Development of new strategies for Area Selective Deposition Processes

---

Rachele Pia Russo

June 15, 2023

International Business Machines Corporation
Politecnico di Torino
Universitè Paris Citè

Development of new strategies for Area Selective Deposition Processes

Rachele Pia Russo

1. Reviewer Maria Luisa Della Rocca
UFR physique
Université Paris Cité

2. Reviewer Carlo Ricciardi
Department of Applied Science and Technology
Politecnico di Torino

Supervisors Krystelle Lionti
Research Staff Member
International Business Machines Corporation (IBM)

June 15, 2023

Rachele Pia Russo
Development of new strategies for Area Selective Deposition Processes
Selective Area Deposition, June 15, 2023
Reviewers: Maria Luisa Della Rocca and Carlo Ricciardi
Supervisors: Krystelle Lionti

International Business Machines Corporation

Almaden Research Lab
650 Harry Road, 95120
San Jose, California (USA)

Contents

Part I

Introduction

1 Introduction

The continuous scaling of the electronic technologies and the development of nonplanar device architectures requires to pattern materials more precisely for the creation of specific features. This is becoming extremely challenging, causing the development of new nanoscale fabrication techniques. More and more precise control over atomic scale processes is required.
Nowadays, lithography is still the most widely used process in device fabrication. However, patterning, overlay errors and non conformal deposition on critical features can heavily affect the device fabrication. Self-alignment processes are required to solve this kind of issues, so the demand for the development of bottom-up approaches, to deposit selectively a film on top of another surface, is increasing.

One of the bottom-up approaches that attracts more the attention of the semiconductor community is Atomic Layer Deposition (ALD). It is a thin film deposition technique that involves the sequential exposure of a substrate to two half reactions. Once the first half-reactor is introduced in the ALD chamber, it reacts with the substrate, resulting in the formation of a monolayer of the desired material.
The reaction is self-limiting, meaning it stops once the surface is fully covered with the desired material. After the first half reaction, the reaction chamber is purged with an inert gas, such as nitrogen or argon, ensuring a clean environment for the subsequent reaction.

The substrate is then exposed to the second half reaction. This involves the introduction of a second precursor gas in the chamber, labeled as a co-reactant. It reacts with the previously deposited monolayer, resulting in the growth of an additional atomic layer. Again, the reaction is self-limiting. The chamber is purged once again to remove any excess reactants, reaction by-products or residues.

The process then is repeated by alternating between the first and second half reactions. Each cycle adds one atomic layer to the thin film.
By repeating these steps, layer by layer, the ALD process allows for precise control of the film thickness, through the number of cycles, and of the film composition, through an accurate choice of reactants and co-reactants, while ensuring excellent conformality and uniformity.

This layer-by-layer growth can be engineered to deposit a vapor phase material just on a predetermined surface in a selective way. This is known as Area Selective Atomic Layer Deposition (AS-ALD) or simply Area Selective Deposition (ASD).

To make the process selective a blocking agent is required.
The choice of the inhibitor depends on the specific materials and surface chemistry involved [1]: it has to selectively bind or react with certain functional groups on pre-selected surface sites, preventing subsequent reactions with the precursor gas, while leaving the other regions untouched.
Moreover, it should form strong covalent bonds, to remain firmly bounded to the substrate during subsequent ALD or solvent cleaning steps, and it should not decompose or react with the precursor gases.

A variety of inhibitors can be used: self-assembled monolayers (SAMs), silane-based compounds, with alkyl or halide functional groups, polymeric films, metal oxide thin films or functionalized nanoparticles with appropriate organic ligands or surface modifiers. Between them, organic inhibiting materials are the most attractive and versatile deactivation chemistry, showing the highest level of selectivity.

In particular, organic self-assembled monolayers (SAMs) [9] [5], are the most widely exploited. They are made by a single layer that forms through the assembly of organic molecules on a solid surface in a regular and ordered manner. They constitute the thinnest possible blocking layer.
Also polymers are widely studied [7] [4], offering a wide range of options in terms of materials and characteristics. The polymeric structure can be tailored depending on the specific ALD process and substrate requirements, to ensure good adhesion and compatibility between the inhibiting layer and the substrate.

The main ASD limitation is related to adhesion issues. They can be caused by poor compatibility between the substrate material and the inhibitor, so that they are unable to form a strong and durable bond, or by residual stresses and impurities at their interface. ASD can be also impacted by insufficient selectivity, that leads to defects in the inhibitor layer, causing the ALD material growth on the deactivated area.

Providing a complete understanding of inhibitor structure/properties relationships is one of the main goal to make advancements in the ASD research, enabling a wide-spread understanding of the selectivity mechanism. This can be achieved with a purely experimental work or with the support of 'Accelerated discovery' tools, that help to speed up the way to obtain information about the deposition process.

2

Research lab and team presentation

IBM Research in Almaden is IBM's Silicon Valley innovation lab. Scientists ranging from chemists to physicists, as well as mathematicians, engineers and designers at Almaden are pioneering breakthroughs across disruptive technologies including AI and machine learning, hybrid cloud, quantum computing, security and storage. Their published works and contributions to the scientific community cross several industries, including healthcare, semiconductors, sustainable energy, renewable materials, retail and data privacy. Accomplishments are recognized through thousands of publications, citations, invited talks, industry and academic partnerships and conference chair appointments.

My group deals with semiconductor fabrication and packaging.
The main focus is to develop new materials for the smallest length scale (<20 nm) of semiconductor fabrication.
Bottom-up additive strategies, in particular Atomic Layer Deposition, are becoming increasingly more important. The ALD inhibition in a controllable area selective manner, which provides a critical means of depositing a film without subtractive processes, is the main subject of interest. The research is centered on the identification and synthesis of selective surface-binding inhibitors to enable area selective processes and provide a more complete understanding of inhibitor structure/properties relationships.

ASD has several applications in microelectronics and nanoelectronics, including transistor and integrated circuit (IC) fabrication, storage systems and semiconductor devices. It enables the creation of customized contact layers, electrodes or interconnections, tailored to the specific needs of the device.

