

# **Accelerating Metal-Organic Framework Optimization through Automation and Machine Learning**

Zur Erlangung des akademischen Grades eines  
DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften  
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

**Lena Pilz**

aus Überlingen

1. Referent: PD Dr. Manuel Tsotsalas

2. Referent: Prof. Dr. Stefan Bräse

Tag der mündlichen Prüfung: 22.04.2024



This document is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>

# Honesty Declaration

This work was carried out from July 1st, 2021, to March 7th, 2024, at the Institute of Functional Interfaces (IFG), Faculty of Chemistry and Biosciences at the Karlsruhe Institute of Technology (KIT) under the supervision of PD. Dr. MANUEL TSOTSALAS.

Die vorliegende Arbeit wurde im Zeitraum vom 01. Juli 2021 bis 7. März 2024 am Institut für Funktionelle Grenzflächen (IFG) an der Fakultät für Chemie und Biowissenschaften am Karlsruher Institut für Technologie (KIT) unter der Leitung von PD. Dr. MANUEL TSOTSALAS angefertigt.

Hereby, I, LENA PILZ, declare that I have independently authored the present work and have not used any resources other than those specified, and that all material published by others is adequately cited. This dissertation has not been submitted to any other college or university previously.

Hiermit versichere ich, LENA PILZ, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet, sowie Zitate kenntlich gemacht zu haben. Die Dissertation wurde bisher an keiner anderen Hochschule oder Universität eingereicht.

# Abstract

This thesis explores the path towards fully autonomous laboratories, focusing on the enhancement of AI-driven material optimization processes. Through the application of machine learning (ML) algorithms, the integration of automated synthesis systems, and effective research data management, this work addresses key challenges in accelerating materials science research. Through three distinct projects, each varying in complexity and goals, this work optimizes specific properties of Metal-Organic Frameworks (MOFs) using an AI-integrated approach. Each project simultaneously adjusts multiple parameters, speeding up the optimization process and uncovering new insights into the interplay of various parameters in MOF synthesis. The use of AI methods reveals broader parameter space insights, challenging some established knowledge by demonstrating that certain widely accepted claims may not hold universally.

The first project achieves a targeted 111-orientation in HKUST-1 surface-anchored MOF (SURMOF) films, emphasizing the significance of automated synthesis for reproducibility and efficiency in machine learning optimizations. It also uncovers that high crystallinity can be reached without water as a modulator, challenging the traditional belief in water's necessity for such processes. The second project explores the DASBDC-pillar-layered SURMOF system, resulting in the creation of a highly structured polymer referred to as SURGEL. This optimization showcases the machine learning tool's effectiveness and highlights the possibilities introduced by post-synthetic modifications. These adjustments, made possible by the earlier ML optimization, led to the creation of a successful gas separation membrane. The third project challenges the ML tool with a complex task of reducing surface roughness in HKUST-1 SURMOFs, aiming for applications in electronic devices. This endeavor reflects on the importance of the initial synthesis cycles and shows the potential for further optimization processes to reduce roughness in other SURMOF systems.

Further, the thesis compares the effectiveness of genetic algorithms with Bayesian Optimization for nanoparticle size optimization, confirming the efficiency of genetic algorithms. It also suggests that Bayesian Optimization could become beneficial for future optimizations with further refinement and tailoring to specific requirements. A significant contribution of this work is the adaptation of the ML tool SyCoFinder employed in the optimization projects to handle categorical vari-

ables, broadening its applicability for complex optimization tasks. The last project deals with the creation of a new input mask for robotic syntheses in a research data management system, to promote openly accessible and reproducible data in line with the FAIR principles.

In summary, this thesis not only advances the field of materials science through evaluating and enhancing innovative optimization strategies and ML tools but also lays a foundation for future research towards the realization of fully autonomous laboratories for materials development.

# Zusammenfassung

Diese Arbeit erforscht Aspekte auf dem Weg zu vollständig autonomen Laboren und konzentriert sich auf die Verbesserung von KI-gesteuerten Materialoptimierungsprozessen. Durch die Anwendung von Algorithmen des maschinellen Lernens (ML), die Integration von automatisierten Synthesystemen und eine effektive Verwaltung von Forschungsdaten werden in dieser Arbeit zentrale Herausforderungen bei der Beschleunigung der materialwissenschaftlichen Forschung angegangen. In drei verschiedenen Projekten, die sich in ihrer Komplexität und ihren Zielen unterscheiden, werden spezifische Eigenschaften von metallorganischen Gerüsten (MOFs) mit Hilfe eines KI-integrierten Ansatzes optimiert. In jedem Projekt werden mehrere Parameter gleichzeitig angepasst, wodurch der Optimierungsprozess beschleunigt wird und neue Erkenntnisse über das Zusammenspiel verschiedener Parameter bei der MOF-Synthese gewonnen werden können. Der Einsatz von KI-Methoden offenbart breitere Erkenntnisse über den Parameterraum und stellt einige etablierte Erkenntnisse in Frage, wobei gezeigt wird, dass bestimmte weithin akzeptierte Behauptungen nicht universell gültig sind.

Im ersten Projekt wird eine gezielte 111-Orientierung in HKUST-1 oberflächenverankerten MOF-Filmen (SURMOFs) erreicht, was die Bedeutung einer automatisierten Synthese für die Reproduzierbarkeit und Effizienz bei Optimierungen durch maschinelles Lernen unterstreicht. Es zeigt auch, dass eine hohe Kristallinität ohne Wasser als Modulator erreicht werden kann, was die traditionelle Annahme, dass Wasser für solche Prozesse notwendig ist, in Frage stellt. Das zweite Projekt erforscht das DASBDC-pillar-layered SURMOF system, das zur Schaffung eines hochstrukturierten Polymers namens SURGEL führt. Diese Optimierung zeigt die Effektivität des maschinellen Lernwerkzeugs und verdeutlicht die Möglichkeiten, die sich durch nachträgliche synthetische Modifikationen ergeben. Diese Anpassungen, die durch die vorangegangene ML-Optimierung ermöglicht wurden, führten zur erfolgreichen Entwicklung einer Gastrennmembran. Im dritten Projekt wird das ML-Tool mit der komplexen Aufgabe konfrontiert, die Oberflächenrauheit von HKUST-1 SURMOFs zu reduzieren, die für Anwendungen in elektronischen Geräten vorgesehen sind. Dieses Unterfangen betrachtet die Bedeutung der anfänglichen Syntheszyklen und zeigt das Potenzial von Optimierungsprozessen zur Verringerung der Rauheit in anderen SURMOF-Systemen.

Darüber hinaus vergleicht diese Arbeit die Effektivität genetischer Algorithmen mit der Bayes'schen Optimierung am Beispiel einer Optimierung von einheitlichen Nanopartikelgrößen und bestätigt die Effizienz genetischer Algorithmen. Sie legt auch nahe, dass die Bayes'sche Optimierung für künftige Optimierungen von Nutzen sein könnte, wenn sie weiter verfeinert und besser auf spezifische Anforderungen zugeschnitten wird. Ein wesentlicher Beitrag dieser Arbeit ist die Anpassung des in den Optimierungsprojekten eingesetzten ML-Tools SyCoFinder

an kategorische Variablen, wodurch seine Anwendbarkeit für komplexe Optimierungsaufgaben erweitert wird. Das letzte Projekt befasst sich mit der Erstellung einer neuen Eingabemaske für Robotersynthesen in einem Forschungsdatenmanagementsystem, um offen zugängliche und reproduzierbare Daten im Einklang mit den FAIR-Prinzipien zu fördern.

Zusammenfassend lässt sich sagen, dass diese Arbeit nicht nur den Bereich der Materialwissenschaften durch die Evaluierung und Verbesserung innovativer Optimierungsstrategien und ML-Tools vorantreibt, sondern auch eine Grundlage für künftige Forschungen zur Realisierung vollständig autonomer Labore für die Materialentwicklung schafft.

# Contents

1. <i>Motivation</i> . . . . .	1
2. <i>Introduction</i> . . . . .	4
2.1 Metal-Organic Frameworks (MOFs) . . . . .	4
2.2 Surface-anchored MOFs (SURMOFs) . . . . .	6
2.2.1 Layer-by-Layer Method . . . . .	7
2.2.2 SURGELS . . . . .	8
2.3 Machine Learning . . . . .	8
2.3.1 Machine Learning in Material Science . . . . .	10
2.3.2 SyCoFinder . . . . .	11
2.3.3 Genetic Algorithm . . . . .	12
2.3.4 Bayesian Algorithm . . . . .	13
2.4 Automation . . . . .	13
2.5 Research Data Management . . . . .	15
3. <i>Methods</i> . . . . .	16
3.1 X-Ray Diffraction (XRD) . . . . .	16
3.1.1 Theory . . . . .	16
3.1.2 XRD Measurement . . . . .	18
3.2 Infrared Spectroscopy . . . . .	19
3.2.1 Theory . . . . .	19
3.2.2 IRRAS Measurement . . . . .	21
3.3 Scanning Electron Microscopy (SEM) . . . . .	21
3.3.1 Theory . . . . .	22
3.3.2 SEM Measurement . . . . .	22
3.4 Ellipsometry . . . . .	22
3.4.1 Theory . . . . .	23
3.4.2 Ellipsometry Measurement . . . . .	24
3.4.3 Fitting and Models . . . . .	24
3.5 Dynamic Light Scattering (DLS) . . . . .	26
3.5.1 Theory . . . . .	26
3.5.2 DLS Measurement . . . . .	26
4. <i>Machine Learning Optimizations</i> . . . . .	28
4.1 MOF thin film Optimization towards Orientation . . . . .	28
4.1.1 Introduction . . . . .	28
4.1.2 ML Optimization . . . . .	31
4.1.3 Results and Discussion . . . . .	34
4.1.4 Summary . . . . .	38



4.2	MOF thin film Optimization for Gasseparation Application . . . .	39
4.2.1	Introduction . . . . .	39
4.2.2	ML Optimization . . . . .	41
4.2.3	Results and Discussion . . . . .	43
4.2.4	Summary . . . . .	51
4.3	MOF thin film Optimization towards Roughness Reduction . . . .	51
4.3.1	Introduction . . . . .	51
4.3.2	ML Optimization . . . . .	52
4.3.3	Results and Discussion . . . . .	54
4.3.4	Summary . . . . .	58
5.	<i>Development of Software Toolsets for AI Integration in MOF Synthesis .</i>	59
5.1	Comparison of Genetic Algorithm and Bayesian Optimization . . .	59
5.1.1	Introduction . . . . .	59
5.1.2	Results and Discussion . . . . .	61
5.1.3	Summary . . . . .	63
5.2	Modification of SyCoFinder . . . . .	63
5.2.1	Introduction . . . . .	63
5.2.2	Modification . . . . .	64
5.2.3	Summary . . . . .	68
5.3	Development of Chemotion input mask . . . . .	69
5.3.1	Introduction . . . . .	69
5.3.2	Results and Discussion . . . . .	70
5.3.3	Summary . . . . .	73
6.	<i>Summary and Outlook . . . . .</i>	74
7.	<i>Experimental . . . . .</i>	79
7.1	Substrate Preparation . . . . .	79
7.2	Solvents . . . . .	79
7.3	MOF thin film Optimization towards Orientation . . . . .	79
7.3.1	SURMOF synthesis . . . . .	79
7.3.2	Measurements . . . . .	80
7.4	MOF thin film Optimization for Gasseparation Application . . . .	80
7.4.1	Sample holder 3D-printing . . . . .	80
7.4.2	SURMOF synthesis . . . . .	80
7.4.3	Crosslinking . . . . .	81
7.4.4	Transformation to SURGEL . . . . .	81
7.4.5	Measurements . . . . .	82
7.5	MOF thin film Optimization towards Roughness Reduction . . . .	83
7.5.1	SURMOF synthesis . . . . .	83
7.5.2	Measurements . . . . .	83
7.6	Comparison GA vs BO . . . . .	83
7.6.1	ZIF-8 Nanoparticle synthesis . . . . .	83
7.6.2	Measurements . . . . .	84
8.	<i>References . . . . .</i>	85
9.	<i>Abbreviations . . . . .</i>	99

<i>Appendix</i>	104
A.1 MOF thin film Optimization towards Orientation . . . . .	105
A.1.1 Synthesis Parameter . . . . .	105
A.1.2 Fitness . . . . .	106
A.1.3 Possible program sequences emerging from the robotic set-up and the suggested parameters. . . . .	107
A.1.4 X-Ray Diffraction . . . . .	109
A.2 MOF thin film Optimization for Gas separation Application . . . . .	120
A.2.1 Synthesis Parameter . . . . .	120
A.2.2 Fitness . . . . .	122
A.2.3 X-Ray Diffraction . . . . .	124
A.3 MOF thin film Optimization towards Roughness Reduction . . . . .	129
A.3.1 Synthesis Parameter . . . . .	129
A.3.2 Fitness . . . . .	130
A.3.3 X-Ray Diffraction . . . . .	132
A.4 Comparison of Genetic Algorithm and Bayesian Optimization . . . . .	138
A.4.1 Provided training dataset from GA Optimization for BO . . . . .	138
A.4.2 Suggested Parametersets and Fitness Evaluation Values of BO . . . . .	139
A.4.3 X-Ray Diffraction . . . . .	140
A.5 Modification of SyCoFinder . . . . .	142
A.5.1 Final Code of the adapted Diverse Set . . . . .	142
A.6 Publications . . . . .	147
A.7 Conference Contributions . . . . .	148
A.8 Curriculum Vitae . . . . .	149

# 1

## Motivation

*"A computer would deserve to be called intelligent if it could deceive a human into believing that it was human."*

*Alan Turing*

The common application of artificial intelligence in science is still relatively new, with the possibilities that have emerged over the last two decades once considered impossible and at times even unimaginable in preceding years. However, the concept of a world infused with artificial intelligence in various forms as part of everyday life has a much deeper historical and imaginative lineage. In 1950, Alan Turing authored "Computing Machinery and Intelligence", delving into the idea of computers replicating human behavior to the point where recognizing intelligence as artificial became challenging for human beings, also referred to as "Imitation Game". Likewise, Isaac Asimov explored the realm of robotics and artificial intelligence (AI) in his writings as early as the 1940s. Nevertheless, when Asimov formulated the three laws of robotics, the world was still a long way away from the advanced state of artificial intelligence we now know. And for many years, these ideas remained firmly within the realm of science fiction. But with ideas developed by Turing and Asimov established, some of the initial foundations of AI development were paved, and in the subsequent years, alongside the advancement of computers, scientists continued to explore this field.

Today, these concepts manifest all around us, in the form of voice assistants, recommendation systems that analyze user behavior, and seemingly magical creations like large language models (e.g., GPT-#) that occasionally exhibit human-like capabilities. This underscores the fact that the field of artificial intelligence has become more relevant than ever. And so it has found application in science across multiple disciplines, enhancing the speed and accuracy of data analysis, and enabling new discoveries.

With the increasing integration of AI into scientific research, a particularly promising avenue is emerging within material science, where machine learning algorithms are revolutionizing the processes involved in designing, synthesizing, and optimizing novel materials, including porous materials such as metal-organic frameworks (MOFs). These materials are integral to a wide range of applications, from filtra-

## 1. MOTIVATION

tion and catalysis to storage solutions, thanks to their unique porous structure. The interest in MOFs is driven by their adaptability; they can be engineered at a molecular level to exhibit specific properties, making them highly versatile.

To fully harness the potential of MOF materials, initial screening of the chemical space is crucial to identify the optimal metal-linker pair for a specific application. Once this pair is determined, optimizing and tailoring their characteristics to align with the intended use becomes essential. This necessitates a comprehensive understanding and control of the synthesis process's influencing factors. These factors are often interdependent, requiring exhaustive experimental efforts for a thorough understanding. Traditionally, this involves adjusting one variable at a time to determine its impact, a method that is both time-consuming and costly. Besides, these conventional methods only explore a small portion of the potential parameter space, limiting the ability to fully understand the complex interactions between variables. Here, the application of machine learning (ML) methods presents a promising solution. ML can handle multiple interdependent parameters simultaneously, enabling a more rapid and focused exploration of the parameter space to identify optimal conditions. Moreover, ML methods investigate a much larger portion of the parameter space, uncovering possible unconventional parameter combinations that might remain unnoticed with conventional methods. However, a common hurdle in effectively training ML models in material science is the significant expense involved in generating practical training data. Therefore, data to train ML models is often scarce. This challenge is further accompanied by a bias in literature towards reporting successful experiments. But ML algorithms significantly improve in effectiveness when they can analyze both successful and unsuccessful outcomes, since such comprehensive data helps to highlighting zones of the parameter space where success is probable and areas where it is not.

In 2019, Moosavi et al. addressed this issue by developing an ML tool for optimizing metal-organic frameworks named SynthesisConditionFinder (SyCoFinder), designed to operate without any initial data. It begins by generating a statistical array of parameter combinations, thereby creating its own dataset that includes both successes and failures. This approach already narrows down the parameter space for subsequent exploration, allowing for the efficient training of an ML model. This uniquely constructed dataset, derived entirely from empirical experiments, does not rely on existing literature or assumptions. In the first half of my thesis, I explore several optimization challenges of metal-organic framework thin films using the SyCoFinder, rigorously investigating its effectiveness and universal applicability.

