real process. For this initial set of experiments, we focused on the optimization of ALD growth in the reactor using trimethylaluminum and water as precursors. This process is the "MNIST" of ALD and one of the best characterized processes in the literature. Still, trimethylaluminum is a pyrophoric liquid, so this first series of experiments were carried out under direct human supervision. For this specific example, we used the rule-based model.

In Figure 2 we show an snapshot of the user interface while carrying out the optimization process. The results showed that the optimization algorithm transferred successfully to the experimental reactor, demonstrating the self-driving capabilities. Compared to the time that it would take for a human operator, we estimate a x100 acceleration through the AI-driven approach.

5 Discussion and conclusions

We extracted the following lessons from the process of developing our self-driving AI tool:

The client-server model between the AI back-end and the experimental equipment was the least disruptive approach to integrate the self-driving AI component into an existing tool. Working on a piece of equipment shared with human users was one of the key constraints in our work. We essentially replaced a human user with an AI user, outsourcing all the decision making to the AI server. We believe that this could be a common use case scenario of self-driving AI tools in the future, where the self-driving AI component is used as an auto-tune capability along with human users.

From a safety standpoint, leaving the reactor in charge was essential to ensure safe operating conditions. Having the reactor request new operating conditions on demand requires additional work to adapt existing machine learning algorithms. However, it ensures that go/no-go decisions are handled by the growth tools and it leads to well defined behaviors if the communication breaks down.

Finally, while the use of in-situ techniques can really accelerate the optimization and search process, it also brings up interesting challenges from an algorithm perspective: the cost of doing an additional in-situ run is much lower than the cost of having to transfer a wafer back or deciding to start with a clean material. In some circumstances, these costs are clear and objectively imposed by experimental constraints, but in other cases these costs have to be inferred by the agent (i.e what if I overshoot and have to start all over again?). Digital twins incorporating some of these objective costs and undesirable outcomes could be extremely valuable for the development of novel algorithms capable of handling this type of scenarios.

References

[1] Elsa Alvaro and Angel Yanguas-Gil. Characterizing the field of atomic layer deposition: Authors, topics, and collaborations. *PLOS ONE*, 13(1):1–19, 01 2018.

[2] E Langereis, S B S Heil, H C M Knoops, W Keuning, M C M van de Sanden, and W M M Kessels. In situ spectroscopic ellipsometry as a versatile tool for studying atomic layer deposition. *Journal of Physics D: Applied Physics*, 42(7):073001, mar 2009.

[3] Anil U. Mane and Jeffrey W. Elam. Atomic layer deposition of w:al2o3 nanocomposite films with tunable resistivity. *Chemical Vapor Deposition*, 19(4-6):186–193, 2013.

[4] R. Methaapanon, S. M. Geyer, C. Hagglund, P. A. Pianetta, and S. F. Bent. Portable atomic layer deposition reactor for in situ synchrotron photoemission studies. *Review of Scientific Instruments*, 84(1):015104, 2013.

[5] Riikka L. Puurunen. Surface chemistry of atomic layer deposition: A case study for the trimethylaluminum/water process. *Journal of Applied Physics*, 97(12):121301, 2005.

[6] Angel Yanguas-Gil, Joseph A. Libera, and Jeffrey W. Elam. Reactor scale simulations of ald and ale: Ideal and non-ideal self-limited processes in a cylindrical and a 300 mm wafer cross-flow reactor. *Journal of Vacuum Science & Technology A*, 39(6):062404, 2021.