Specifically, the group is focused on the exploitation of ASD in BEOL (Back-End-of-Line) steps, the final stages of the integrated circuit manufacturing process. They involve the fabrication of interconnects and metal layers that connect the various components of the IC. In BEOL, ASD can be used to perform selective deposition on patterned surfaces, with both planar or high aspect ratio features. The goal is to bind the inhibitor to the metallic area (Cu in our case), enabling successive depositions just on the dielectric subsections (SiCOH in our case).

3 Internship subject overview

The experimental section of my project (described in part II) focuses on the specific class of organic compounds of amines, characterized by a nitrogen atom bounded to one or more alkyl or aromatic groups carbon atoms. The amines are classified as primary, secondary or tertiary if the nitrogen atom is bounded to one alkyl (or aromatic group) and two hydrogen atoms, or to two alkyl (or aromatic groups) and one hydrogen atom or to three alkyl (or aromatic groups), respectively. In order to maximize the bonding between the amine and the substrate, we are focusing exclusively on primary amines.

The aim is to study their inhibiting properties and understand the possible advantages of using this specific class of organic compounds as inhibitors with respect to others. The amines have been selected through a detailed search among a very large number of compounds, different in terms of chemical composition as well as chemical and physical properties, taking advantage of searching and filtering tools that expedite the process. Facilitate the work is precisely what 'Accelerated discovery' aims.

Possible means to increase the efficiency and effectiveness of the research processes can be: increasing the research and development effort by investing more resources, enhancing the collaboration among researches from different disciplines or institutions and adopting innovative research methodologies to test and validate hypothesis, such as high-performance computing, big data analytics and artificial intelligence tools.

In particular, "Accelerated discovery for ASD" wants to define a process window for achieving more spatial control in the ALD process on patterned surfaces. This reduces the number of experimental trials, optimizing the workflow, and favours advancements in the classes of materials that can be used.

Different classes of materials can already be exploited as inhibitors, resulting in a broaden repertoire of ALD materials that can be selectively deposited. Nevertheless, some issues can impact the effectiveness and the reproducibility of the process, such as adhesion issues or insufficient selectivity. Moreover, the reaction geometry, the precursor flow conditions and all the process parameters, such as fluctuations in

temperature, pressure and deposition rate, need to be optimized to guarantee a successful and reproducible experiment.

My personal work on Accelerated Discovery for ASD, described in part III, consists in creating an artificial intelligence (A.I.) model that can predict new sets of experimental conditions and new classes of inhibitors to enable selective area deposition.

Machine learning algorithms can identify relationships between process parameters and properties of the deposited material, allowing for the prediction of the behavior of precursor gases, their diffusion and their distribution on the substrate surface [10]. This can compensate any experimental deviation, aiding in identifying potential experimental solutions.

The A.I. model input data are gathered from Scanning Electron Microscopy (SEM) images. They were taken after the deposition of both the inhibitor layer and the ALD material. A combination of looking at them, and looking up their corresponding experimental conditions written down in notebooks, allows to gather all the experimental parameters that can influence the ASD process and to visualize the effective growth rate of the ALD material.
From all the information collected from the SEM images, the frequency of success for each set of unique experimental conditions can be calculated and used as input data for the A.I. model.

The A.I. models chosen to process the data are based on classification and regression algorithms. They can predict, for a new set of the experimental conditions, its frequency of success and what is the range of values for a certain experimental parameter that enables selective deposition, so a successful result.
Moreover, by linking the inhibitors already used to their chemical composition and their chemical or/and physical properties, is possible also to predict new classes of inhibitors that give rise to selective deposition.

Part II

Experimental section

4 Amines as functionalizing agents of metallic surfaces

4.1 Amines searching process

The project focuses on studying the inhibiting properties of the specific organic compounds class of primary amines. The choice is based on previous experimental results obtained using Propargylamine as inhibitor [3].
Propargylamine was shown to be an effective inhibitor toward TaN, enabling TaN deposition just on SiCOH (dielectric) and not on Cu (metal), up to almost 100 TaN ALD cycles. The effective binding of Propargylamine on the metallic surface was driven by the primary amine, used as ligand.

Previously, selective deposition of TaN was demonstrated by exploiting as inhibitor other classes of organic compounds, such as phosphonic acids.
However, phosphonic acids were deposited on the substrate from liquid phase.

The aim is to develop an "all integrated process", where both the inhibitor and the ALD material are deposited through ALD (i.e. vapor phase), without exposure of the wafer to air after the inhibitor deposition.
This requires a very low vapor pressure compounds: the physical properties of amines make them attractive candidates for this purpose.

Based on the previous requirements and results, a research for a possible new inhibitors was conducted through four big classes of organic compounds: phosphonic acids, hydroxamic acids, primary amines and carboxylic acids functional groups. Only the organic compounds with maximum 10 Carbon backbone were included, excluding all the ones containing F and Cl, being not compatible with microelectronic applications. Furthermore, a vapor pressure of at least of 1mTorr was required, to ensure they could vaporize in the ALD tool.

A first accurate search through a huge variety of organic compounds, applying the searching criteria listed above, was conducted through IBM CIRCA.
CIRCA is an IBM Research platform designed for a rapid and easy search of materials, with straightforward browsing and filtering options to quickly identify them.

Advanced keyword search, including proximity search, chemical similarity, substructure search, reaction search or target assay searches, can be exploited. Powerful filtering options refine the search to give a targeted result set.
CIRCA covers a huge amount of data from patents, abstracts, text articles and public databases. The tool is perfect for working with big data, and linking it with other cognitive applications enables advanced AI capabilities.

It is worth noting that the search was not perfect and a second filtering job was performed through DataWarrior, to further narrow down the list of compounds. DataWarrior is a software tool designed for chemical data visualization and analysis. It enables to import external large chemical databases and offers a wide range of customizable visualization options, from 2D and 3D chemical structure representations to scatter plots, bar charts and heatmaps.

The second filtered search effectively reduced the number of organic compounds to be successively tested as inhibitors for ASD. One of them, belonging to the class of primary amines, was X.
It remains an inhibitor of interest for future IBM researches.

4.2 Metallic surfaces functionalzation

The first experiment was performed to test the possible functionalization of metallic surfaces through X. Blanket Cu coupons were used as substrates. As a start, both liquid and all-integrated vapor phase approaches were investigated.