To truly elevate the efficiency of scientific research and minimize human errors, leveraging machine learning (ML) alone is not sufficient. A comprehensive approach to data driven science is essential. While ML can rapidly process data and yield effective outcomes when properly applied, its iterative nature demands clean, error-free input and thorough evaluation. A key strategy to meet these requirements involves integrating automated methods alongside ML. This dual approach not only reduces human errors, leading to higher reproducibility but also, depending on the automation technique, enables the parallelization of tasks, thereby speeding up the entire research process. In this study, I have employed automated synthesis methods, such as synthesis robots, and automated measurement techniques extensively.

However, envisioning a fully autonomous laboratory that optimizes materials in-

## 1. MOTIVATION

dependently requires a seamless integration of several components. I have outlined these components as interlocking gears in Figure 1.1: Beginning with automated robotic synthesis, which provides samples to automated measurement devices. These devices generate raw data that must be evaluated to assess the sample quality. This step generates a vast amount of data, which needs to be organized, stored, and converted into formats that are readable by machines for ML processing. Once the ML system is furnished with all necessary data in an appropriate format, it can optimize the synthesis conditions, thus initiating the production of the next sample. This cycle continues until the optimal sample is produced. For such autonomous optimization loops to function efficiently, each component must operate in concert, and each step presents its unique set of challenges.

In my thesis, I explore not only the practical ML application through the SyCoFinder but also examine another ML methodology. I conduct all syntheses using automated systems and address data management by enhancing and developing a structured data storage component within the Chemotion data management system, to facilitate machine-readable data transfer in the future. Moreover, in collaboration with Prof. Moosavi, I improve the SyCoFinder tool to accommodate a wider range of variable types in optimization processes.

By adhering to FAIR principles for data publication, we enable the scientific community to derive insights and gain knowledge from accessible, machine-readable data, further paving the way for advancements in research. All these efforts to address the individual challenges of each component are directed toward achieving a fully autonomous material optimization system utilizing advanced AI strategies with roots that trace back to Turing and Asimov, now representing a significant stride towards the future of materials science.

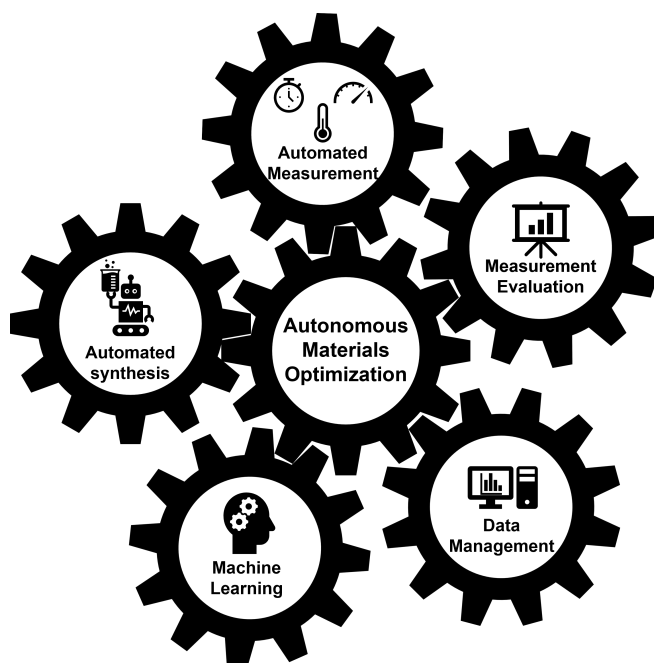


Fig. 1.1: Components of autonomous optimization.

## 2

# Introduction

This chapter introduces the main research topics and the foundational theories necessary for understanding the relationships and contexts within this thesis. However, these introductions are not intended to offer comprehensive insights or complete overviews of each respective topic. Instead, they aim to establish a knowledge base sufficient for comprehending the specific problems addressed in this thesis. The chapter begins with the fundamental understanding of Metal-Organic Frameworks (MOFs) and their relevant production methods. It progresses to an exploration of the implemented machine learning methods, which are crucial for goal-oriented and accelerated optimization of products and synthesis processes. Additionally, the chapter covers the automated robotic systems employed to ensure the reliability and reproducibility of results. Furthermore, the chapter highlights an enhanced data management system, specifically designed to meet FAIR criteria, emphasizing the creation of machine-readable and comprehensive data sets regarding MOF syntheses.

## 2.1 Metal-Organic Frameworks (MOFs)

Metal-Organic Frameworks, commonly abbreviated as MOFs, are characterized by their modular and permanent porous structures. This class of materials was introduced in the 1990s by O. Yaghi, initiating a significant field of research.[1] Prior to Yaghi's popularization of the term 'Metal-Organic Frameworks', these materials were often known as 'porous coordination polymers' (PCPs).[2] This earlier terminology emphasized their porous nature and the coordination bonds linking metal ions with organic ligands, resembling polymers. The adoption of the MOF nomenclature, largely influenced by Yaghi, reflects their distinct structure and properties more accurately, thereby differentiating them from traditional polymers.

These materials find extensive applications in cutting-edge areas including gas adsorption and separation, gas storage, chemical capture (e.g. extraction of carbon from the air), the development of optical devices, and processes of encapsulation and release for purposes like drug delivery.[3][4][5][6][7][8][9]

MOFs are composed of metal nodes and organic connectors, commonly referred to as linkers. A metal node can consist of either metal cations or metal clusters.[10] The extensive variety of these building components leads to an immense flexibility in MOF design, and thus also to a diverse range of shapes and sizes in their pores

## 2. INTRODUCTION

and channels.[11]

These components are assembled into uniformly distributed, porous networks - the Metal-Organic Frameworks. The construction principle is schematically illustrated in Figure 2.1. The metal centers, however, may exhibit varying coordination numbers, as well illustrated in Figure 2.1. This variation, in conjunction with the length, geometry, and functional groups of the linker molecules, is a crucial factor that significantly influences the pore shape and overall geometry of a MOF.

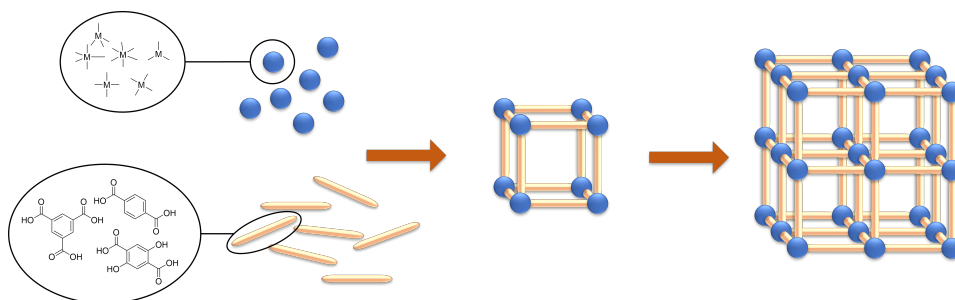


Fig. 2.1: Schematic representation of the construction of a metal-organic framework; the metal nodes (blue spheres) connect with the organic linker molecules (yellow struts) to form a three-dimensional porous network.

The term commonly employed to explain the construction of MOFs is the 'Secondary Building Unit' (SBU), representing the smallest repetitive entity within a MOF structure. Typically, these units exhibit the characteristic of linking multiple metal clusters to M-O-C clusters through the utilization of bidentate ligands, such as carboxylic acids.[12]

A well-known example of an SBU structure is the 'paddlewheel' configuration, where two metal cations are connected by four carboxylic groups (See Figure 2.2a).[13]

This structural pattern is featured in the widely recognized MOF known as HKUST-1, consisting of copper-cations and 1,3,5-benzenetricarboxylate (BTC) as bridging ligand (See Figure 2.2b). First reported by Chui et al. in 1999, HKUST-1 has since been extensively researched due to its straightforward synthesis methods leading to high yields. Besides, as demonstrated with gases such as methane and carbon dioxide, HKUST-1 has outstanding gas storage capabilities.[14][15][16]

To synthesize a bulk or powdered MOF, the solvothermal method is the most common synthesis approach. In this process the concentration of metal salts and linkers, the solubility of these reactants in the chosen solvent, the solution's pH value, and the temperature are the most crucial parameters. Solvents used are typically those with high boiling points and strong polarity, such as water, acetonitrile, dimethyl sulfoxide (DMSO) or dimethylformamide (DMF). While certain established guidelines can suggest starting conditions, achieving the optimal formation of these crystalline structures often necessitates a significant amount of experimentation and a trial-and-error methodology.[17]

## 2. INTRODUCTION

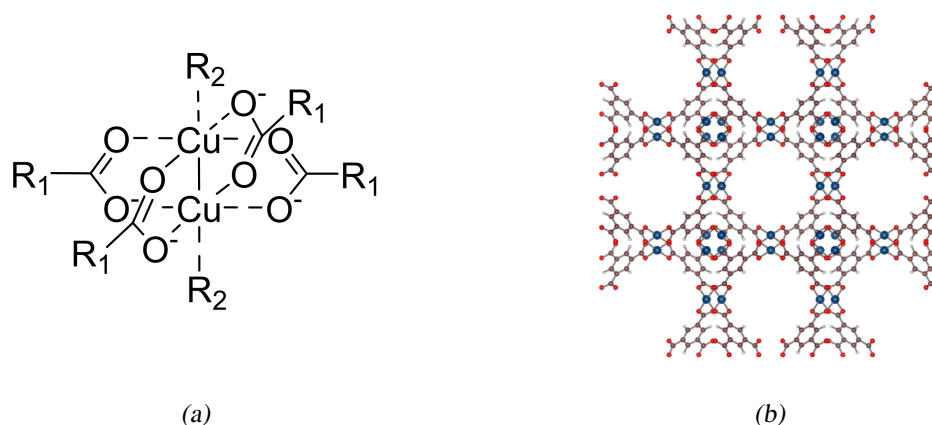


Fig. 2.2: a) Schematic illustration of the copper paddlewheel, a highly popular metal-oxo cluster, featuring a di-copper core coordinated by four carboxylic groups. b) Structural representation of an HKUST-1 MOF, composed of copper (blue) and trimesic acid (brown: carbon, red: oxygen, white: hydrogen).

Over the past century, also the synthesis in microwave reactors has emerged as a promising and reliable method, particularly for generating nanoparticles. This technique is typically much faster than the solvothermal method and is renowned for yielding highly uniform but smaller nanoparticles. One key characteristic is that microwaves provide evenly distributed heating throughout the entire vessel. In contrast, the solvothermal method primarily heats from the outside to the inside, potentially leading to less uniform heating. However, the microwave synthesis is limited by the choice of solvents since not all solvents are suitable. For instance, the use of corrosive solvents is restricted due to the microwave setup. Solvents with boiling points lower than the desired reaction temperature can lead to high pressures. Although microwave reactors can withstand some pressure, they are not as pressure-resistant as solvothermal methods. This limitation stems from the fact that microwave synthesis often employs septums to cap glass vessels, whereas solvothermal methods typically use more robust stainless steel autoclaves.[18][19]

### 2.2 Surface-anchored MOFs (SURMOFs)

Surface-anchored Metal-Organic Frameworks (SURMOFs) represent a specialized subset of MOFs, uniquely constructed directly on substrate surfaces.[20] The application of SURMOFs has gathered significant interest due to their potential in areas such as sensing, and molecular separation via selective membranes, leveraging their unique surface interactions and structural properties. The utility of SURMOFs in creating thin films is especially remarkable, as these films have potential applications in electronic devices and various optical technologies.[21][22][23] This versatility not only positions them as a key focus in advanced material science research but also underscores the need for precise control over the growth mechanism to ensure the production of flawless and reproducible end products.

The applied substrates may vary widely, with common choices including silicon wafers and alumina substrates. Often, these substrates are pre-coated with mate-



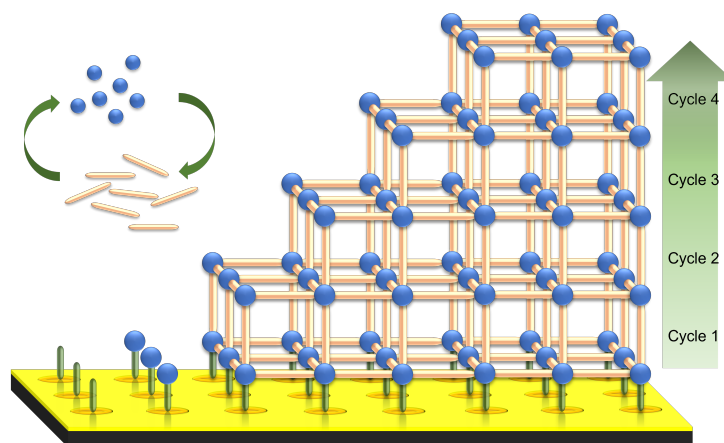
## 2. INTRODUCTION

rials like gold, which serve to create a smooth and uniform surface, providing the base for optimal MOF growth. In this thesis, the investigations focus on these critical control factors, emphasizing the meticulous selection and optimization of these parameters, which can be a laborious yet essential process for achieving desired results.

To establish the initial MOF layer on the substrate, functional groups are required for binding. Various methods exist for this purpose: a common approach involves activating the surface via ozone treatment or oxygen-plasma, both of which result in terminal OH-groups that enable MOF attachment. Another method utilizes self-assembled monolayers (SAMs) – molecules with dual functional ends. One end binds to the substrate (thiol-groups are often used for gold surfaces), while the other end binds the MOF. SAMs provide considerable flexibility in pre-adjusting some of the SURMOFs' properties, such as orientation. By varying the SAM molecules and thus for example their terminal groups or their chain length, it is possible to change the spacing and angle of these molecules on the surface. This variation may influence the preferred orientation of the growing SURMOF, yet it isn't the only factor influencing the direction in the resulting patterns.[24][25]

### 2.2.1 Layer-by-Layer Method

The fabrication of SURMOFs on substrates typically yields thin film structures, offering exceptional control over their thickness and structural characteristics. A pivotal advancement in the field was the introduction of the layer-by-layer technique in 2007 by Shekhah et al., which has proven instrumental in achieving well-defined and precisely structured SURMOFs.[26][27]



*Fig. 2.3:* This figure illustrates the formation of surface anchored metal-organic frameworks. Metal nodes (blue spheres) interact with organic linker molecules (yellow bars) to produce highly oriented SURMOFs in a layer-by-layer fashion. This method commences with a self-assembled monolayer (SAM, green bars), which forms the initial layer for SURMOF growth. [Adapted and reproduced from: [28]]

This method entails a sequential immersion process. Initially, the substrate is dipped into a metal salt solution. Subsequently, it is rinsed in a vessel with pure solvent to remove excess reactants. Next, the substrate is immersed in a separate

## 2. INTRODUCTION

vessel containing only the dissolved linker. This is followed by another cleaning step in a new solvent vessel. Each complete sequence of these steps forms one cycle. Repeating this cycle a specific number of times allows for the gradual construction of an epitaxial SURMOF thin film. This process achieves precise control over the film's thickness and structural properties, ensuring accuracy and consistency in the final product. Figure 2.3 displays the construction of such a thin film: A gold-coated silicon wafer serves as substrate, which is covered with a SAM represented by the green bars, serving as a connection layer between the substrate and the thin film. The latter is then built on top, following the described layer-by-layer method connecting metal nodes (blue spheres) with linker molecules (yellow bars) in an alternating pattern. In the case of the previously mentioned HKUST-1, one of the most commonly used SAMs is MHDA (16-Mercaptohexadecanoic Acid). For the metal solution, the SAM-coated substrate is typically immersed in dicopper acetate dissolved in ethanol and the corresponding linker is trimesic acid (Benzene-1,3,5-tricarboxylic acid, BTC), also dissolved in ethanol.[26][28][29]

### 2.2.2 SURGELS

When constructing bio-applicable thin films, the presence of metal atoms often presents a limitation or even serves as a criterion for exclusion, given their general toxicity to cells and bacteria. In this context, the synthesis of SURGELS offers a viable approach for creating structured, oriented, and metal-free thin films.

This process begins when linker molecules in the SURMOF structure contain functional groups that remain unbound to the metal and are available after the synthesis. These available functional groups can be exploited to crosslink the SURMOF structure. For this, crosslinking molecules are introduced into the SURMOF pores, where they bind to these functional groups, interconnecting the SURMOF's linker molecules throughout the structure. This can be achieved, for example, through click chemistry, which is activated under a specific wavelength.[30]

Alternatively, the crosslinking molecules can be designed to bind to each other, again occurring through the windows of the SURMOF pores.

The underlying concept in both approaches is to utilize the highly structured and oriented framework of the SURMOF as a template for constructing a metal-free polymeric substructure. The actual SURGEL is formed in a subsequent step, wherein the metal nodes of the SURMOF are removed, for instance, via an etching process. This leaves behind a polymeric network that retains the structure of the original SURMOF framework.[31][32]

## 2.3 Machine Learning

A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .

*Tom M. Mitchell, 1997, in "Machine Learning", Professor at Carnegie Mellon University[33]*

The terms 'artificial intelligence', 'machine learning', 'deep learning', and 'neural

## 2. INTRODUCTION

networks' are frequently used in contemporary discourse. To facilitate a clear understanding and accurate usage of these terms in the correct context, Figure 2.4a offers a concise overview that categorizes and delineates their relationships.[34]

In the past few years, machine learning has significantly impacted various sectors of science and everyday life. Notably, it plays a pivotal role in developing autonomous vehicles and is also employed in search engines, voice assistants, and product recommendations.[35] Machine learning, as a sub branch of artificial intelligence, involves creating algorithms designed to solve complex tasks. One of the key functions of these algorithms is to develop predictive models based on training data. The machine learning definition of Tom M. Mitchell leads to the assumption that a computer program can be able to gain experience by learning from data and thereby enhance its performance or the predictive models it uses.[33]

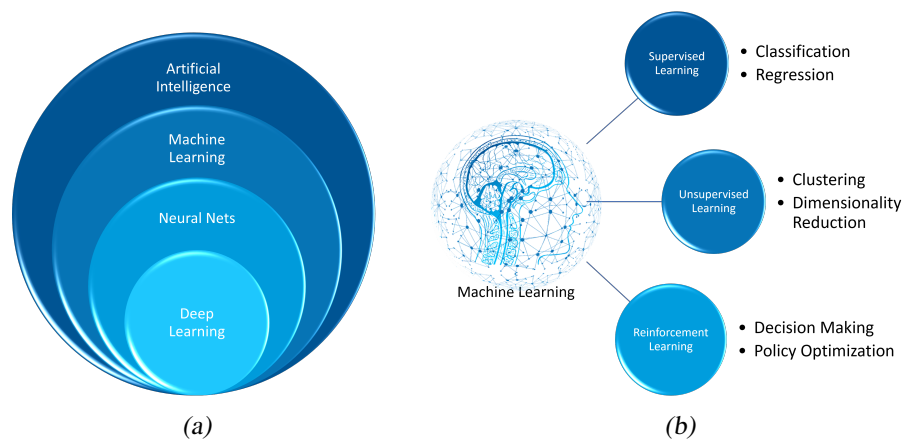


Fig. 2.4: a) Classification of the terms artificial intelligence, machine learning, neural networks and deep learning. [Graph inspired by Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials, Yulia Gavrilova, April 08, 2020 [34]] b) Illustration of the subtopics of machine learning: supervised learning, unsupervised learning and reinforcement learning. They are differentiated by their respective tasks.

Machine learning can be classified into three main categories: supervised learning, unsupervised learning, and reinforcement learning. Each category fundamentally differs in the types of problems it addresses, as illustrated in Figure 2.4b.

The goal of supervised learning is to learn from, or create a model using, training data in order to make quantitative predictions about future data. This training data is known as labeled data, meaning that the desired output values of these data are known or marked, as seen in applications like a spam filter. Such a filter can be trained to identify future spam emails through the manual labeling of unwanted emails.[36]

The classic tasks of supervised learning include classification and regression. Classification refers to the sorting of training data into specific, pre-established categories. Regression involves identifying relationships between independent parameters to achieve or predict outcomes related to a target variable.

Unsupervised learning, on the other hand, deals with unlabeled data, where both the output values and the structure of the data are unknown. The objective here is to discover the underlying structure within the data. One method is clustering, which

## 2. INTRODUCTION

involves dividing information into subgroups by exploratively finding commonalities in the properties of the data being studied. This can lead to the derivation of relationships between the data and data groups.

The third category, reinforcement learning, aims to enhance the performance of a system (an agent) through its interactions with the environment. Key tasks in this category include decision making, where the agent must choose the best actions to achieve its goals, and policy optimization, which involves refining the strategy that guides these decisions. Each interaction results in either positive or negative feedback, which assesses the action taken. The feedback is not necessarily binary (simply good or bad) but offers a measure of quality to the system. The system can then improve by adapting according to the most favorable feedback.

This thesis exclusively focuses on supervised learning methods, as the optimization tasks considered are inherently situated within the domain of supervised learning.[37][36]

### 2.3.1 Machine Learning in Material Science

To synthesize new materials in classical synthetic chemistry and materials science or to modify their properties, extensive and repetitive series of experiments are typically required. These experiments often involve separately testing various parameters to understand their individual influences. Common variables adjusted in these experiments include synthesis duration, temperature, additives, or solvent choice, all aimed at tailoring the specific properties of a material or enabling its synthesis. To circumvent the time-consuming and costly trial-and-error methods, data-driven approaches offer a way to expedite the modification process. This can be achieved, for instance, by mining existing literature for previously explored synthesis conditions and deriving insights from these findings via machine learning approaches.[38]

Another compelling reason for integrating machine learning (ML) algorithms into material science, particularly in the study of Metal-Organic Frameworks (MOFs), is the capability to predict and simulate structures. Given the vast number of potential MOFs beyond the nearly 90,000 that have already been reported, navigating through these materials to identify the most suitable one for a specific application becomes a formidable challenge.[39][40] Machine learning methods offer a promising pathway to efficiently screen these materials and pinpoint the optimal candidates for various applications.[41]

In this thesis, a machine learning algorithm is employed to predict synthesis parameters, with the ultimate goal of optimizing both the synthesis pathway and the resulting synthesis product. The training data for this algorithm are derived from an initial statistical series of experiments where several parameters are varied concurrently. This approach not only aids in deciphering complex correlations but also paves the way for more targeted and efficient material synthesis, demonstrating the growing importance and potential of machine learning in the field of material science.

## 2. INTRODUCTION

### 2.3.2 SyCoFinder

Integrating Machine Learning into the fields of chemistry and materials science, particularly within synthesis processes, presents complex challenges. The acquisition of sufficient datasets, a fundamental requirement for such integration, is particularly challenging since these datasets typically originate from experimental work. These experiments are constrained by financial and human resource limitations. Additionally, the tendency in scientific publications to highlight successful experiments poses a challenge for Machine Learning algorithms, which also benefit from data on experiments that were unsuccessful or only partially successful.

Given these constraints, one viable approach to incorporating Machine Learning is the application of genetic algorithms. These algorithms fall under the term of evolutionary algorithms and are inspired by the 'survival of the fittest' principle.

The 'materialscloud' platform, a collaborative project by the Swiss Federal Institute of Technology Zurich (ETH) and the École Polytechnique Fédérale de Lausanne (EPFL), features the 'SynthesisConditionFinder' (SyCoFinder) tool.[42] The SyCoFinder represents a practical method for employing genetic algorithms in synthesis optimization, thereby embedding Machine Learning principles into chemistry and material science research. A notable aspect of SyCoFinder is that it requires no initial data, making it feasible to start entirely from scratch.[43]

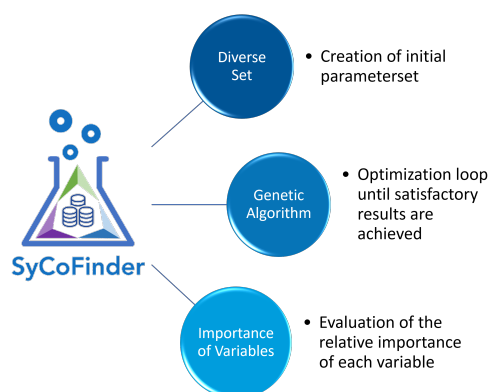


Fig. 2.5: This figure provides an overview of the three main tools of the synthesis condition finder (SyCoFinder). [SyCoFinder Logo reproduced from [42]]

As depicted in Figure 2.5, the SyCoFinder is structured into three main steps. The first is the 'Diverse Set,' a phase that allows for initiation without any prior data by generating an initial statistical set of parameters. Besides, this approach helps in effectively exploring the parameter space for comprehensive and efficient optimization. In this step, users specify the number of experiments they wish to conduct, select variables considered relevant in the process, and set their respective lower and upper limits. It is also possible to assign different weights to each variable. Based on this input, the 'Diverse Set' tool employs a MaxMin method to create an initial set of parameters. The MaxMin method seeks to maximize the minimum distance between different points in the parameter space, ensuring that an initial dataset for an optimization process covers a broad and diverse range of the selected parameters.[43]

This formed parameter set comprises combinations of the different variables, with each combination corresponding to a specific experiment to be conducted. Fol-

## 2. INTRODUCTION

Following the execution of these experiments, they are assessed and assigned a fitness value, a score ranging between zero and one. In the context of chemical experiments, this often involves measurements using various characterization methods. It is important to establish a consistent method for converting these measurement results into fitness values to maintain comparability. Therefore, devising a specific fitness formula is beneficial, a procedure, that will be detailed for each project in the upcoming chapters.

In the second step, based on the fitness values generated for each experiment in the Diverse Set, the genetic algorithm undertakes the actual machine learning optimization. Adhering to the principle of 'survival of the fittest,' the algorithm selects the best-performing syntheses, recombines them, and applies a mutation factor. As a result, SyCoFinder evolves a new, more optimized set of parameters, effectively employing evolutionary processes.[43] The subsequent step involves conducting these newly designed experiments, characterizing, and evaluating them to provide fresh data back to the genetic algorithm, as illustrated in the loop in Figure 2.6. This iterative process can be repeated as needed until the user achieves the desired level of satisfactory results, the criteria of which are user-defined.

The final component of the SyCoFinder, as depicted in Figure 2.5 and at the end of the loop in Figure 2.6, involves calculating the importance of variables. This process generates a bar chart that illustrates the relative importance of each selected variable in relation to one another, in the context of the optimization goal. To achieve this, data from the entire optimization process are taken into account. On one hand, this analysis can inform the weighting of variables in future, similar optimizations - particularly in cases where the chemical system being optimized is analogous, thereby potentially accelerating the optimization process. On the other hand, when combined with the variable values from the conducted experiments, this analysis may reveal insightful information regarding chemical dependencies.[43] These aspects will be explored in greater depth in subsequent chapters.

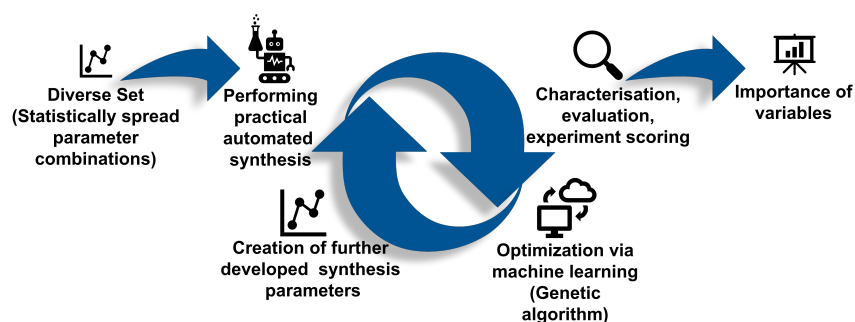


Fig. 2.6: This figure provides an overview of the actual optimization process provided by the synthesis condition finder (SyCoFinder). [Adapted and reproduced from: [30]]

### 2.3.3 Genetic Algorithm

Genetic algorithms (GA) are employed to identify the optimal solution within a vast parameter space, also termed global optimum. A parameter space includes the full array of potential solutions for an optimization problem, made up of chosen variables that are thought to impact, along with their respective ranges.

## 2. INTRODUCTION

These algorithms deploy a set of explorers, known as chromosomes, each possessing a unique combination of genes representing various potential solutions. In the context of SyCoFinder, these genes correspond to the different parameter combinations used in each experiment. The performance of each explorer in addressing the problem is evaluated. Regarding SyCoFinder, fitness values are attributed to each experiment, providing information about their performance. Following this evaluation, the most successful explorers are chosen. Their genes are then combined to form new explorers, or children, a process termed 'crossover'. Additionally, GAs introduce mutations - spontaneous, random changes - to discover previously untested solutions. As this process of selection and gene combination progresses, the GA incrementally approaches the most effective solution.[44] Figure 2.7 provides an overview of the steps of the genetic algorithms evolutionary process.

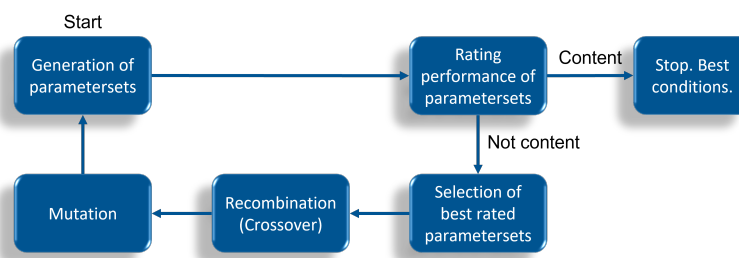


Fig. 2.7: This diagram depicts the sequence of steps processed by a genetic algorithm to predict and subsequently optimize parameter sets.

### 2.3.4 Bayesian Algorithm

In optimization tasks, such as those addressed by GAs, it's sometimes impractical to directly evaluate the objective function — the target of optimization, like reducing costs or increasing efficiency. This is where surrogate model methods prove valuable. These methods employ an alternative, or proxy model, to approximate the objective function. The most famous approach in this category is Bayesian Optimization (BO), renowned for its effectiveness. BO excels in navigating complex and elusive 'black-box' objective functions, whose internal mechanics are not entirely transparent. Its strength lies in adeptly managing noisy function evaluations, a frequent hurdle in optimization scenarios.

'Noisy function evaluations' in optimization refer to inconsistent or unreliable results from evaluating the objective function due to factors like measurement errors, data variability, or system randomness. Bayesian Optimization adeptly handles this noise by employing statistical techniques to model the function and its associated noise, effectively balancing exploration of new solutions and exploitation of known ones to find optimal solutions in these unpredictable environments.[45][46][47]

## 2.4 Automation

The interplay between machine learning and automation forms a crucial thematic thread throughout this thesis, highlighting the significant benefits of integrating these technologies in scientific research. Automation in the laboratory setting is

## 2. INTRODUCTION

transformative, primarily due to the precision, consistency, and tireless operation of machines compared to human counterparts. Machines are inherently less prone to errors and can deliver more accurate and reproducible results, an aspect that is particularly vital in machine learning applications.[48] This accuracy is decisive when multiple variables are simultaneously altered in experiments, as it ensures that uncontrolled influences are minimized, thereby preventing potential misdirection of machine learning algorithms.

The continuous operation of machines, with their ability to work around the clock, further enhances productivity and efficiency in research environments. This thesis emphasizes the implementation of automated systems in the laboratory, primarily focusing on six-axis industrial synthesis robots. Automated setups not only streamline the synthesis processes but also provide a high degree of control and repeatability, essential for the reliable data needed in machine learning models.[49] At the Institute of Functional Interfaces at KIT, we have access to two industrial six-axis robots dedicated to automated syntheses, as depicted in Figure 2.8a. These robots are designed to operate under inert conditions, allowing for the reproducible execution of epitaxial layer-by-layer syntheses aimed at producing homogeneous thin films. Moreover, they are equipped to simultaneously synthesize four samples using the sample holder illustrated in Figure 2.8b, featuring a built-in mechanism for ensuring reproducibility.

However, automation extends beyond the synthesis processes; it also encompasses measurement processes. The integration of machine learning with advanced automation opens new horizons in material science, allowing for more complex, innovative experiments and faster, more reliable outcomes. Additionally, it promotes safety by offering the possibility of delegating the handling of hazardous chemicals to robotic systems. This synergy promises to accelerate discoveries and innovations, making lab automation not just a convenience, but a cornerstone of modern scientific inquiry.[49][50][51]

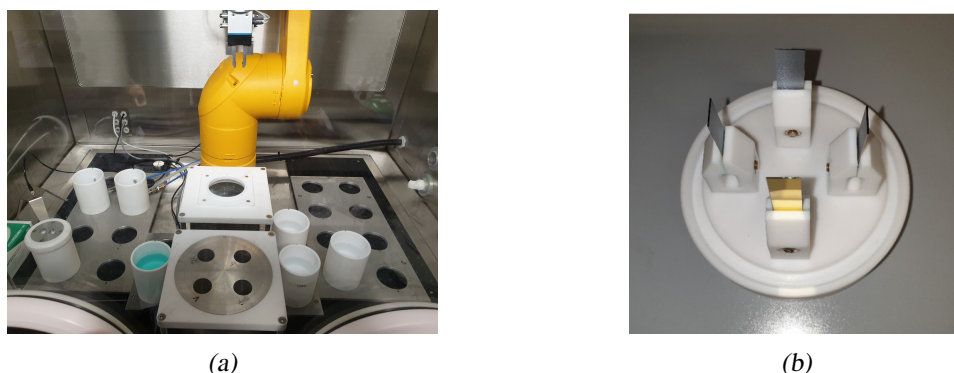


Fig. 2.8: a) Photograph of the robotic set-up at the Institute of Functional Interfaces at KIT.[Adapted and reproduced from [52] - Manuscript accepted for publication] b) Exemplary sampleholder for the robotic set-up, capable of carrying four samples per synthesis.



## 2.5 Research Data Management

To advance the transition from automated, machine-learning-guided synthesis to fully autonomous synthesis, a crucial component is necessary: effective data management. This involves not only documenting the parameters and conditions used in each synthesis but also organizing all characterization and evaluation data. In essence, successful fully automated machine learning optimizations, including automated practical experiments, necessitate well-curated, machine-readable data. Consequently, there is a need for an interface that connects synthetic output and input for new experiments.

To facilitate this, we utilize Chemotion, a data management system specifically designed for chemical data.[53] Chemotion offers two key functionalities. Firstly, it features an electronic lab journal (ELN), which provides chemists with all the necessary tools for tracking their experiments, thereby eliminating the tedious task of later transferring handwritten data into electronic formats. Secondly, it includes a repository that is directly linked to the ELN. This repository enables the publication of data in accordance with the FAIR principles established in the last decade. These principles are depicted in Figure 2.9.