- From liquid phase: The substrates were functionalized with X as-is or with a solution at a concentration of 0.1 wt% in 4-methyl-2-pentanol (4M2P). The substrates, without any surface preparation, were immersed in the liquid solution for a period of 30 minutes. After removing the samples from the solution, they were rinsed with 4M2P and isopropanol and dried under nitrogen (N).

- From vapor phase: X was inserted into a stainless steel bubbler and after plumbed into an ALD tool. First, a 10 min N remote plasma treatment was executed on the samples, followed by evacuation. Then, X was pulsed for 3.5 s, pursued by a dwell time of 80 s and by 8 s of evacuation. The ALD chamber was maintained at the constant temperature of 35°C. The same procedure was repeated 80 times.

4.3 X-rays photoelectron spectroscopy (XPS)

The functionalized surfaces were then characterized by X-rays photoelectron spectroscopy (XPS) to confirm that the inhibitor was effectively bonded to the metal surface. XPS is a powerful technique that can provide valuable information aiding in characterizing the quality, purity and uniformity of deposited material and in optimizing the ALD deposition parameters.

It is possible to determine the elemental composition of the deposited film by measuring the binding energies of the core electrons of the different elements present. Looking at the binding energy shifts allows us to distinguish different chemical species. By comparing the spectra before and after ALD, any changes in the surface chemistry can be identified, indicating the also the presence of possible contamination and impurities. XPS can also provide an estimate of the film thickness by comparing the X-ray signal intensity from the film to that of the substrate.

The XPS analysis was performed by comparing four different metallic substrates. The results are showed in the Table 4.1:

- Cu reference substrate not functionalized
- Cu substrate functionalized with X from liquid phase
- Cu substrate functionalized with a 0.1 wt% solution in 4M2P
- Cu substrate functionalized with X from vapor phase

Consistent data with the formation of an X monolayer are: a decrease in Cu at% and an increase in C at% (up to 35%-45%) and in N at%, compared to the Cu reference substrate. Looking at the chemical composition of the inhibitor X, a larger increase in the C concentration with respect to the N one is expected.

	C1s	N1s	O1s	Si2p	Cu3p
Cu ref	18.6	1.2	31.7	0.2	48.4
(1) Cu + liquid X	17.5	0.2	37.3	42.6	2.3
(2) Cu + X in 4M2P	29	1.4	23.2	0	46.5
(3) Cu + vapor X	78.9	15.6	1.2	0.9	4.1

Tab. 4.1: Liquid phase vs Vapor phase inhibitor

Looking at the Table 4.1, the substrates (1) and (2), functionalized with X in liquid phase, show an amount of C lower than expected: X is not attached to the substrate. Moreover, for (1), the Cu concentration is extremely low, while the Si one, supposed to be not detectable, is extremely high. This indicates a possible corrosion of the Cu

substrate, showing the underlying Si layer.

The substrate (3), functionalized with X from vapor phase, exhibits less amount of Cu compared to the reference and a C relative atomic concentration of 78%, about twice that of what is observed for a monolayer. So, the deposited layer can be thicker than expected. Sample (3) will be used in the following, referred to as Cu + X (1).

5 X from vapor phase as inhibitor for ZnO ALD

To assess the inhibiting properties of X, it was again pulsed into the ALD tool. We used the same ALD experimental parameters of the first attempt, that will be used in all the subsequent ALD runs.

We performed first the X substrates coating and after, without breaking the chamber vacuum, a Zinc Oxide (ZnO) atomic layer deposition. The chamber temperature was changed from 35° C to 150° C. A half cycle of dimethyl Zn, pulsed for 20 s was followed by a water flow pulsed for 0.03 s, both with a dwell time of 20 s.

We are searching for an inhibitor that binds just to a metallic surface and not to a dielectric one, to enable selective deposition of ZnO through ALD only on the dielectric subsection of a patterned substrate.

As a consequence, in the ALD chamber were inserted a blanket Cu coupon (metallic surface), a blanket carbon-rich silicon oxicarbide (SiCOH) coupon (dielectric surface) and a coplanar patterned Cu/SiCOH substrate, labeled as P. Different experiments were performed, by keeping constant all the deposition parameters, except for the number of ZnO ALD cycles, and by using each time new substrates (i.e. replaced, not reused).
The experiments are listed below, chronologically ordered:

- 50 ZnO cycles
- 200 ZnO cycles
- 114 ZnO cycles
- 150 ZnO cycles

Blanket coupons were characterized through XPS.
Looking at the Cu substrates results in Table 5.1, the C relative atomic concentration is always between 35% and 45%, consistent with the X monolayer formation.
The X layer thickness is confirmed by the 1:1 ratio between C and N percentages, expected to be around 6:1. The XPS tip might have been reached the N layer, underlying the X one, formed during the N plasma treatment done before the X pumping.

The Zn concentration remains 1% - 2% up to 150 ZnO ALD cycles, while it increases to 40% for 200 ZnO ALD cycles. X can block the ZnO growth on Cu only for a limited number of ZnO cycles.

	C1s	N1s	O1s	Si2p	Cu3p	Zn3p
Cu ref	18.6	1.2	31.7	0.2	48.4	0
Cu + X (1)	78.9	15.6	1.2	0.9	4.1	0
Cu + X + 50 ZnO	44.8	44.9	3.4	0.2	5.3	1.5
Cu + X + 114 ZnO	35.9	40.7	7.1	0.1	14.3	1.9
Cu + X + 150 ZnO	37.8	37.1	9.3	0	14.4	1.3
Cu + X + 200 ZnO	11.5	0.5	47.3	0	0	40.6

Tab. 5.1: X and ZnO ALD on Cu substrates

Looking now at the XPS results on SiCOH substrates, the reference substrate and the one subjected just to 80 X ALD cycles, show the same C concentration. The high N concentration is caused by the N plasma treatment preformed before X pumping.
The SiCOH substrates subjected also to ZnO ALD, show a gradual increase in the Zn relative atomic concentration, to which corresponds a gradual decrease in the C one. These data describe exactly what we want: X does not attach on SiCOH, so that ZnO growth can occur.
The conclusion is that X adheres to Cu substrates, deactivating them from ZnO deposition, leading to a blanket deposition only after 150 ZnO ALD cycles. Instead, it does not attach on SiCOH substrates, enabling ZnO depositon even for a low number of ALD cycles performed.