Adhering to these principles not only brings the advantage of easily publishing successful experiments, but also promotes curating and publishing failed or sub-optimal attempts, which is invaluable for various machine learning methods but is regrettably not yet widespread in the literature. Implementing a data management system like Chemotion results in a substantial database that provides added value, as it is accessible to everyone.[54][55]



Fig. 2.9: This figure illustrates the principles of FAIR Data.

# 3

## Methods

### 3.1 X-Ray Diffraction (XRD)

X-Ray Diffraction (XRD), also known as Röntgen diffraction, is a valuable technique used for the analysis of crystalline materials. It provides a wealth of information about the properties of these materials from a single measurement. This includes the quantification of crystallinity, insights into crystal distortion and stress, and determination of crystal orientation. Its primary function, however, is the precise identification of substances.

XRD is a widely favored method of material characterization because it is non-destructive and can be carried out under normal environmental conditions. Moreover, it is particularly well-suited for the analysis of crystalline thin films due to its high resolution, which operates at the atomic level. This means it can provide accurate results even when examining films with only a few layers. The following section offers basic information about X-Ray diffraction, presented at a level sufficient for comprehending the data utilized in this thesis.

#### 3.1.1 Theory

X-ray diffraction patterns originate from the interaction of X-rays with the atoms in a material under examination. In simplified terms, these X-rays are generated when electrons are removed from the innermost shell of an atom due to irradiation from a radiation source. To grasp this phenomenon, a basic shell model is considered in which electrons are arranged in shells around the atomic nucleus. When these electrons are struck by sufficiently energetic electrons, such as  $K_{\alpha}$ , electrons from inner shells are ejected. Subsequently, an electron from a higher shell transitions to fill the vacancy, emitting energy in the form of X-rays. Because these ejected electrons are close to the nucleus, this process is referred to as  $K_{\alpha}$  radiation, also illustrated in Figure 3.1a.

### 3. METHODS

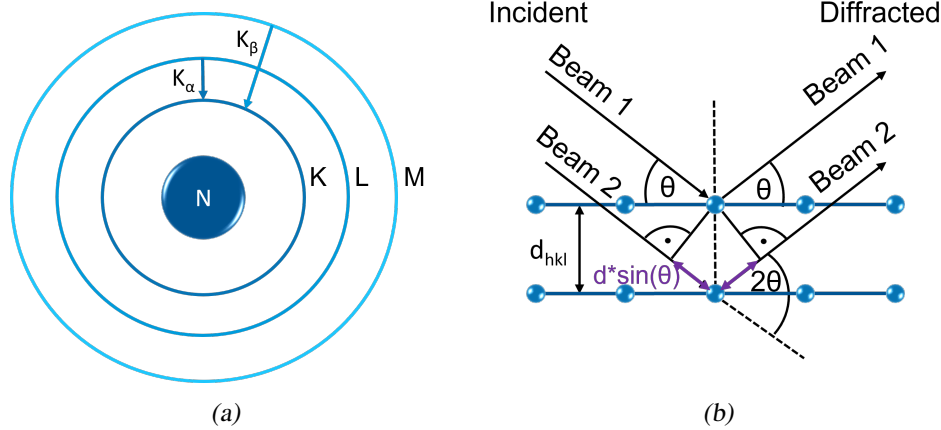


Fig. 3.1: a) This figure illustrates the principles of the origin of X-rays, with K, L, M representing the electron shells around the nucleus N. When electrons from shells near the nucleus are displaced by the impact of sufficiently energetic electrons, electrons from outer shells transition to these vacated positions, emitting X-rays in the process (eg.  $K_{\alpha}$ ). b) Illustration of Bragg's law and the diffraction of X-Rays respectively, with the blue spheres representing atoms in lattice planes.

When X-ray radiation is directed at a crystalline sample, it undergoes diffraction as it interacts with the regularly arranged atoms within the crystal lattice. The relationship between the spacing of lattice planes, denoted as  $d_{hkl}$ , the wavelength  $\lambda$ , and the diffraction angle  $\theta$ , is governed by Bragg's Law:

$$n\lambda = 2d_{hkl} \sin(\theta_{hkl}) \quad (3.1)$$

Here, 'n' is an integer representing the order of reflection.

The incident beam, in Figure 3.1b referred to as "beam 1", and "beam 2", is diffracted at different lattice planes as they exit the crystalline sample towards a detector. Notably, the X-ray beam that penetrates deeper into the sample follows a slightly longer path compared to "beam 1." When the difference in their paths is an integer multiple of the wavelength  $\lambda$ , these two beams undergo constructive interference as they exit the material. This constructive interference occurs only at specific values of the angle, known as the "Bragg angle"  $\theta$ , which represents the angle between the incident beam and the lattice plane.

A detector records the signal resulting from this constructive interference, which is subsequently represented as a peak in a diffractogram. The distinct diffraction patterns, specific to each substance and derived from Bragg's Law, enable the unambiguous identification of a material.

To interpret these reflections or deduce information about the crystal's structure and orientation, Miller indices  $h, k, l$  are employed. A unit cell is conventionally defined and represented by vectors 'a,' 'b,' and 'c' as exemplary depicted in Figure 3.2b. In the case of a cubic crystal system,  $a_0$  denotes the lattice constant. By using the Miller indices  $h, k, l$ , they can be divided into the smallest possible reciprocal sections with consistent ratios, like  $(\frac{a}{h}, \frac{b}{k}, \frac{c}{l})$ . This approach facilitates the description of lattice planes within the crystal. By applying the provided Miller indices

### 3. METHODS

and lattice constant, one can calculate the lattice plane spacing  $d_{hkl}$  according to Equation 3.2.[56][57]

$$d_{hkl} = \frac{a_0}{\sqrt{h^2 + k^2 + l^2}} \quad (3.2)$$

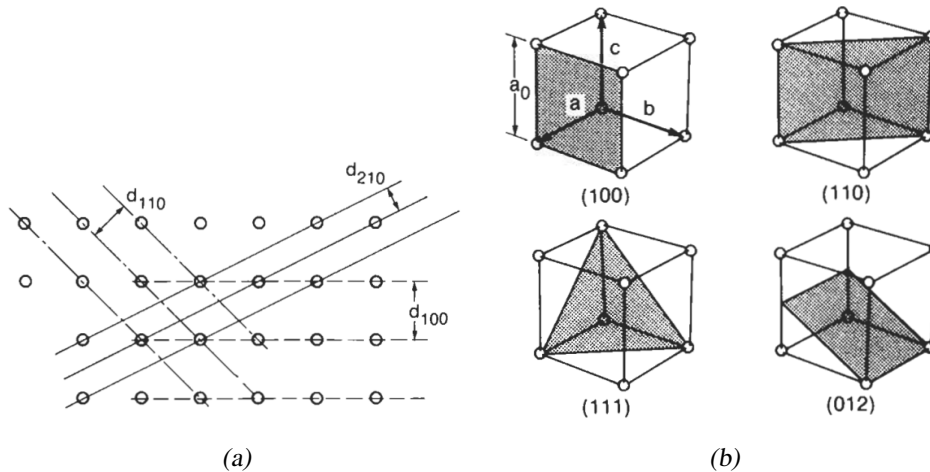


Fig. 3.2: a) Illustration of various atomic layers and the spacing between them within a cubic crystal structure. b) Depiction of Miller indices of different lattice planes. [Reprinted from Encyclopedia of Materials Characterization, Vol 14, Brundle, C. Richard, Evans, Charles A., XRD - X-Ray Diffraction, Page No. 201, Copyright (1992), with permission from Elsevier][56]

#### 3.1.2 XRD Measurement

To elucidate the structural characteristics of solid samples, particularly MOF (Metal-Organic Framework) thin films in the context of this thesis, X-ray diffraction (XRD) measurements are indispensable. A prevalent configuration utilized in diffractometers is the Bragg-Brentano setup, characterized by a  $\theta - \theta$  geometry. In this arrangement, the sample remains stationary while both the column also referred to as tube, and detector orbit the sample, their relative movements precisely coordinated by a goniometer. The principle is illustrated in Figure 3.3 The column houses the X-ray anode, typically fashioned from copper, although molybdenum is also a common choice. X-rays are produced following the principle described earlier (see Figure 3.1a), collimated through slits, and then directed at the sample. The diffracted waves from the sample are subsequently captured by the detector.

### 3. METHODS

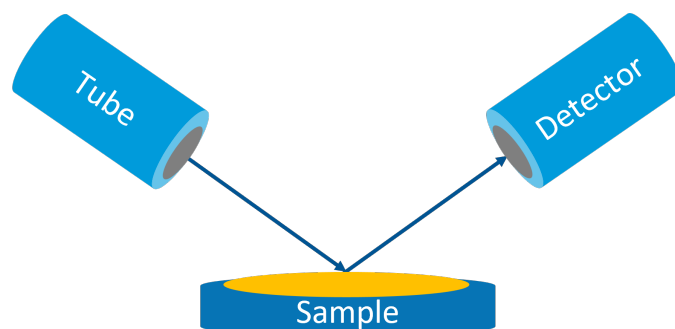


Fig. 3.3: Schematic XRD measurement set-up.

To compile a diffractogram, the intensity of the diffracted X-rays is measured as a function of  $2\theta$ , which varies according to the sample's orientation. Here,  $\theta$  denotes the angle between the incident or reflected X-rays and the lattice planes of the sample, as delineated by Bragg's law (see 3.1) and illustrated in Figure 3.1b.[58][56]

## 3.2 Infrared Spectroscopy

Infrared Spectroscopy is a method within vibrational spectroscopy, a category that also includes Raman spectroscopy. Both techniques explore how radiation within a specific range interacts with molecules, inducing vibrational excitation. The primary difference lies in the mechanism of energy transfer during vibrational state alterations. Infrared (IR) spectroscopy is sensitive to changes in a molecule's dipole moment, while Raman spectroscopy is concerned with molecular polarizability. Both methods are versatile, allowing for both qualitative and quantitative analysis of a wide variety of sample types.

In this thesis, the focus is primarily on the application of IR measurement techniques, particularly in the study of thin films. Consequently, the ensuing chapter will concentrate on furnishing essential information about IR and Infrared Reflection-Absorption Spectroscopy (IRRAS) measurement techniques. This knowledge is crucial for understanding the research and analyses presented in the subsequent chapters.[59]

### 3.2.1 Theory

The essence of IR spectroscopy lies in its ability to detect absorption bands within the IR spectra, which are directly correlated to the characteristic vibration and rotation bands specific to each molecule. These bands provide a unique molecular fingerprint, allowing for the identification and analysis of various substances. This phenomenon is intricately modeled through the concepts of harmonic and anharmonic oscillators, which serve as the mathematical foundation for understanding how molecules interact with IR radiation.

### 3. METHODS

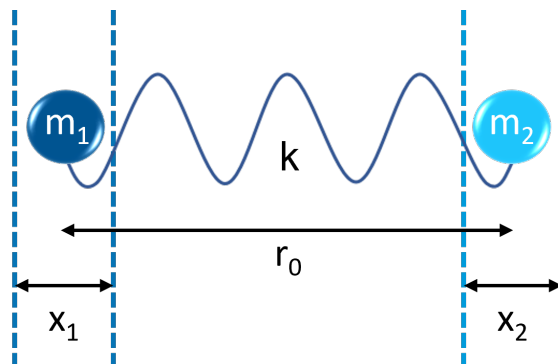


Fig. 3.4: Depiction of a two-atomic molecule simplified as a mechanical model comprising two masses linked by a spring.

Visualizing vibrational oscillation involves examining a two-atomic molecule composed of masses  $m_1$  and  $m_2$ , interconnected by a spring with a spring constant  $k$ . This constant represents the bond strength. Figure 3.4 illustrates the equilibrium or rest position of the molecule's atoms, around which oscillation may occur. The spring constant  $k$  is indicative of the binding strength within Hooke's law, expressed as  $F = -k * \Delta r$ , where the force  $F$  is proportional to the displacement  $\Delta r$  from the equilibrium position when the molecule is either stretched or compressed, with  $\Delta r = x_1 + x_2$ .

Assumptions of smooth and regular oscillation, characteristic of a harmonic oscillator, are depicted in Figure 3.5a. Energy levels, indexed by quantum numbers ( $n = 0, 1, 2, \dots$ ) maintain equidistant spacing and adopt only discrete values. The energy associated with oscillation is computed by:

$$E_n = \left( n + \frac{1}{2} \right) * h * \nu \quad (3.3)$$

Here,  $h$  signifies the Planck constant, and  $\nu$  represents the oscillation frequency. Absorption of a specific energy quantum results in the elevation of the molecule to a higher energy level. According to the harmonic oscillator model, the transitions are limited to  $\pm 1$ , which corresponds to the energy difference between neighboring levels when a photon is absorbed or emitted. Despite the theoretical infinitude of energy levels, suggesting perpetual oscillation without dissociation, the energy levels increase in magnitude with higher quantum numbers.

In contrast, the anharmonic oscillator model, depicted in Figure 3.5b, offers a more realistic approximation of molecular behavior, particularly in how nuclei repel each other more strongly than predicted by the harmonic (parabolic) model as they approach one another. This model suggests the possibility of molecular dissociation at certain energy levels. This model's potential curve conforms to the Morse potential, thereby rendering the energy levels non-equidistant. Energy calculation adheres to the Morse equation:

$$E_n = \left( n + \frac{1}{2} \right) * h * \nu - \left( n + \frac{1}{2} \right)^2 * h * \nu * x \quad (3.4)$$

Within this equation,  $x$  delineates the anharmonic constant.[60][61][59]

### 3. METHODS

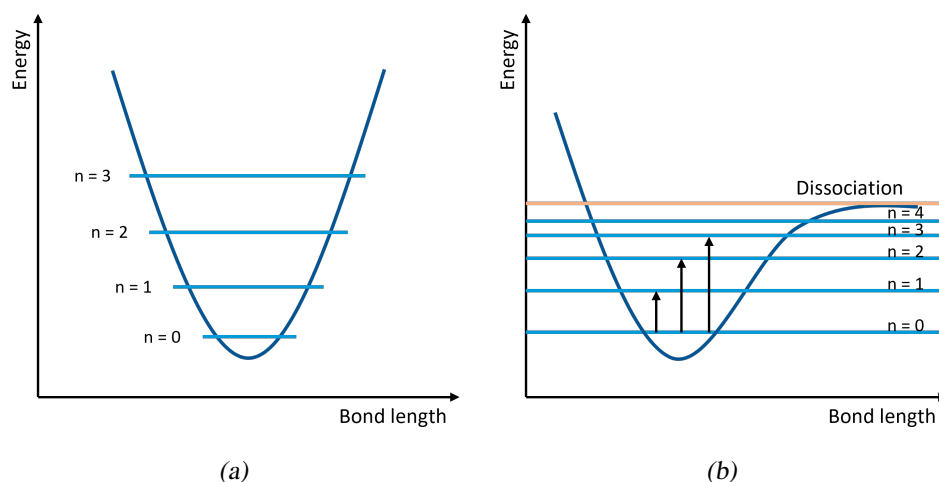


Fig. 3.5: a) Model of the harmonic oscillator. b) Model of the anharmonic oscillator.

#### 3.2.2 IRRAS Measurement

Infrared Reflection-Absorption Spectroscopy (IRRAS) is tailored specifically for thin film analysis, necessitating a unique measurement setup that deviates from traditional infrared spectroscopy methods. In standard setups, the detector is typically placed behind the sample to directly measure absorption. However, IRRAS utilizes a reflective substrate positioned beneath the sample being studied. This configuration is essential as the infrared radiation reflects off the substrate and passes through the sample twice - once before and once after reflection - enhancing the sensitivity of the measurement. This double pass of the IR radiation, crucial for analyzing thin films, results in distinct absorption characteristics being more prominently featured. The substrate and sample are strategically oriented in the measurement chamber to facilitate this reflective process, allowing for precise analysis of the sample's absorption properties based on the intensity variations of the reflected IR radiation.

Prior to the actual analysis in IRRAS, it's common practice to conduct a one-channel measurement on a blank substrate. This step is essential for recording a reference or background spectrum. Following this, the specimen itself is measured, typically within the MID-IR range, spanning from  $4000\text{ cm}^{-1}$  to  $200\text{ cm}^{-1}$ . This range is chosen to capture the most informative spectral data for the analysis of the sample's properties.[62]

### 3.3 Scanning Electron Microscopy (SEM)

The scanning electron microscope (SEM) is a powerful imaging technique that offers the capability to examine surfaces with significant magnification and high depth of field. This method not only provides excellent material contrast and high resolution, making it invaluable in scientific research, but it is also adept at applications like damage analysis, including the characterization of fracture behavior and patterns in various materials. SEM creates magnified images that unveil microscopic details about the size, shape, composition, crystallography, and a range of physical and chemical properties of a specimen. However, it's important to note

### 3. METHODS

that this method can have a destructive impact on the samples, with the extent of this effect varying depending on the material being analyzed.[63]

#### 3.3.1 Theory

Scanning Electron Microscopy functions on the principle of generating a beam of electrons with significant energy, usually ranging between 0.1 and 30 keV. This beam originates from an electron source and is subsequently refined and directed using a combination of apertures, lenses, and electromagnetic coils. As this electron beam methodically scans the specimen in a detailed raster pattern, it interacts with the material, resulting in the emission of two distinct types of electrons. The first type, backscattered electrons (BSEs), are characterized by retaining a large portion of their initial energy following scattering processes within the specimen's atomic structure. The second type, secondary electrons (SEs), are liberated from the specimen's surface, carrying significantly lower energy levels, generally in the range of 0 to 50 eV.

In the measurement phase of SEM, the emitted electrons – both BSEs and SEs – are captured using specialized detectors. The Everhart-Thornley detector is commonly employed for this purpose, capable of registering both types of electrons, alongside dedicated detectors specifically for BSEs. These detectors convert the electron interactions into digital signals, which are then meticulously mapped onto a computer display, each signal corresponding to a pixel in the generated image.[63] [64]

#### 3.3.2 SEM Measurement

The operation of SEM necessitates a highly controlled environment, especially under high vacuum conditions, to mitigate scattering effects and improve image resolution. In the case of insulating specimens, applying a conductive coating, typically achieved by sputtering them with metals such as platinum or gold, is crucial to prevent the accumulation of electrical charge.

For the measurement process, the specimen is positioned within the SEM chamber, which is then evacuated to achieve the necessary vacuum level. Once the ideal vacuum state is reached, the electron beam is activated and directed towards the sample with a probe of approximately 1 to 10 nm. The resulting image's quality is further enhanced by meticulously adjusting parameters such as the position of the sample, magnification levels, acceleration voltage, and exposure duration.[65]

### 3.4 Ellipsometry

Ellipsometry, an advanced optical measurement technique, explores the characteristics of materials and thin films by analyzing the changes in polarized light upon reflection or transmission. Operating mainly within the visible and ultraviolet spectrum, ellipsometry boasts several key advantages. Firstly, it is non-destructive, preserving the integrity of samples. Additionally, it offers rapid data acquisition, providing results in mere seconds.



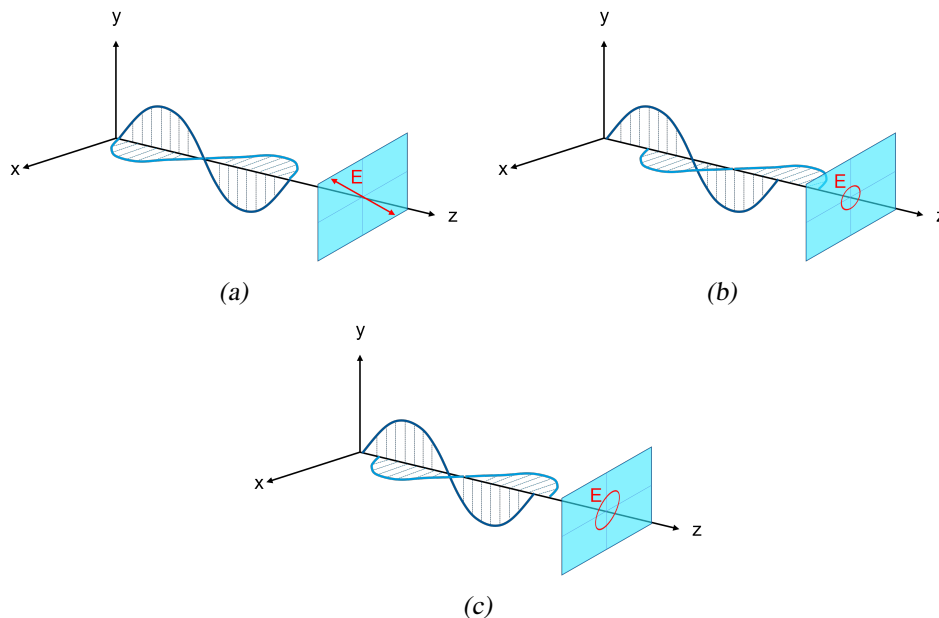
### 3. METHODS

One of ellipsometry's standout features lies in its ability to reveal crucial material properties, including the optical constants refractive index ( $n$ ) and extinction coefficient ( $k$ ). Moreover, it excels at determining parameters like thickness and surface roughness. However, it's important to note that ellipsometry's indirect approach does introduce complexity, necessitating the use of intricate optical models for precise characterization. [66]

To provide a quick overview, the principles of ellipsometry are presented in the upcoming chapter, concerning its measurement technique and the process of data analysis, involving the construction of precise models and their fitting to experimental data.

#### 3.4.1 Theory

The term "ellipsometry" originates from the concept of elliptically polarized light used in its measurements. In this process, the amplitude of transverse light waves, which propagate in the  $z$ -direction, is split into orthogonal components along the  $x$  and  $y$  axes as illustrated in Figure 3.6. These components can also be conceptualized as waves with electric fields oriented perpendicularly to the direction of the wave's transverse motion. When the  $x$  and  $y$  components are in phase, leading to complete constructive interference, the resultant light is linearly polarized (See Figure 3.6a). Conversely, if there is a phase difference of exactly 90 degrees between these components, the light becomes circularly polarized (See Figure 3.6b). However, the most frequently encountered scenario is an intermediate state, where the polarization is neither purely linear nor purely circular, resulting in elliptically polarized light (See Figure 3.6c).



*Fig. 3.6:* Illustration of transverse light propagating in the  $z$ -direction, decomposed into components along the  $x$ - and  $y$ -axes for three scenarios: a) linearly polarized light, b) circularly polarized light, and c) elliptically polarized light. [The three Subfigures are inspired by [67].]

### 3. METHODS

To gather insights into the material properties of a sample, ellipsometric measurement hinges on detecting the changes in polarization between the incident light and the light that is re-emitted or reflected from the sample.

In ellipsometric measurements, the key parameters captured are the amplitude ratio, denoted as  $\Psi$ , and the phase difference,  $\Delta$ . These are integral to the equation of the Fresnel reflection coefficient  $\rho$  (See Equation 3.5). This coefficient is derived from the ratio of two specific reflection coefficients,  $r(s)$  and  $r(p)$ , which are characteristic for each layer being measured.

Considering light as an electromagnetic wave, its electric field can be decomposed into two orthogonal, and thus independent, components: one perpendicular (s-polarized) and the other parallel (p-polarized) to the interface.

$$\rho = \frac{r_s}{r_p} = (\tan\Psi)e^{i\Delta} \quad (3.5)$$

#### 3.4.2 Ellipsometry Measurement

Ellipsometric measurements begin with light emitted from a source, which is then passed through a polarizer. This process polarizes the initially non-polarized light. The now polarized light strikes the sample under investigation and reflects off of it. During this interaction, the state of polarization of the light is altered due to interference with the material properties of the sample. The modified light waves are subsequently captured by a polarization analyzer, which essentially functions as a continuously rotating polarizer. This analyzer controls the amount of light that passes through to the detector. The detector then converts the received light signal into an electrical signal. By comparing the initial state of polarization with the state after interacting with the sample, the amplitude ratio  $\Psi$  and phase difference  $\Delta$  are determined. The setup is schematically illustrated in Figure 3.7

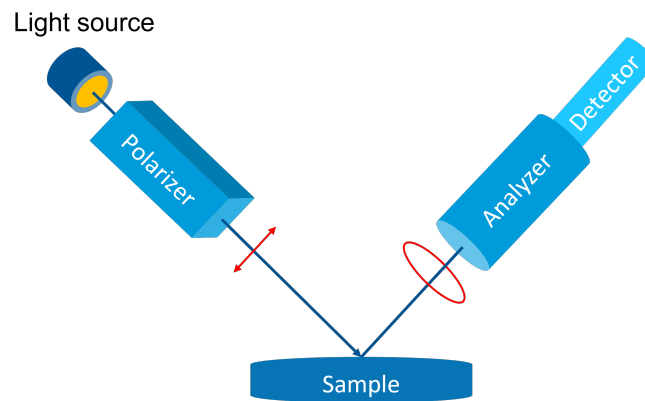


Fig. 3.7: Schematic illustration of measurement set-up in Ellipsometry.

#### 3.4.3 Fitting and Models

The amplitude ratio  $\Psi$  and phase differences  $\Delta$  in ellipsometric measurements are recorded over a defined wavelength range, typically from 370 to 1000 nm. These measurements are taken at various angles of incidence and the corresponding reflection angles. The most suitable angle can be experimentally determined, but for

### 3. METHODS

common substrates like silicon and gold, the appropriate angles are already well-established.

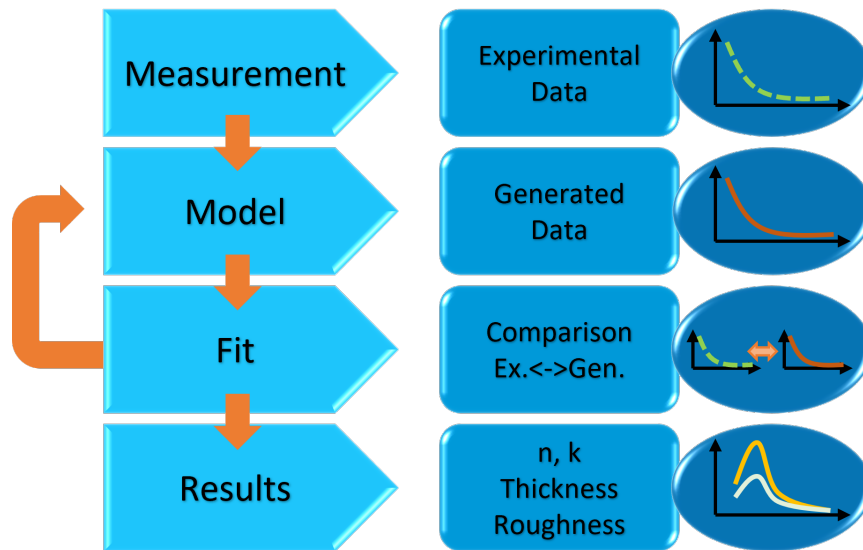


Fig. 3.8: Description of the fundamental sequence of modelling and fitting in ellipsometry for evaluating measurement data. [This Figure is inspired by J.A. Woollam, Software Manual for Spectroscopic Ellipsometers, Figure 3-1, SE Data Analysis Flowchart.[68]]

The fundamental sequence of spectroscopic ellipsometry is illustrated in Figure 3.8. To extract parameters such as film thickness, optical constants ( $n$ ,  $k$ ), or surface roughness from the material-dependent curves of  $\Psi$  and  $\Delta$  post-measurement, a mathematical model of these curves is initially developed. This model is then compared with the physical measurement data. Based on assumptions about the parameters  $n$  and  $k$  as well as the film thickness, these are adjusted until the deviation between the model curve and the measurement data curve is minimized. The closer the model aligns with the actual data, the more accurate the values for the sought-after parameters.

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \quad (3.6)$$

Depending on the specific application, different complexity levels of models are employed in generating these model curves. For transparent organic films deposited on a substrate with a known refractive index, the "Cauchy-layers" model (Equation 3.6) is suitable. In this model, assumptions are made for parameters  $A$ , the amplitude of the Fresnel material index, and  $B$  and  $C$ , which dictate the shape or curvature of the curve per wavelength.  $A$  can initially be approximated as the expected refractive index and then adjusted as needed.

By selecting the appropriate model and making a reasoned initial estimate of the desired parameters, spectroscopic ellipsometry produces realistic and mathematically robust evaluations through a completely non-destructive method.

### 3. METHODS

## 3.5 Dynamic Light Scattering (DLS)

Dynamic Light Scattering (DLS) is a non-destructive technique used to assess macromolecular properties, including the particle size and size distribution of nanoparticles and their diffusion behavior. This method provides insights into potential sedimentation and agglomeration tendencies by measuring the scattering of light by macromolecules in solution.[69]

### 3.5.1 Theory

DLS is a powerful technique that utilizes a laser as a monochromatic light source to illuminate a liquid sample containing molecules or small particles. When this laser light is directed at the sample, it is scattered by the particles within the solution as illustrated in Figure 3.9, a phenomenon known as Rayleigh scattering. The scattered light may interfere constructively or destructively, forming a speckle pattern, which is then analyzed by an autocorrelator to gather information about the particles.

A key aspect of DLS is its sensitivity to the Brownian motion of particles in solution. This motion arises as molecules collide with solvent molecules, leading to their random movement throughout the solution, causing the intensity of the scattered light to vary over time. By measuring these intensity fluctuations at short, regular time intervals, DLS provides detailed information on the dynamics of particle movement. This fluctuation data over time is crucial for determining the size distribution and diffusion behavior of the particles.[70]

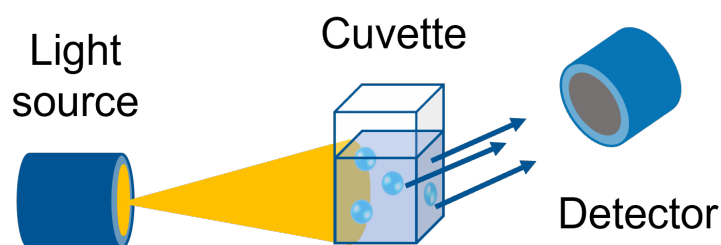


Fig. 3.9: Schematic illustration of measurement set-up in DLS.

### 3.5.2 DLS Measurement

A DLS measurement can be conducted under environmental conditions. Typically, a cuvette containing a diluted sample solution is placed into a measurement chamber that blocks all external light. It's important to consider the light absorbance or scattering effects of the cuvette during the measurement. Additionally, it's crucial that the sample is not overly concentrated, as this can lead to multiple scattering effects, complicating the interpretation of the results. Since the Brownian motion of particles is influenced by temperature—owing to the temperature-dependence of solvent viscosity—the measurement chamber is equipped with temperature control. The sample undergoes a period of equilibration to stabilize temperature before the measurement begins. The process includes several repeated measurements to capture the time-dependent fluctuations of the particles. Additionally, the solvent's

### 3. *METHODS*

refractive index is factored into the data recording process. In the end the auto-correlation function derived from the intensity fluctuations is used to calculate the particle size.[70]

# 4

## Machine Learning Optimizations

### 4.1 MOF thin film Optimization towards Orientation

The orientation of a MOF plays a pivotal role in determining both the quality and the resultant properties of the MOF, making its control essential for any application. Contrary to common assumptions in literature (See Chapter 2.2), factors influencing MOF orientation extend beyond merely the substrate and SAM. The subsequent chapter delves into the intricacies of orientation control. Additionally, it represents the first instance where I have applied the SyCoFinder for a machine learning guided optimization, effectively demonstrating its utility to greatly accelerate the research process.

It is to be noted that the findings and underlying data presented in this chapter have been published in [28]. However, they have been completely rewritten and placed into a broader context for the purpose of this thesis. The data from this optimization have also been published in the Chemotion repository as part of the mentioned publication.[71][72][73]

#### 4.1.1 Introduction

Most classical MOF syntheses, such as solvothermal or microwave-assisted methods, yield MOFs in powder form or small crystallites.[19][74] These powders are inherently non-oriented, with each particle positioned in a macroscopically distinct direction from the others, resulting in an X-ray diffractogram that displays all possible orientations (See blue diffractogram in Figure 4.2a). In contrast, thin films exhibit a different behavior. Since they are anchored to a substrate through functional groups on the surface, the molecules have limited orientations for attachment, enabling the formation of oriented SURMOFs as exemplarily depicted in Figure 4.2a by the red and black diffractograms. However, it is not guaranteed that every thin film will grow oriented spontaneously. If multiple orientations are feasible, which is often the case, the outcome is likely to be a SURMOF with mixed orientations.

To achieve a high-quality SURMOF, characterized by a homogeneously covered surface, absence of island formations, minimal defects in the volume, and uniform material properties throughout the entire thin film, careful consideration of the orientation is crucial.[75] For instance, when integrating SURMOFs into electronic devices, such as sensors or filtration membranes or employing them as photocon-

#### 4. MACHINE LEARNING OPTIMIZATIONS

ductors, it is not only essential to avoid cracks and holes in the thin film, but controlling the oriented growth to ensure consistent performance.[22][76][77] For example, in scenarios involving anisotropic transport of excitons generated through light absorption, the orientation emerges as the decisive factor influencing functionality.[78]

Several previous studies revealed a strong influence of the substrates surface functionalization on orientation, yet they did not identify a singular decisive influencing factor. Furthermore, Liu et al. investigated the impact of dissimilarly functionalized SAMs on the orientation of thin films on gold-coated substrates.[24] Zhuang et al. explored the role of temperature in determining crystal orientation.[79] And additionally, Vello et al. assessed the effect of chain length variations in SAMs on crystal orientation, with the functional group remaining constant.[25] This ambiguity led me to explore synthesis influences beyond substrate and surface functionalization, which I will detail in this upcoming chapter.

Considering the vast array of variables that can affect synthesis outcomes - such as concentrations, temperatures, reaction times, solvents, and more - the resulting parameter space becomes unwieldy, especially when considering the possible interdependence of these variables. Figure 4.1a visualizes a possible idea of such a parameter space, where all variables provide an unspecified contribution. To navigate this complexity efficiently and cost-effectively, I opted for the application of a machine learning tool, known for its capability to accelerate investigative processes particularly under complex conditions. This approach enabled me to vary several parameters simultaneously, significantly enhancing my ability to explore the extensive parameter space compared to classical approaches varying only one parameter at a time. As a MOF system to investigate I chose the HKUST-1, since the starting materials are cheap and easily available.