	C1s	N1s	O1s	Si2p	Cu3p	Zn3p
SiCOH ref	19.4	0.1	52.5	28	0	0
SiCOH + X	20	11	46.2	21.5	0	1.4
SiCOH + X + 50 ZnO	44.7	39.8	8.7	4.9	0	2
SiCOH + X + 114 ZnO	34.9	29.2	21	9.3	0	5.4
SiCOH + X + 150 ZnO	13.7	3.9	45.6	0.3	0	36.6
SiCOH + X + 200 ZnO	10	0.2	47.8	0	0	42

Tab. 5.2: X and ZnO ALD on SiCOH substrates

A good selectivity window, plotted in Figure 5.1, can be defined. It can be confirmed by looking at the SEM images on patterned coupons. SEM images reveal a good agreement between the results obtained for blanket and patterned coupons, in Figure 5.2 (a),(b) and (c). However, for 150 ZnO ALD cycles, ZnO selective growth is esxpected while ZnO is deposited everywhere (Figure 5.2 (d)).

This assess that the results obtained on blanket coupons could not be necessarily transferred on patterned coupons.

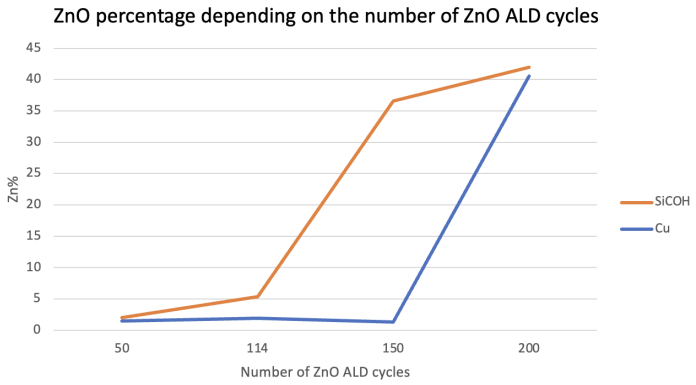

Fig. 5.1: ZnO selectivity window

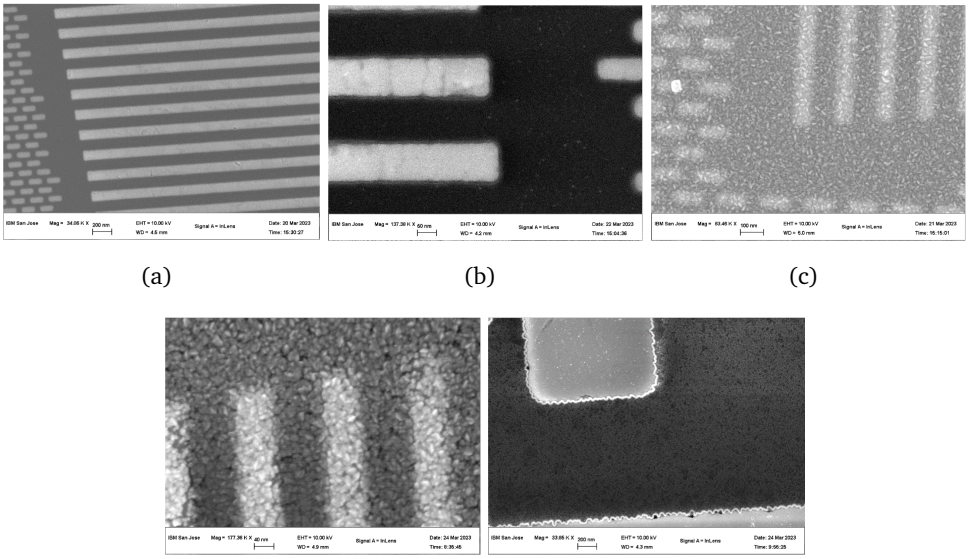

Fig. 5.2: (a) 50 ZnO ALD cycles, (b) 114 ZnO ALD cycles, (c) 200 ZnO ALD cycles, (d) 150 ZnO ALD cycles on P, (e) 150 ZnO ALD cycles on K

To ensure the reproducibility of the previous results obtained for X from vapor phase, a further investigation about X as functionalizing agent on metallic surfaces was performed.

Each Cu substrate, functionalized with X, was then characterized by XPS. Almost all of them show relative atomic concentrations consistent with the formation of a X monolayer. Follow-up investigations are focused on ensuring reproducilibity to this process, involving also quantum mechanical ab initio simulations [8].

Part III

Accelerated Discovery for Area Selective Deposition

6 Accelerated Discovery for Area Selective Deposition

'Accelerated discovery' involves finding ways to expedite the process of uncovering new information to make scientific advancements. This can be achieved through various means including cutting-edge technologies, such as quantum mechanical simulations [6] and artificial intelligence [2], that can enhance the efficiency and effectiveness of the research processes, allowing for faster analysis, data processing and hypothesis testing.

Until now, in order to test the inhibiting properties of a certain material, that can enable selective deposition of a vapor phase precursor on a predetermined surface, purely experimental 'trial and error' approaches have been exploited.

The aim of this project is to leverage an AI model that can recognize and suggest a material, that has not been tested experimentally before, as successful inhibitor or not. By specifying a set of experimental conditions, the model can predict if it will give rise to selective deposition and it can predict range of values for a certain experimental parameter that lead to a successful result.

All the information, to built the input data for the A.I. model, are obtained by looking at SEM images, showing a substrate, which may or may not undergo a pre-treatment, on which a particular inhibitor is deposited and a subsequently ALD deposition is performed.

The input data are split in training and testing datasets, essential components of the machine learning process. They are collection of data points, each of them consisting of input features and a target output. In this case a single data point has as input features a unique set of experimental conditions used in a previous ASD experiment and the target output are the results obtained, labeled as successful or not successful, and the frequency at which successful ones are achieved.

The AI model learns from the testing dataset, by iteratively presenting the data to the model. Its performance is evaluated and its internal parameters are adjusted through optimization algorithms, to minimize the difference between its predictions and the true labels.