To implement the selected machine learning tool, the SyCoFinder, as introduced in the opening chapter (refer to Chapter 2.3.2), the initial step involves selecting variables believed to influence the desired outcome. This entails two primary tasks: firstly, defining a clear goal by specifying the target measurements, and secondly, ensuring that only the chosen variables are altered during the synthesis process. This precision is crucial because the optimization algorithm relies on the correlation of variable values with their corresponding quality criteria or fitness values. Any undocumented changes could mislead the algorithm. These considerations led to the adoption of an automated and environmentally resilient synthesis setup illustrated in Figure 4.1b. For this purpose, I utilized a Stäubli six-axis dipping robot under inert conditions, which has demonstrated reliability in producing high-quality MOF thin films via the layer-by-layer method, thus eliminating uncontrolled factors other than the parameters I intended to adjust.

The variables selected partly emerged from the robotic setup (See Figure 4.1b): ultrasonication and spray cleaning times, which can be set for the respective stations in this setup, were chosen to represent two types of cleaning between dips in reaction solutions. Additionally, variations were made in the concentrations of the metal-salt and linker solutions, along with the amount of modulator in the linker solution. Defining these variables necessitated setting upper and lower limits, which were determined based on solubility for the metal and linker solutions, literature values for the modulator (which, in the case of HKUST-1, is water [80]), and the physical capabilities of the synthesis setup regarding ultrasonication and

#### 4. MACHINE LEARNING OPTIMIZATIONS

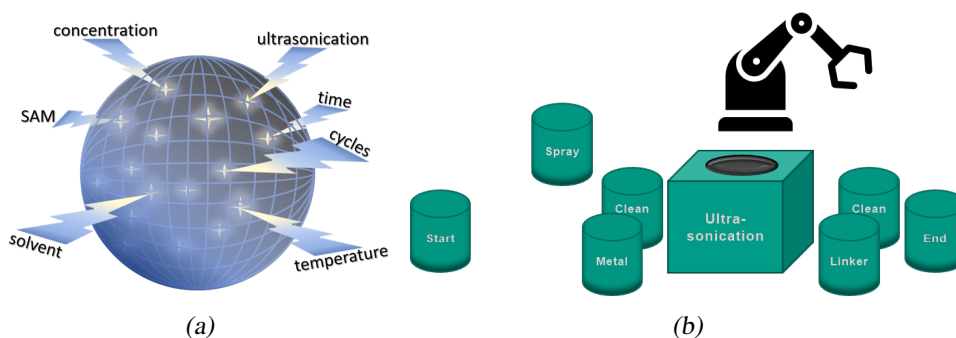


Fig. 4.1: a) Illustration of the parameterspace built from several parameters in which the algorithm is to search the optimum, while effects on the synthesis outcome are both unknown and potentially intersecting when varying parameters simultaneously. [Reproduced from [28]] b) Schematic illustration of the robotic set-up for the dipping layer-by-layer synthesis under inert conditions. [Reproduced from [28]]

spray cleaning times. It is also important to mention that throughout the entire optimization process, I consistently used the same type of substrates - gold-coated silicon wafers - and the same SAM, MHDA.

The final preparatory step for the SyCoFinder was to decide on the number of experiments per generation, which I set at ten. The SyCoFinder then generated a first batch of ten parameter combinations, primarily to test the boundaries of the variable space by exploring edge cases. I conducted all recommended experiments and evaluated them using X-ray diffraction measurements, in line with my predefined objectives.

The goal, motivated by an interest in controlling the orientation of the MOF, was to produce a highly crystalline HKUST-1 SURMOF with a singular phase and a distinct orientation in the [111] direction. This orientation was chosen to challenge the machine learning method, considering that HKUST-1 SURMOF typically favors forming in the [100] direction on MHDA-SAM on a gold substrate. To visually illustrate these two particular orientations, representative examples are presented in Figures 4.2a and 4.2b providing exemplary XRD patterns and corresponding SEM recordings.

My defined objectives however, were based on the measurands - also referred to as quality criteria - obtained from X-ray diffraction, specifically crystallinity, phase purity, and the [111] orientation. To quantify the success of each experiment, I converted these metrics into a fitness value ranging from zero to one using a designated fitness formula (refer to Equation 4.1). These parameter combinations, along with their corresponding fitness values, were then fed back into the SyCoFinder for the subsequent phase of machine learning, utilizing the genetic algorithm. Mirroring evolutionary principles, as detailed in Chapter 2.3.3, the SyCoFinder generates a new, refined set of parameter combinations guided by the initial fitness values provided.

These newly suggested syntheses are then to be executed and assessed. This iterative process can be continued until satisfactory outcomes are achieved, with the definition of "satisfactory" varying according to the user's definition. In my scenario, I determined that reaching a completely pure [111]orientation was the pri-



## 4. MACHINE LEARNING OPTIMIZATIONS

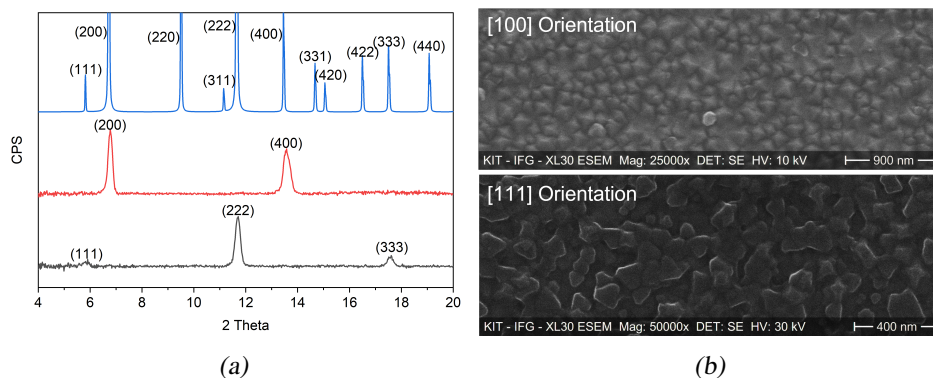


Fig. 4.2: a) Comparative X-ray diffractograms showcasing simulated bulk-MOF (in blue) versus HKUST-I SURMOFs grown with orientation in the [100] direction (in red) and the [111] direction (in black). [Reproduced from: [28]] b) top: a representative SEM image of an HKUST-I SURMOF oriented in the [100] direction; bottom: a representative SEM image of an HKUST-I SURMOF oriented in the [111] direction. [Adapted and reproduced from: [28]]

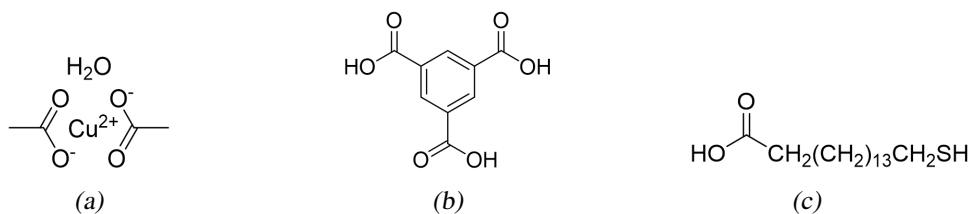


Fig. 4.3: This figure depicts the structural components of HKUST-1. a) The metal-salt providing the nodes; copper(II) acetate. b) The linker molecule forming the connection bonds between the metal centers; trimesic acid (BTC). c) The SAM which is the connection compound between surface substrate and first SURMOF layer; MHTDA.

primary objective, with the ambition of achieving 100% purity alongside high crystallinity. Drawing from my experience, I set the threshold for acceptable crystallinity at 80%. Therefore, for a sample to be considered suitable to end the optimization loop, its overall fitness needed to meet or exceed this 80% benchmark. Upon completing the optimization the SyCoFinder allows an analysis of the importance of variables. This step can uncover valuable insights and dependencies within the growth process. In the following chapter, I will elaborate on the experimental procedure, the evaluation methods, and the insights gained from this concluding phase in detail.

### 4.1.2 ML Optimization

#### Set up

For all syntheses, I utilized the previously mentioned six-axis TX-60 robot from Stäubli, which operated within a nitrogen-filled glovebox. The layer-by-layer synthesis process involved sequential immersions of four functionalized substrates into various solutions, following a predetermined sequence (Refer to Figure 4.1b). The standard procedure for one layer-by-layer cycle included:

#### 4. MACHINE LEARNING OPTIMIZATIONS

1. Immersion in metal solution,
2. A first cleaning step involving spray cleaning and/or ultrasonication and/or dip cleaning,
3. Immersion in linker solution,
4. A second cleaning step involving spray cleaning and/or ultrasonication and/or dip cleaning.

This process was repeated for 40 cycles and the respective times are noted in Chapter 7.3. Besides the actually possible sequences of the robots programs emerging from the suggested parameters are more detailed in the Appendix 4. The variables selected for this study, along with their specified limits, are detailed in Table 4.1. The HKUST-1 structure is built from copper(II) acetate as the metal salt and trimesic acid as the linker (See Figure 4.3). I prepared solutions for each component separately, applying the respective concentrations and amounts provided by the SyCoFinder as noted in the appendix in Tables A.1, A.2 and A.3. All syntheses and subsequent measurements for characterization were carried out according to the procedure described in Chapter 7.3.

<b>Variable</b>	<b>Limits</b>
n (Metal salt) [mmol]	0.02 – 10.00
n (Linker) [mmol]	0.02 – 10.00
Modulator (Water) [mL]	0.00 – 40.00
Ultrasonication time [s]	0 – 100
Spraycleaning time [s]	0 - 5

Tab. 4.1: Definition of variables and their respective limits in which they were varied.[Reproduced from [28]]

#### **Fitness Evaluation**

To furnish the SyCoFinder with the necessary fitness value for conducting evolutionary optimization, I devised a fitness formula. This formula is designed to consistently and accurately convert data obtained from XRD measurements into a value ranging from zero to one. This step is crucial to ensure the comparability of the diffractogram data by avoiding user dependent influences in the evaluation.

The principal fitness formula, which combines all criteria into a value ranging between zero and one, is outlined as follows:

$$fitness = fitness(phase\ identity) * fitness(crystallinity) * fitness(orientation) \quad (4.1)$$

The  $fitness(phase\ identity)$  was determined using a straightforward binary criterion. I compared the diffractograms to a simulated HKUST-1 powder diffractogram (See Figure 4.2a) to investigate whether the signal pattern indicated the formation of HKUST-1 without evidence of doubled phases. If this condition was met, I assigned the  $fitness(phase\ identity)$  a value of one; otherwise, it received

#### 4. MACHINE LEARNING OPTIMIZATIONS

a zero. This approach guaranteed that experiments not yielding the desired compound were immediately disqualified from further consideration, as a zero in this criterion would render the overall fitness score zero. However, these unsuccessful attempts are not without value; they provide useful feedback to the genetic algorithm by indicating areas within the parameter space where further exploration is unlikely to be fruitful.

To ensure reproducibility, I included four substrates in each synthesis. I routinely measured three of these substrates and reserved a randomly selected fourth as a backup in case any sample sustained damage. This strategy allowed me to consistently calculate the average from three measurements  $m_1$ ,  $m_2$  and  $m_3$  for each criterion. Consequently, the calculation of  $fitness(phase\ identity)$  can be illustrated as demonstrated in Equation 4.2:

$$fitness(phase\ identity) = \begin{cases} 1 & \text{if } \frac{m_1+m_2+m_3}{3} = 1, \\ 0 & \text{else.} \end{cases} \quad (4.2)$$

The  $fitness(crystallinity)$  is derived as a percentage value from a routine embedded within the Bruker software. This routine calculates the proportion of amorphous material by comparing the global and reduced areas beneath the measurement curve of a diffractogram, as delineated in Equation 4.3. The global area is the integral under the entire measurement curve, whereas the reduced area is the integral after removing the amorphous portion. It's critical to perform this analysis on the diffractogram before background correction; otherwise, the area calculations, particularly for the global area, could be distorted. I determined the respective  $fitness(crystallinity)$  subsequently according to Equation 4.4, receiving a value between zero and one to provide the first quality criterion.

$$Amorphous[\%] = \frac{Global\ Area - Reduced\ Area}{Global\ Area} * 100 \quad (4.3)$$

$$fitness(crystallinity) = \frac{100 - \% Amorphous}{100} \quad (4.4)$$

The determination and evaluation of the orientation posed the most complex challenge, as the formation is not limited to a single orientation but often involves mixed orientations. Consequently, it became necessary to identify a metric capable of quantifying the ratio between the two primary orientations and a possible resulting predominance respectively in HKUST-1: [100] and [111]. Orientations less likely to occur were not considered in this analysis.

To address this issue, I focused on the most dominant signal corresponding to each orientation, specifically the (222) signal for the [111] orientation and the (400) signal for the [100] orientation. I extracted three simulated intensity values for each signal from the literature for simulated powder patterns, which contain all orientations, and constructed a ratio  $v_s$  (refer to Equation 4.5) based on their averages. Subsequently, I formulated the analogous ratio  $v_m$  (see to Equation 4.6) for the measured signals from the thin films. Through a straightforward case distinction formulated in Equation 4.7, I was able to quantify the preferred orientation by comparing the simulated data to the measured data.

#### 4. MACHINE LEARNING OPTIMIZATIONS

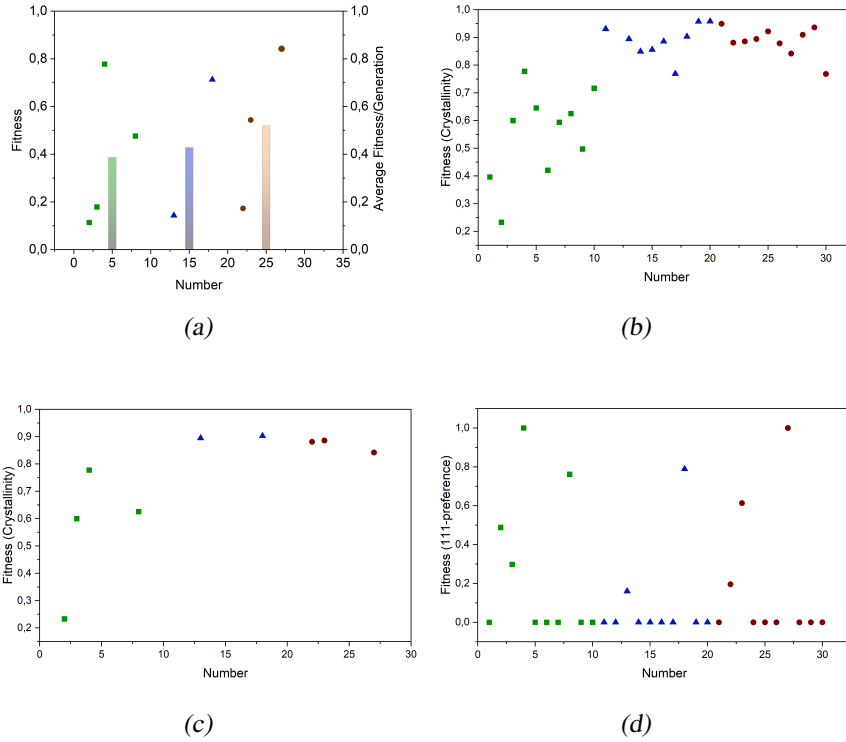
$$v_s = \frac{v(222)}{v(400)} = 5, 19 \quad (4.5)$$

$$v_m = \frac{m(222)}{m(400)} \quad (4.6)$$

$$fitness(orientation) = \begin{cases} 1 - \left(\frac{v_s}{v_m}\right) & \text{for } v_m > v_s \quad \text{resp. } \left(\frac{v_s}{v_m}\right) \leq 1 \\ 0 & \text{for } v_m < v_s \quad \text{resp. } \left(\frac{v_s}{v_m}\right) > 1 \end{cases} \quad (4.7)$$

### 4.1.3 Results and Discussion

#### Evolution



*Fig. 4.4:* This figure is annotated with a consistent legend across all parts: Green squares represent data from the first generation (Diverse Set), blue triangles indicate the second generation (Genetic Algorithm 1), and brown dots correspond to the third generation (Genetic Algorithm 2). a) This panel displays the fitness values of successful trials within the machine learning-guided optimization process, with bars indicating the average fitness of successful trials per generation, color-coded to match the legend. [Reproduced from [28]] b) This section depicts the crystallinity trends among all trials producing HKUST-1, regardless of achieving the desired [111]-orientation, over the generations. [Reproduced from [28]] c) Illustrated here is the evolution of crystallinity in successful trials—those achieving a preferred 111-orientation and thus a fitness value greater than zero—across the three generations. [Reproduced from [28]] d) The progression of achieving the desired [111]-orientation throughout the three generations is detailed here. [Reproduced from [28]]

#### 4. MACHINE LEARNING OPTIMIZATIONS

Figure 4.4a depicts the progression of all successful experiments throughout the entire optimization process, excluding those rated with a fitness value of zero. Each data point signifies an experiment's fitness value, with color and shape coding indicating generation affiliation (green squares for the first generation (Diverse Set), blue triangles for the second generation (Genetic Algorithm 1), and brown dots for the third generation (Genetic Algorithm 2)). The bars, color-coded accordingly, represent the average fitness of successful experiments within each generation. All the foundational data for phase identity, crystallinity, and the percentage of [111] orientation preference, essential for calculating the fitness value, are documented in the Appendix, specifically in Tables A.4, A.5, and A.6. Additionally, the measured X-Ray diffractograms, from which this data was extracted, are detailed in the Appendix Chapter A.1.4.