Once optimized, the model makes predictions on the testing dataset based on the patterns and relationships it learned during its training. The model's predictions are then compared with the true testing set labels to assess its precision.

All the main steps, that characterize my work, are resumed in Figure 5.1:

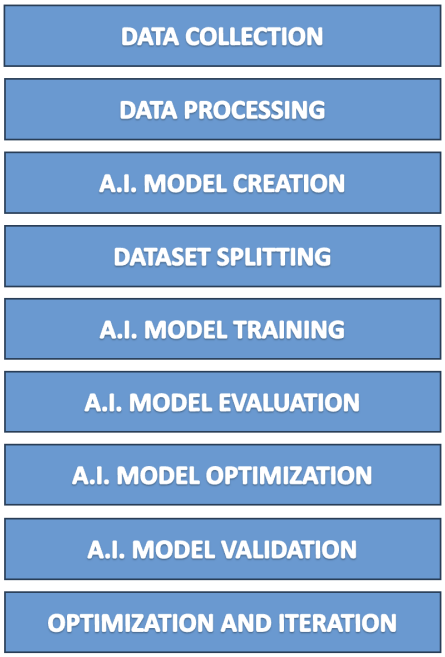

Fig. 6.1: Project overview

7 Data collection and Data processing

7.1 SEM images characterization

Scanning Electron Microscopy (SEM) analysis provides high-resolution images of the surface of the deposited thin film, allowing for visual and quantitative evaluation of its morphological and structural characteristics.

Top-down SEM images contain information about roughness, porosity and conformality of the depostied film.
The presence of layering and surface defects, such as grains, aggregates or holes, give us information about the structure of the deposited film, enabling a measure of their size, shape and distribution.
Additionally, SEM analysis allows the examination of the interface between the substrate and the deposited thin film, identifying (if large enough) any contaminants that may affect the material's properties.

Gathering a large and diverse set of SEM images, that covers the majority of the ASD experiments performed in this lab in the last few years, constitute the starting point to build a consistent dataset to train and test the AI model[**empty citation**]. We need to prepare the images by pre-processing them, ensuring that they are annotated with the desired inputs and outputs labels.
Figure 6.1 is a good example of selective ZnO deposition. ZnO grows in a polycrystalline fashion, in the way that different grains orientations can be recognized where the film is deposited.

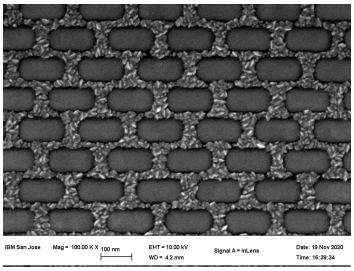

Fig. 7.1: Example of SEM image

7.2 Creation of the dataframe

The categorization is performed by considering all the experimental procedures and the correspondent experimental parameters that can differentiate one image from another.

Each image has a specific name, but is the timestamp that is used to distinguish uniquely them. It consists in the day and the exact time (00:00:00) in which the image was taken through the SEM tool.

First, we specify the type of substrate: three blanket coupons (SiCOH, Cu and Co capped Cu coupons) and two patterned coupons (PORTER and KATLA). We specify if the substrate is subjected to any substrate preparation or to a plasma treatment, by characterizing its duration and the type of plasma.

The inhibitor molecule is one of the decisive parameters that determines the results of the experiments. 14 different inhibitors were identified.
The inhibitor can be deposited on the substrate from liquid or vapor phase.
For liquid phase inhibitors we distinguish between inhibitors deposited as-is or in solution, the type of solvent and the solution concentration (wt%). We take note also of the the immersion time. For vapor phase inhibitors, the specified parameters are the temperature of the ALD chamber, the number of ALD cycles, the pulse duration and the dwell time.

When we are dealing with organic inhibitors that contains a polymerizable thermal or photoreactive component, after their deposition some procedures can be exploited to 'cross-link' the functionalized surface, so that the ADL material nucleation is further delayed or blocked. Even this procedures and their details are classified.

The material deposited through ALD and all the ALD procedure parameters are labeled: chemical precursor and co-reactant, temperature of the ALD chamber, pulse duration and the dwell time and the number of ALD cycles for both the 'half-reactions'.

The images are differentiated also by specifying the dimensions of the visible features.

After listing all the "Experimental conditions", the images are classified with respect to the growth rate of the ALD material, that determines the success of the experiment. Patterned coupons are labeled by specifying the ALD material growth rate on both the dielectric and metallic areas.

The information collected are grouped in a two-dimensional tabular data structure, called dataframe, in which the number of rows corresponds to the number of SEM images analyzed, while each column is a specific feature that characterizes them. The table contains 1343 rows and 36 columns and it constitutes the starting point for the data processing.

7.3 Data processing with Python

Each column of the dataframe is converted as a "category" data type. The code counts the unique labels (categories) present in each column and assigns a numerical value (index) to each category, as shown in Figure 6.2.

SUBSTRATE	Substrate
SiCOH	0
Co capped Cucoupon	1
Cu	2
KATLA	3
PORTER	4
	-1

Fig. 7.2: Example of data indexing

The columns of the dataframe are divided in two sections: 'Experimental Conditions' and 'Results'. In particular, the second group describes the growth rate of the ALD material on the dielectric and the metal that make up both the patterned and the blanket coupons: 'Results on SiCOH', 'Results on Cu' and 'Results on Co on Cu'.

Dealing with blanket coupons means that just one out of the three result is present, while for patterned coupons the result is specified for both the metallic (Cu) and dielectric (SiCOH) substrate subsections. I ended up with 2 possible outcomes for 'Co capped Cu coupons' and 14 possible outcomes for both 'SiCOH' and 'Cu'.

After indexing the data, GroupBy objects are created, representing the grouping of data based on specific columns. The first columns specified are all the ones belonging to the 'Experimental Conditions' section, so that we can group the data with respect them and have each row that represents a unique set of experimental conditions used. In the same time we can count how many times each set of conditions is repeated.
Another GroupBy object is created, by specifying the 'Results' section columns. In this way, for each set of experimental conditions, it is possible to list all the results combinations and count how many times each combination occurs.
This can be understood better by looking at Figure 6.3.

index	Substrate	Substrate_preparation	Plasma_treatment	...	water_dweel	area	Res_SiCOH	Res_Cu
0	-1	-1	-1	...	-1	1	-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1...	-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1...
1	0	2	0	...	0	1	6, 6, 6	-1, -1, -1
2	0	3	0	...	0	1	7, 7, 12, 12, 12, 12, 7	-1, -1, -1, -1, -1, -1, -1
3	1	3	0	...	0	1	-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1...	-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1...
4	1	3	0	...	0	1	-1	-1
...	...	...	...	...	...	...	...	...
193	4	3	1	...	0	2	1	1
194	4	3	1	...	0	1	8, 7, 4, 4	7, 7, 7, 7
195	4	3	1	...	0	2	8, 8, 8, 8, 8	7, 7, 7, 7, 7
196	4	3	1	...	0	1	8, 8, 8, 4, 8, 8, 7, 7, 7	7, 7, 7, 7, 7, 7, 7, 7, 7
197	4	3	1	...	0	2	8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 7, 7...	7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7...

Fig. 7.3: Data grouping

7.4 Data visualization: Heatmaps

Heatmaps are graphical representations used to visualize and explore patterns and relationships between the data. Individual values within a matrix are represented as colors. The intensity of the color represents the value of the data points, typically ranging from light colors for low values to dark colors for high values. However, the color scale can be chosen also to highlight particular visual aspects, to emphasise some data points.

One of the most decisive parameters that can affect the ALD material growth rate on a specific substrate, even without modifying any other parameter, is the features size. It is interesting to observe what is the relation between all the possible results on SiCOH and Cu with respect to all the possible features dimensions.
In the Figure 6.4 the color scale is chosen to emphasize the unique outcomes, that occur just one time, plotted in red.

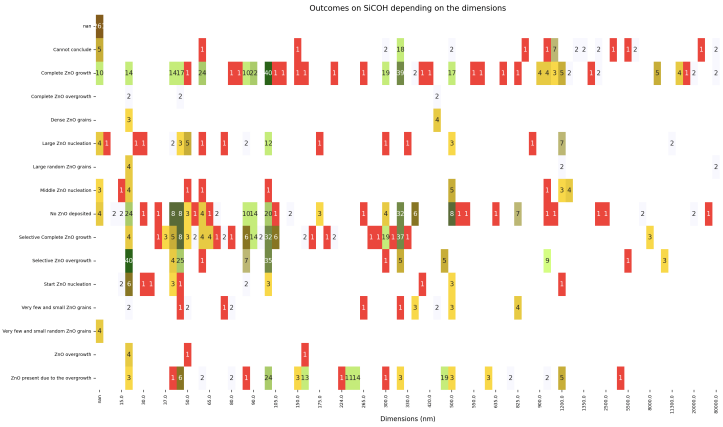

Fig. 7.4: Results on SiCOH with respect to all the possible features dimensions

The Figure 6.4 shows a large number of unique outcomes. In other words, if we consider a set of unique experimental conditions, based on the results/dimensions relationship, we expect just one result that occur just one time.

This can make the modeling more cumbersome, passing in input data that are

not well distributed. The problem domain is related to the frequency of success of a certain set of experimental conditions, so unique outcomes are not so representative. We want as many results as possible for each set of experimental conditions, divided in good or bad results, in order to assess overall whether the experiment is successful and with what frequency.

To better redistribute the data, I decided to regroup all the possible features sizes in three broad categories, The lower and upper limits can be easily tailored, depending on the initial data distribution and the specific aim.

- 0 nm - 100 nm
- 100 nm - 1 μm
- $>$ 1 μm

After grouping the dimensions, as visible in the Figure 6.5, the number of unique outcomes has significantly decreased and, in relation to the intervals boundaries, the data appear well balanced/distributed with a significant sample size for each interval.

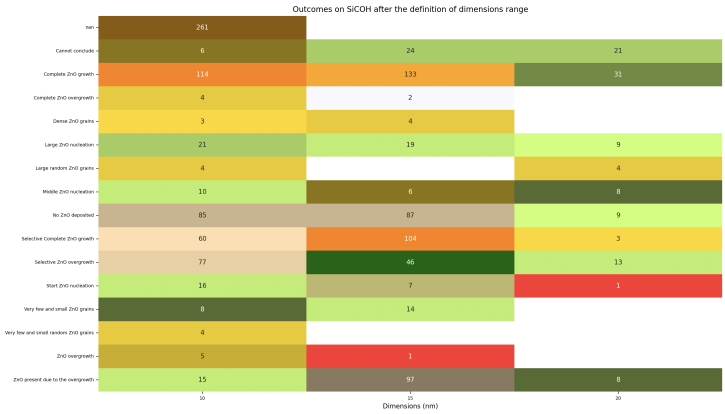

Fig. 7.5: Results on SiCOH after features dimensions class definition

8 Results categorization

The success of an experiment, involving a blanket coupon as substrate, relies solely on the growth or not of the ALD material on the substrate, depending on the specific objective of the experiment; the growth rate does not count.
However, for patterned coupons, an experiment can be deemed successful only if the ALD material selectively deposit on one of the two substrate subsections and, depending on its growth rate, there are different degrees of selectivity that determines different degrees of success. A cross-validation between the result on SiCOH and on Cu is needed.

I labeled 14 possible results for both SiCOH and Cu, so 196 results pairs are possible. All the results pairs that appear in our experiments are visualized in the Figure 7.1. Even if not all the possible results pairs are present, and the data points are not really balanced, because some pairs occur much more frequently than the majority of others, overall there is a sufficient quantity of well distributed combinations available.