Analysis of the average bars reveals a notable increase in overall fitness throughout the optimization process, from 0.38 to 0.52 by the last generation. This trend aligns with expectations, as the evolutionary mechanism of the genetic algorithm aims to converge towards a global optimum within the parameter space, thus enhancing the overall quality of generations with each iteration. However, a detailed examination of individual data points unveils a noteworthy fact: in the first generation, the diverse set, four out of ten experiments fall within the evaluation range, appearing widely distributed. In contrast, the second generation appears to exhibit a decline, with only two successful experiments. Yet, a closer look, particularly at the graph depicting the crystallinity of successful experiments (see Figure 4.4b), reveals that the crystallinity, previously varied, has now improved to approximately 90% for both samples, marking a significant advancement from the first to the second generation. I could further underscore this progression by drafting the development of crystallinity across all samples (see Figure 4.4c), indicating that almost all experiments resulted in crystalline HKUST-1, albeit not all meeting my more stringent fitness criterion of preferred [111] orientation, with only nine out of thirty experiments meeting this standard. Nevertheless, the overall crystallinity has seen improvement from the first to the second generation.

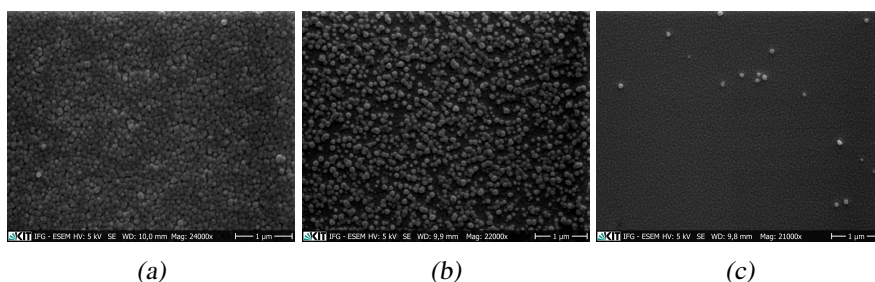


Fig. 4.5: SEM recordings for the best three HKUST-1-thin films in order of evolution over the optimization process. a) Sample 4 (04-DS); Fitness: 0,78. b) Sample 18 (08-GA1); Fitness: 0,71. c) Sample 27 (07-GA2); Fitness: 0,84. [Adapted and reproduced from [28]]

Returning to the overall fitness (Figure 4.4a), the final generation includes three successful experiments, notably one exceptional case (07-GA2) achieving an 84% fitness with a perfect 100% [111] orientation and 84% crystallinity, thus satisfying all my predefined criteria. However, the first generation already showcased a

#### 4. MACHINE LEARNING OPTIMIZATIONS

promising parameter set (04-DS) with a fitness of 0.77, though its crystallinity was only 78%, falling short of the minimum 80% threshold. Besides, opting to stop after the first generation would have precluded further valuable data points essential for understanding the growth process, as I will discuss in the following chapter.

Additionally, I compared SEM images for the three best samples (04-DS, 08-GA1, and 07-GA2) kindly recorded by Dr. Ilona Wagner at the Institute of Functional Interfaces, KIT, with respective fitness values of 0.78, 0.71, and 0.84, to demonstrate the visible improvement from the initially fortunate sample in the first generation to the optimized sample in the last generation (07-GA2) (Refer to Figure 4.5).

#### Importance of Variables

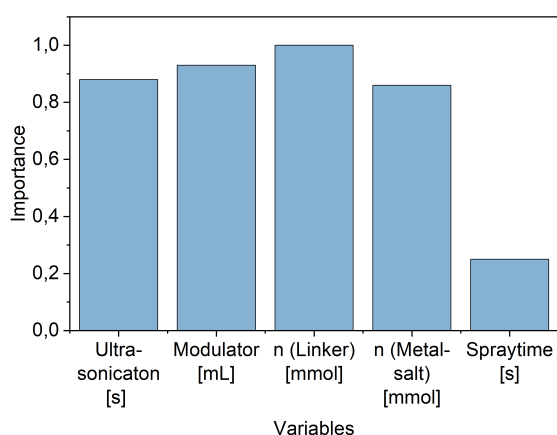


Fig. 4.6: The significance of each variable employed in the machine learning-guided optimization process is depicted for every variable. [Reproduced from [28]]

Once I completed the optimization process, I gathered the parameter sets and their corresponding fitness values from all generations and entered them into the SyCoFinder to assess the significance of the selected variables. This part of the SyCoFinder provides the results in form of a bar diagram as depicted in Figure 4.6. This approach allowed me to quantify the impact of each variable on the target outcome, taking into account not only direct effects but also the interplay between variables. Essentially, it illuminated how strongly each variable influenced the fitness values. The tool determines the importance of these variables by tracking how prediction errors changed when the values of one variable are shuffled while the others are kept constant. For those interested in the specifics of the random forest method employed in this tool called Importance of variables, I recommend consulting the works of Liaw and Wiener, as well as Moosaavi et al.[43][81]

Upon initial examination, the graph appears to suggest that spray time is the lone parameter of minor importance, with all other parameters playing crucial roles. Diving deeper into the data from the nine successful syntheses, which are drafted in Table 4.2, I uncovered further insights into the HKUST-I SURMOF system and the adjustments needed for its orientation. Ranking the parameter sets by decreasing fitness revealed that the paramount importance of ultrasonic cleaning time necessitates its maximization. Drawing on Gu et al.'s findings, I concluded that ultrasonication plays a key role in removing unreacted materials, thereby enhancing the

#### 4. MACHINE LEARNING OPTIMIZATIONS

Tab. 4.2: All successful synthesis parameter sets with their respective fitness values ordered by decreasing fitness.[Reproduced from [28]]

<b>Fitness</b>	<b>Crystallinity [%]</b>	<b>[111]-Orientation [%]</b>	<b>Ultrasonication [s]</b>	<b>Modulator [mL]</b>	<b>n (Linker) [mmol]</b>	<b>n (Metal) [mmol]</b>	<b>Spray [s]</b>
0.84	84	100	100	0.0	7.25	0.02	3
0.78	78	100	100	0.0	10.00	0.02	3
0.71	90	79	100	0.0	9.32	2.77	5
0.54	89	61	25	0.0	6.59	8.92	4
0.48	63	76	60	0.0	10.00	10.00	0
0.18	60	30	0	0.0	5.01	10.00	5
0.17	88	20	82	0.0	9.74	4.95	2
0.14	89	16	25	0.0	6.59	8.92	4
0.11	23	49	100	40.0	10.00	10.00	5

MOF's surface smoothness.[82] Moreover, since ultrasonication cleans each layer during synthesis, I deduced it significantly affects crystallinity.

The individual significance of the metal and linker source amounts was less straightforward. However, combining these insights with an analysis of the top three parameter sets, it became evident that a substantial excess of the linker source over the metal salt is crucial for directing growth towards the [111] orientation. The role of the water parameter was the most straightforward: optimal conditions involve the complete absence of water, a consensus reached in eight of the nine successful syntheses. This observation is supported by Müller et al. and Zhang et al., who noted that water forms solvation shells, slowing the bonding between copper atoms and carboxylic groups.[80][83] Growth towards the [100] orientation occurs with just two simultaneous bonds, in contrast to the three required for an [111] orientation. Therefore, if water hinders bond formation, as previously thought, it logically favors the [100] orientation with only two bonds to form by default. Although these studies suggested that water's presence is essential for high crystallinity, my research demonstrates that adjusting multiple parameters can also yield highly crystalline SURMOFs in the [111] orientation.

#### Increasing the Cyclennumbers

Following the finalised optimization, I explored whether the preferred orientation for SURMOF growth would be maintained beyond 40 growth cycles. The study by Njem et al. provided a clear indication that as deposition cycles increase beyond this number, the influence of self-assembled monolayers (SAMs) diminishes, and the preferential [111] orientation tends to be lost.[84] Challenging this notion, I conducted an additional experiment with 80 deposition cycles, doubling the previous count, under the optimized conditions I had established. By adjusting the

#### 4. MACHINE LEARNING OPTIMIZATIONS

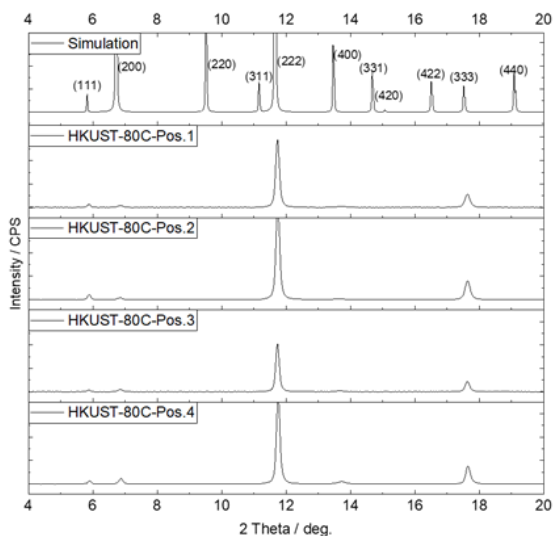


Fig. 4.7: X-Ray diffractograms of four samples from a synthesis conducted for 80 cycles with the best performing parameter combinations from the optimization; This proves that the orientation in [111] direction is preserved also for higher cycle numbers. [Reproduced from [28]]

orientation without taking SAMs into account and solely varying the synthesis parameters, I discovered that even at these higher cycle numbers, the resulting X-ray diffractogram displayed a pronounced [111] orientation (Refer to Figure 4.7), with a clarity of 92% and a crystallinity of 95%. This finding suggests that preferred orientation can be sustained through careful parameter management, even in extended deposition cycles.

#### 4.1.4 Summary

In this chapter, I successfully demonstrated a proof of concept by applying the SyCoFinder to optimize synthesis conditions, achieving the desired goals in material science and expressing the properties of materials in their most effective forms. I conducted a successful machine learning optimization that resulted in the synthesis of a purely 111-oriented HKUST-1 SURMOF with high crystallinity. This was accomplished by varying only the synthesis parameters - concentration of metal, linker, amount of modulator, ultrasonication time, and spray cleaning time - while keeping the substrate and SAM constant. This approach provided concrete evidence that it is indeed possible to control the orientation independently of SAM and substrate, challenging the statement suggested by previous studies that orientation is dictated by a single parameter.

By employing a genetic algorithm to simultaneously vary the five synthesis parameters, I was able to obtain an optimized product in just 30 experiments, a method far more efficient and cost-effective than traditional trial-and-error approaches. The analysis of variable importance not only shed light on the mechanisms and dependencies among the varied variables but also allowed me to challenge previous assumptions that water is necessary for achieving high crystallinity.

An additional experiment conducted with 80 cycles under optimized conditions



## 4. MACHINE LEARNING OPTIMIZATIONS

further demonstrated that, contrary to findings from earlier studies, the preferred orientation is maintained even at higher cycle numbers. This chapter, through its findings and methodologies, contributes valuable insights into the optimization of synthesis conditions and emphasize especially the potential of the SyCoFinder as a practical machine learning method on problems, where only few or even no previous data is available.

### 4.2 MOF thin film Optimization for Gasseparation Application

Following the successful application of SyCoFinder in my initial machine learning project, which yielded significant results, I ventured into a more complex area, investigating the DASBDC pillar-layered SURMOF. This project not only presented a leap in complexity but was also markedly application-oriented. The chosen SURMOF system stands out for its potential in crafting gas separation membranes, a promising avenue for practical deployment.

This upcoming chapter delves into my second optimization project, exploring both the intricate challenges encountered and the insights gained as well during optimizing the SURMOF but also during the subsequent transformation into a SURGEL. It aims to provide a comprehensive overview of the methodologies employed, the optimization process, and the consequent findings, emphasizing the practical implications and the potential of the DASBDC SURMOF/SURGEL system in real-world applications.

The findings and foundational data featured in this chapter were previously published in [30]. For inclusion in this thesis, the material has undergone extensive rewriting and contextual expansion to align with the broader thematic scope. The data from this optimization have also been published in the Chemotion repository as part of the mentioned publication.[85]

#### 4.2.1 Introduction

In 2014, Tsotsalas et al. introduced a three-dimensional, highly porous, covalently bound polymer film, demonstrating its effective use in loading bioactive compounds for drug-release applications [31]. The focus of this study was on an epitaxially grown SURMOF, utilizing copper-paddlewheel nodes and 2,2'-Diazido-4,4'-stilbenedicarboxylic acid (DASBDC) as the primary linker and 1,4-Diazabicyclo-[2.2.2]octane (DABCO) for constructing its pillar-layered structure (Refer to Figure 4.8). The resulting SURMOFs' structure is illustrated in Figure 4.9 from the top and in side view.

The inclusion of two azide groups per linker, while forming the SURMOF, which can serve as functional groups for post-synthesis processing such as crosslinking, renders the system particularly interesting (See Chapter 2.2.2). For the process of transformation into a SURGEL, trimethylolethane tripropiolate (See Figure 4.8d) was used as a crosslinker, integrating into the pores to interconnect the DASBDC linker molecules throughout the structure. Subsequently the metal centers were removed to yield a regular, well-defined polymeric thin film. This general transformation process ensured the formation of a SURGEL with uniformly sized pores,

#### 4. MACHINE LEARNING OPTIMIZATIONS

utilizing the SURMOF as a template.

Considering this uniform pore structure, the possibility to build a network through these pores and the possibility to even transform it into a structured polymer raised the question whether this MOF or one of its subsequent post synthetic states might be applicable as gas separation membrane. Given that structural defects can markedly affect performance, uniformity becomes essential, especially for gas separation membrane applications. Therefore, this highly structured and post-synthetically modifiable material appeared to be promising.

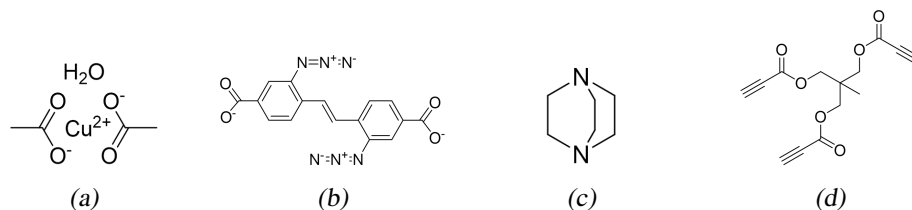


Fig. 4.8: This figure depicts the structural components of the DASBDC-SURMOF and SURGEL. a) The metal-salt providing the nodes of the SURMOF; copper(II) acetate. b) The linker molecule forming the connection bonds between the metal centers; 2,2'-Diazido-4,4'-stilbenedicarboxylic acid (DASBDC). c) The second linker molecule forming the perpendicular connection bonds between the metal centers to form a pillar layered structure; 1,4-Diazabicyclo-[2.2.2]octane (DABCO). d) The crosslinker to interconnect the structure in a postsynthetic procedure; trimethylolethane tripropiolate.

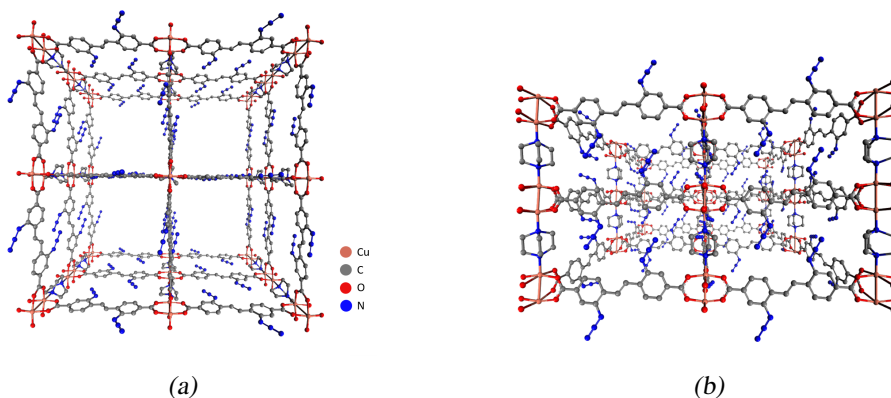


Fig. 4.9: Visualization of the DASBDC-DABCO pillar layered SURMOF structure. [Adapted and reproduced from: [30]] a) top view; b) side view. The graphs were kindly created by Simon Spiegel at the Institute of Functional Interfaces, KIT.

Prioritizing the development of a gas separation membrane, I considered it crucial to first ascertain the high quality of the SURMOF, which would serve as the template for membrane formation. This meant that in terms of the planned machine learning optimization, the main objective would be to achieve a pure phased and high crystalline SURMOF. The principal optimization procedure was the same as in the first optimization:

1. Diverse Set: Calculation of the initial parameter combinations by choosing

#### 4. MACHINE LEARNING OPTIMIZATIONS

relevant variables, respective limits and amount of experiments.

2. Genetic Algorithm: Optimization loop until satisfactory results are achieved.
3. Importance of Variables: Evaluation of the significance of the chosen variables in terms of the defined goal.

However, the synthesis process described by Tsotsalas et al. in 2014 relied on a pump system not enclosed or operated under inert conditions, making it vulnerable to environmental influences and challenging in terms of ensuring exact reproducibility for machine learning methods. In response, I adapted the synthesis setup to utilize my robotic system, but modified it slightly compared to the first optimization to address the fragility of the DASBDC system (Refer to Figure 4.10a). Specifically, harsh cleaning steps such as ultrasonication or spray-cleaning which I had applied to optimize the orientation in HKUST-1, I omitted in favor of gentler dip-cleaning baths after each reactant solution to prevent cross-contamination.