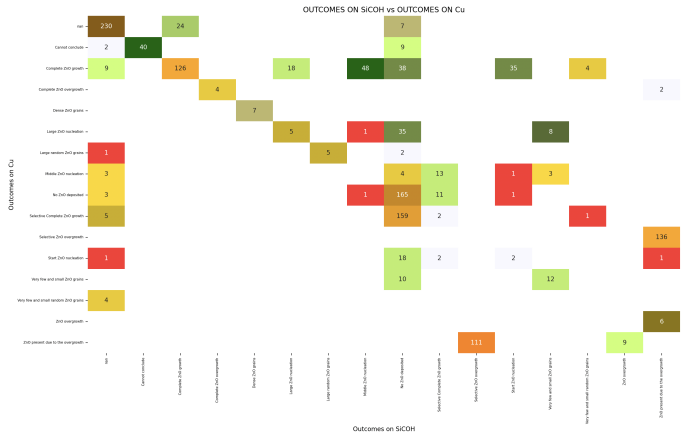

Fig. 8.1: Results on SiCOH vs Results on Cu

The results pairs are divided first into seven classes:

- Complete selectivity;
- Almost complete selectivity;
- Good selectivity;
- Not good selectivity;
- Almost no selectivity;
- No Selectivity;
- Cannot Conclude.

This first categorization exhibits not only positive or negative results, but also neutral results, that neither strongly indicate success nor failure. To remove this gray area, the classes are split just in two macro sections: the first three classes are grouped under successful results, while all the others are grouped under not successful results.

For blanket coupons the outcomes are labeled simply by considering if the ALD material is deposited or not on the substrate, so just two classes are defined, and the categorization in successful or not successful results is determined by the goal of the specific experiment.

For all the unique sets of experimental conditions we know all the verified results, the occurrence of each result and if it is labeled as successful or not successful. For instance, for each of them we can calculate the frequency of success, so the number of successful results out of the total, showed in Figure 7.2.
The frequency of success represents the milestone for the artificial intelligence (A.I.) model.

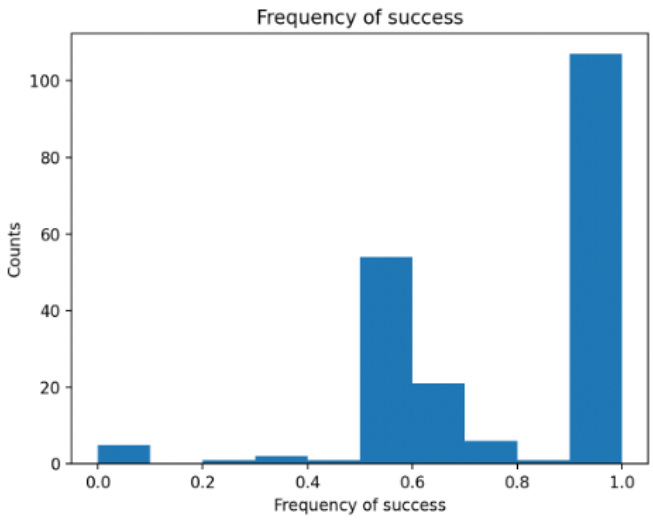

Fig. 8.2: Frequency of success

8.1 Dataset splitting

The input data are divided in two distinct subsets: training dataset and testing dataset.
The purpose of having separate training and testing datasets is to evaluate the model's performance on independent data, getting an unbiased assessment of the model's predictive abilities.
A reasonable splitting of the input data is crucial. Once the datasets sizes are determined, the data points are randomly assigned to either the training or testing set. This random assignment helps ensure that both datasets have a similar distribution of points and avoids any bias that may arise from a specific ordering of the data.

2/3 of the input data, from the frequency of success, are assigned to the training dataset and 1/3 to the testing dataset, randomly shuffled.

The training dataset is used to train the machine learning model, so to adjust the model's parameters and weights through various optimization algorithms. The testing dataset serves as unbiased evaluation set to verify the performance of the trained model. It helps determine how well the model generalizes to unseen data and provides an estimate of its predictive accuracy, detecting possible overfitting or underfitting.

Underfitting occurs when the model is too simple to capture the underlying patterns in the training data. Signs of underfitting include high training and testing errors and poor performance on both the training and testing datasets, resulting in oversimplified or generalized predictions.

Overfitting happens when a machine learning model becomes too complex and starts to memorize the training data instead of learning the underlying patterns. The model becomes too specific to the training dataset, capturing noise or random fluctuations in the data, making it less effective in generalizing to new data. Signs of overfitting include a low training error but a high testing error, excellent performance on the training dataset but poor performance on new, unseen data.

9 A.I. Model creation

The most appropriate A.I. models for processing my input data are based on classification and regression.

Classification models are used when the target variable is categorical or discrete. The goal is to assign input data points to predefined classes or categories, in our case successful results or not successful results. The model learns from labeled training data, where each data point is associated with a specific class and during training, it identifies patterns and relationships in the input features to make predictions on unseen data.

Regression models are used when the target variable is continuous or numerical. The goal is to estimate or predict a numerical value based on input features. The model learns from labeled training data, where each data point has both input features and a corresponding numerical target value, and captures the relationship between the input variables and the target one, to make predictions on new data.

Different algorithms and techniques can be employed to build these models. I used XGBoost, a decision tree algorithm based on gradient boosting, to boost the performances of weak predictive models.

9.1 Hyperparameters tuning

Hyperparameters are the configuration settings or parameters that control the overall behavior, architecture, and configuration of the A.I. model. They are external parameters of the model, in the sense that they are not learned from the data, and they are set before the training process. They can be directly set by the model developer or determined by tuning process, to optimize the model performances.

In this case the hyperparameters are not set directly, but I defined ranges of values within which they can be tuned, by exploiting an optimization algorithm based on Bayesian Optimization.

The hyperparamters specified for the models are:

- Learning rate: it controls the step size at which the gradient boosting algorithm updates the model's internal parameters during training.
- Max depth: it determines the maximum depth or number of layers in each tree of the gradient boosting model. Increasing it allows the model to capture more complex relationships in the data.
- Number of estimators: it specifies the number of individual decision trees or estimators in the gradient boosting ensemble. Increasing it can improve the model's performance, but it also increases the computational cost.
- Subsample: it controls the fraction of the training data to be used for each individual tree in the gradient boosting process.
- Colsamplebytree: it specifies the fraction of features (columns) to be randomly sampled for each tree in the gradient boosting process. This introduces further randomness and can prevent overfitting by forcing the model to focus on different subsets of features.

9.2 Model training and testing

The model's training is an iterative process. Each iteration involves a forward pass, loss calculation, backward pass and internal parameter update.

In this case, several training loops occur, each time defining a different set of hyperparameters, to try to optimize them, as explained in the previous paragraph.

At the start of the training loop, the model's parameters are initialized with random or predefined values. These parameters represent the learnable weights and biases that the model will update during training, that influence the relationship between the model's inputs and outputs.

The training loop begins by passing the input data through the model's layers or computational units, which result in a predicted output. During the forward pass the training dataset undergoes cross-validation. Cross-validation involves partitioning the available data into subsets to train and test the model multiple times.
The type of cross-validation used is the k-fold cross-validation:
The data is divided into k equal-sized subsets or "folds". The model is then trained k times, each time using k-1 folds for training and the remaining fold for testing. This process is repeated for each fold, and the performance results are averaged to obtain an overall evaluation of the model.

The next step is to calculate the loss, which quantifies the difference between the predicted output and the true or expected ones. The choice of loss function de-

pends on the specific task and the on nature of the data. In regression problems, the loss function uesd is the mean squared error (MSE), while in classification problems is the accuracy.

After computing the loss, the training loop proceeds with the backpropagation, that involves calculating the gradients of the loss with respect to the model's internal parameters. This step measures the sensitivity of the loss to the changes of each parameter and helps determine how much each parameter contributes to the error. With the gradients computed, an optimization algorithm is employed to update the model's parameters, to optimize the loss.

The steps above are repeated for a predefined number of iterations. Once the training loop is completed, the final model represents the learned knowledge, with optimized loss, and it is ready for evaluation on unseen data.

The testing process begins by performing a forward pass of the testing dataset. Each input sample is fed into the model, that produces a predicted output. Based on the model's output for each input sample, predictions are calculated.

The next step is to evaluate the performance of the model using appropriate metrics. The choice of metrics depends on the specific task. The same training evaluation metrics are exploited also for testing.

The best results obtained until now are:

- Accuracy on training data: 75%
- Accuracy on testing data: 63%
- MSE on training data: 1.63
- MSE on testing data: 0.065

If the model's performance is not satisfactory, further strategies of model improvement may be performed to enhance its execution.

Part IV

Conclusions

10 Conclusions and future perspective

Until now, the level of accuracy and the MSE values obtained for both the training and testing datasets, assess an acceptable model's performance. As first trial, this can be considered a good result.
However, several possible strategies can be applied to enhance the performance of both the Classification and Regression models. The aim is to obtain a higher level of accuracy and a MSE as lower as possible.

One of the most effective and immediate approach is to increase the number of training loops, to try to get more relationships between the given data and perform better on the unseen ones.

Also, a further hyperparameter tuning can be performed. Earlier, a range within which to search for hyperparameters values was defined. The boundaries of the defined hyperparameters ranges are quite standard to be applied to any Classification and Regression problem. The research space can be enlarged, to further optimize the hyperpatameters values and so the models architecture.

Instead of tailoring the models parameters, data engineering can be different tactic. As we can see from Figure 8.2, more than half of the experiments shows 100% of successful results. If each percentage, from 0% to 100% of success, represents a class, the 100% successful class is constituted by a very large amounts of cases than all the others.

With the data randomly shuffled while split in training and testing datasets, it can easily happen that the majority of the training data points belongs to the 100% successful class.
This can be a problem because it can produce biased model predictions. Furthermore, not including all the classes in the training dataset means that the data treated are not representative for the problem.

A possible solution is to normalize the classes and to balance the data, so that the same number of cases are taken from each class to form the training and testing datasets.

In conclusion, great benefits can be gathered from the optimization of this A.I. model.

First, ADS requires many experimental steps before reaching a consistent results. This time-consuming and expensive aspect can be overcome by predicting optimum combinations of experimental conditions to achieve successful results.
Moreover, the chemical specificity requirement necessary for selectivity can be relaxed, suggesting new classes of inhibitors, characterized by precise chemical and physical properties for each ASD requirement.

This may play an important role in significantly reducing the difficult materials challenges in micro and neno electronic and the process complexity associated with new 3D structured devices.

Bibliography

[1] Richard W Johnson, Adam Hultqvist, and Stacey F Bent. „A brief review of atomic layer deposition: from fundamentals to applications“. In: *Materials today* 17.5 (2014), pp. 236–246.

[2] L Li et al. „Towards accurate and efficient process simulations based on atomistic and neural network approaches“. In: *2022 International Electron Devices Meeting (IEDM)*. IEEE. 2022, pp. 15–6.

[3] Krystelle Lionti et al. „Area-Selective Deposition of Tantalum Nitride with Polymerizable Monolayers: From Liquid to Vapor Phase Inhibitors“. In: *Chemistry of Materials* 34.7 (2022), pp. 2919–2930.

[4] Magi Mettry et al. „Extending the compositional diversity of films in area selective atomic layer deposition through chemical functionalities“. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 37.2 (2019), p. 020923.

[5] Fatemeh Sadat Minaye Hashemi, Chaiya Prasittichai, and Stacey F Bent. „Self-correcting process for high quality patterning by atomic layer deposition“. In: *Acs Nano* 9.9 (2015), pp. 8710–8717.

[6] Thomas Jeffrey Lomax Mustard et al. „Quantum mechanical simulation for the analysis, optimization and accelerated development of precursors and processes for atomic layer deposition (ALD)“. In: *Journal of the Korean Ceramic Society* 53.3 (2016), pp. 317–324.

[7] Thomas G Pattison et al. „Surface initiated polymer thin films for the area selective deposition and etching of metal oxides“. In: *ACS nano* 14.4 (2020), pp. 4276–4288.

[8] David Sibanda, Sunday Temitope Oyinbo, and Tien-Chien Jen. „A review of atomic layer deposition modelling and simulation methodologies: Density functional theory and molecular dynamics“. In: *Nanotechnology Reviews* 11.1 (2022), pp. 1332–1363.

[9] Rudy Wojtecki et al. „Fifteen nanometer resolved patterns in selective area atomic layer deposition—defectivity reduction by monolayer design“. In: *ACS applied materials & interfaces* 10.44 (2018), pp. 38630–38637.

[10] Angel Yanguas-Gil and Jeffrey W Elam. „Machine learning and atomic layer deposition: Predicting saturation times from reactor growth profiles using artificial neural networks“. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 40.6 (2022), p. 062408.

List of Figures

List of Tables