#### 4.2.2 ML Optimization

##### Set up

The variables that I selected for the SyCoFinder to calculate the Diverse Set are detailed in Table 4.3. I varied the concentrations of the Metal and Linker, as well as the Modulator, which in this case was also water. Additionally, I decided to adjust the amount of DABCO in relation to the DASBDC linker.

The final variable essentially covered two aspects: It is defined as a ratio of ethanol used as a solvent. Given that the total volume of the reaction vessels in the robotic setup is fixed at 210 mL due to the immersing depth of the sample holder, this variable could not be simply a volume measurement. Instead, the other implicit component of the ratio is the percentage of methanol required to complete the total volume. My intention was to provide the algorithm with the flexibility to choose between two different solvents or a blend of both. However, since the SyCoFinder did not allow for the definition of categorical variables, I found this ratio to be a suitable solution. I set the amount of experiments again to ten per generation, since this had proven to be a good variable to experiment ratio in my first optimization.

<b>Variable</b>	<b>Range</b>
c (Linker) [mM/l]	0.01 - 1.00
c (Metal) [mM/l]	0.05 - 5.00
Modulator (Water) [ml]	0.00 - 40.00
DABCO [eq of Linker]	0.10 - 2.00
EtOH [%] (Ratio to MeOH)	0.00 - 100

Tab. 4.3: Definition of variables and their respective ranges in which they are to be varied.[Reproduced from: [30]]

#### 4. MACHINE LEARNING OPTIMIZATIONS

For the actual SURMOF synthesis the metal-salt solution and the linker solution containing all, the DASBDC linker, DABCO as second linker and water as modulator were prepared separately, but with the same ratio of ethanol to methanol. I carried out all syntheses according to the procedure described in Chapter 7.4 with the respective parameters per generation provided by the SyCoFinder as noted in the appendix (See Appendix A.7, A.8 and A.9). For this optimization towards a highly crystalline SURMOF I adhered to the gold coated silicon wafers and MHDA as SAM.

I placed the prepared solutions, as well as three dip-cleaning baths per reaction solution in the slightly varied set-up, depicted in Figure 4.10a around the dipping robot. The amount of cycles per experiment was set to 40 and kept constant throughout the whole optimization. Since I had chosen not to vary the cleaning steps in this optimization and to refrain from spray cleaning and ultrasonication, the sequence of one cycle was consistently fixed as described in Chapter 7.4.

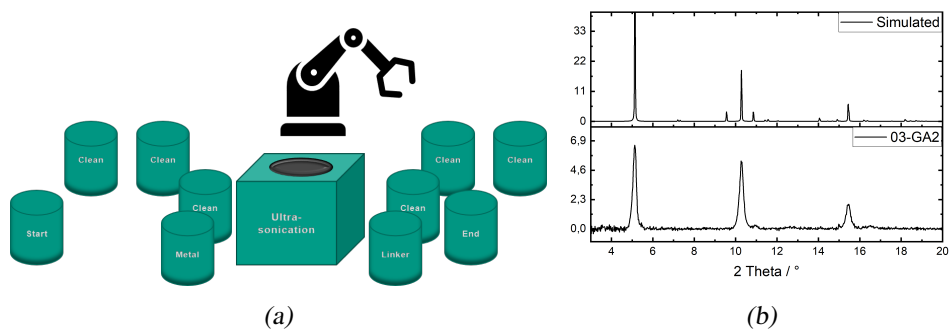


Fig. 4.10: a) Schematic illustration of the applied robotic set-up with three dip-cleaning vessel for each reaction solution. The ultrasonication station is not used in this case. [Reproduced from: [30]] b) Exemplary X-Ray Diffractogram of a measured sample compared to a simulated powder diffractogram of the pillar layered DASBDC SURMOF. [Reproduced from: [30]]

After each synthesis, I conducted X-Ray diffraction measurements on each sample to assess their phase identity, which is exemplary shown in Figure 4.10b. This was achieved by comparing the positions of the signals in the measured diffractogram to those in a simulated powder diffractogram. Samples that exhibited significant shifts in the signals or displayed doubled phases were assigned a zero for this criterion. Conversely, samples that produced diffractograms closely matching the simulated pattern were rated with a one. The actual quality rating was determined based on the crystallinity, which I will explain in the subsequent section.

After generating a fitness value for each parameter combination, I returned the data from the diverse set to the SyCoFinder, enabling it to calculate new, more optimized experiments through the application of the genetic algorithm. I continued this optimization loop until the goal I had set - exceeding 90% crystallinity while maintaining a pure phase - was achieved.

#### 4. MACHINE LEARNING OPTIMIZATIONS

##### Fitness Evaluation

The fitness evaluation again meant the assessment of a fitness formula, to keep the comparability of the results granted. As my defined goal this time was especially the crystallinity, the fitness formula again contained the calculation of the  $fitness(crystallinity)$  according to the built-in routine of the Bruker software. For deeper understanding please refer to Chapter 4.1.2, the underlying equations are again listed below for completeness (See Equations 4.10 and 4.11). Together with the binary exclusion criterion  $fitness(phase\ identity)$  (See Equation 4.9), which I also already applied in the first optimization the main fitness formula resulted to Equation 4.8.

$$fitness = fitness(phase\ identity) * fitness(crystallinity) \quad (4.8)$$

$$fitness(phase\ identity) = \begin{cases} 1 & \text{if } \frac{m_1+m_2}{2} = 1 \\ 0 & \text{else} \end{cases} \quad (4.9)$$

$$Amorphous[\%] = \frac{Global\ Area - Reduced\ Area}{Global\ Area} * 100 \quad (4.10)$$

$$fitness(crystallinity) = \frac{100 - \% Amorphous}{100} \quad (4.11)$$

#### 4.2.3 Results and Discussion

##### Evolution

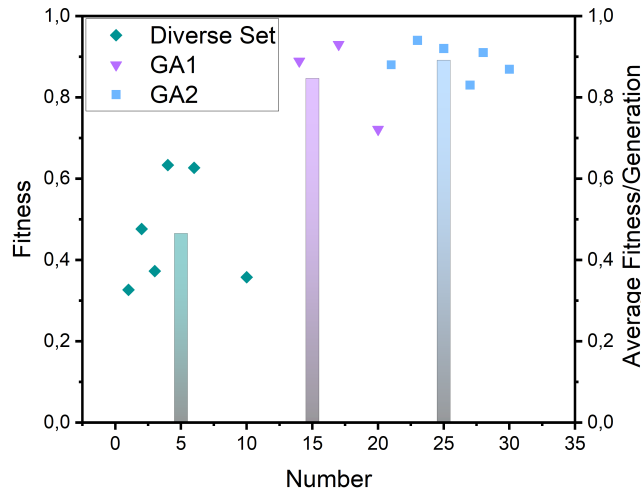


Fig. 4.11: Evolution of the three performed generations throughout the optimization process of the DASBDC SURMOF. [Reproduced from: [30]]

In Figure 4.11, I present the evaluation of the evolution of the machine learning optimization. The data points illustrate the fitness value for each experiment, with green rhombuses representing the diverse set, purple triangles for the first genetic

#### 4. MACHINE LEARNING OPTIMIZATIONS

algorithm (GA1), and blue squares for the second genetic algorithm (GA2). Additionally, the bars indicate the average fitness of the successful experiments per generation, adhering to the same color code. The foundational data used for calculating these fitness values across each generation can be located in the Appendix, specifically within Tables A.10, A.11, and A.12. Moreover, the diffractograms that served as the source of these values are detailed in Chapter A.2.3.

The improvement in fitness values throughout the optimization process is evident, both in the average values of successful experiments and in the individual experiments. Interestingly, the number of experiments deemed qualified for evaluation decreased from the first to the second generation, only to increase again in the third. Moreover, in contrast to the first optimization, where considering only phase purity and crystallinity resulted in 29 out of 30 experiments yielding an HKUST-1 SURMOF, this time only 15 out of 30 experiments were determined to have formed the correct structure. A closer look at the corresponding X-ray diffractograms reveals that many samples were excluded not due to a lack of crystallinity, but rather because of the formation of doubled phases or shifted signals, highlighting the complexity of the SURMOF system.

However, after two generations of machine learning, I was able to identify the best experiment, which attained 94% crystallinity with a pure phase. I thereby considered this optimization to have been successfully completed.

#### Importance of Variables

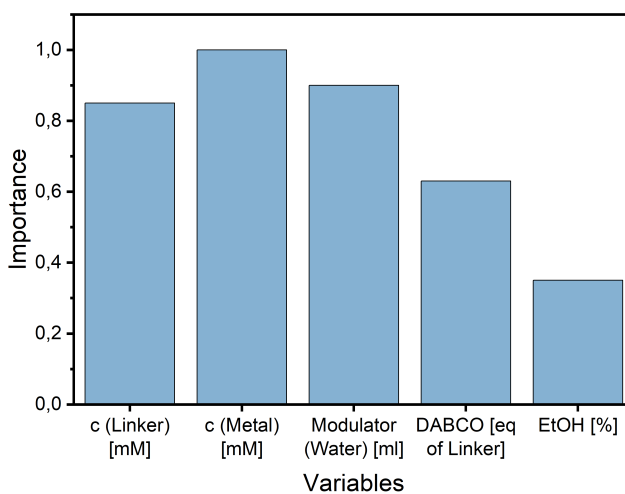


Fig. 4.12: Importance of Variables depicting the significance of the chosen variables in reference to a defined goal. [Reproduced from: [30]]

To understand the impact of the selected variables on the optimization process, I utilized the Importance of Variables tool provided by the SyCoFinder. I supplied it with all the parameter combinations and their respective fitness values from all generations. The significance distribution of the variables is depicted in Figure 4.12. Upon examining the variables and their importance, it becomes immediately clear that all variables, with the exception of the percentage of ethanol used as the reaction solvent, have an importance rating of more than 0.6. This indicates

#### 4. MACHINE LEARNING OPTIMIZATIONS

their significant influence on the successful outcome of the experiments. Notably, the high importance values attributed to the linker and metal salt are fundamental to the system, as they play crucial roles in structure formation and, consequently, crystallinity.

But to gain some more detailed insight I regarded the variables' significance together with the performed parameter combinations ordered by decreasing fitness (See Table 4.4).

When I analyzed the concentration values for each variable individually, it became evident that the "importance" of the linker necessitates a relatively high concentration for success. This observation is based on the fact that almost all experiments achieving a fitness above 0.5 exhibited a linker concentration exceeding  $0.5 \text{ mM L}^{-1}$ , and in many cases, even more than  $0.7 \text{ mM L}^{-1}$ , with  $1 \text{ mM L}^{-1}$  being the maximum observed.

While examining the concentration values for the metal salt, I noticed that they are spread across a broad range without clustering around any specific value. However, there's a strong and approximately linear correlation between the metal salt concentration and the percentage of ethanol, especially for experiments that achieved high fitness. This pattern suggests that even though the experiments were successful with varying ratios of ethanol/methanol mixture - as highlighted by the low significance of this variable - it's plausible that the metal salt concentrations could also exhibit a wide range of values.

In the vast majority of instances, the addition of the modulator, water, exceeded  $20 \text{ mL}$ , often surpassing  $30 \text{ mL}$ , with  $40 \text{ mL}$  marked as the upper limit and  $0 \text{ mL}$  as the lower. Remarkably, only a single experiment conducted without water managed to achieve any success, registering a modest fitness of 0.33, therefore explaining the importance of the modulator variable.

#### 4. MACHINE LEARNING OPTIMIZATIONS

Tab. 4.4: Parameter sets and according fitness values of all successfully rated experiments ordered by decreasing fitness values.[Reproduced from: [30]]

Fitness	c(Linker) [mM/l]	c(Metal) [mM/l]	Modulator (Water) [ml]	DABCO [eq of Linker]	EtOH [%]
0.94	0.70	0.56	40	1.36	24
0.93	0.84	0.05	40	1.43	26
0.92	0.60	2.35	22	0.74	49
0.91	0.78	5.00	36	1.87	100
0.89	1.00	5.00	40	2.00	100
0.88	0.51	5.00	19	1.26	100
0.87	0.92	4.89	15	1.47	91
0.83	0.87	5.00	40	2.00	100
0.72	0.78	5.00	36	1.87	100
0.63	1.00	0.05	40	0.10	50
0.63	0.01	5.00	40	1.05	0
0.48	1.00	5.00	40	2.00	100
0.37	0.01	0.05	20	2.00	100
0.36	0.01	2.52	40	0.10	100
0.33	0.01	0.05	0	0.10	0

When it comes to the additive DABCO, its significance is somewhat lower compared to the concentrations and the modulator, yet it still exceeds 0.6. It's evident that, in this context, 'importance' implies utilizing quantities at the higher end of the variable's limits (refer to Table 4.4). Notably, only three experiments used very low DABCO amounts, around 0.1 equivalents of the linker, and these experiments resulted in lower fitness scores. In contrast, the majority of highly rated experiments incorporated DABCO in amounts ranging from more than one to two equivalents.

#### Transfer of Conditions

For the intended application as a gas separation membrane, the gold-coated silicon wafer substrates posed a significant challenge, as no gas molecule, regardless of its size, could pass through them. In the work of Tsotsalas et al., the authors had produced free-standing membranes using mica substrates, from which the final membrane could be detached and transferred to any other substrate through a membrane transfer process.[31] However, this mechanical release method is prone to damaging the membranes, which is not a good baseline scenario for gas separation applications. This led to the consideration of a porous alternative to directly grow the thin film on top of a substrate, avoiding the damage prone membrane transfer: a gold-coated alumina oxide substrate (See Figure 4.13b). The pores of this material were substantially larger than those of the SURMOF, ensuring they would not obstruct gas flow towards the thin film (See Figure 4.13c).



#### 4. MACHINE LEARNING OPTIMIZATIONS

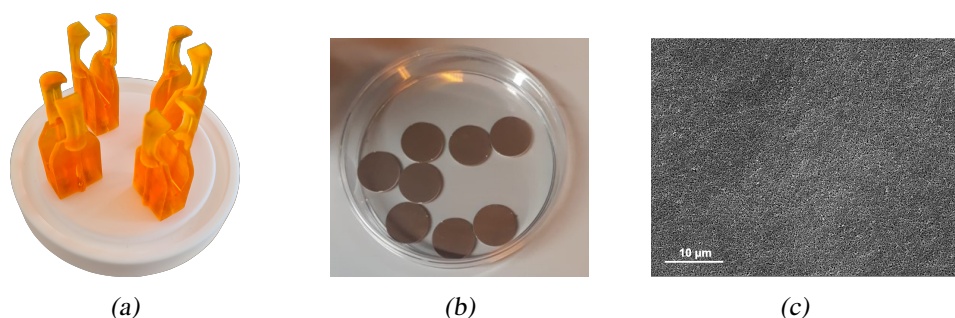


Fig. 4.13: a) 3D printed sampleholder for porous  $\text{Al}_2\text{O}_3$ -substrates. [Reproduced from: [30]] b) Photograph of the gold coated aluminaoxide substrates. c) SEM recording of the porous  $\text{Al}_2\text{O}_3$ -substrates. [Reproduced from: [30]]

These  $\text{Al}_2\text{O}_3$  substrates, being round and approximately one mm thick, required complete coverage for use in gas separation applications. Consequently, the previously used sample holder (See Figure 2.8b), which merely gripped the silicon wafers at one end - leaving that end uncovered - was no longer suitable. The round substrates had to expose as much surface area as possible while still being securely held during the dipping synthesis process. To address this challenge, Dipl.-Ing. Jonas Wohlgemuth and Nina Scheuermann from Prof. Matthias Franzreb's group at IFG, KIT, graciously designed a new sample holder using a 3D-printable material (See Figure 4.13a). The geometry of the samples and the new holder required an adjustment of the immersing depth and therefore respectively the filling volume of the vessels to 260 mL to ensure full coverage of the substrate. This new holder was flexible enough to easily place the substrate inside, yet sufficiently rigid to securely fixate the substrate and inert to any reactant or solvent. However, we encountered an issue with the plasticizer in the material not being resistant to methanol. This led me to decide against using the optimal synthesis conditions, which achieved 94% fitness in a methanol-ethanol mixture, and instead opt for the best parameter combinations using 100% ethanol as the solvent, still yielding 91% fitness. The transition of adapting the conditions optimized for gold-coated silicon wafers to the new porous  $\text{Al}_2\text{O}_3$  substrates was seamless, resulting in clear XRD signals that exhibited almost identical intensity and shape as shown in Figure 4.14.

#### 4. MACHINE LEARNING OPTIMIZATIONS

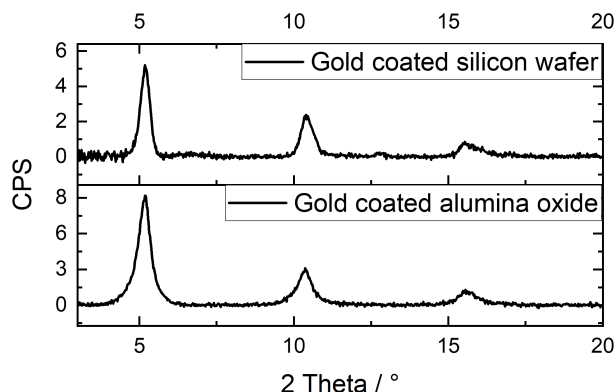


Fig. 4.14: X-Ray diffractograms of the SURMOFs with the best synthesis conditions in pure ethanol applied on a) gold-coated silicon wafer substrate and b) gold coated aluminaoxide substrate.

#### Transformation to SURGEL

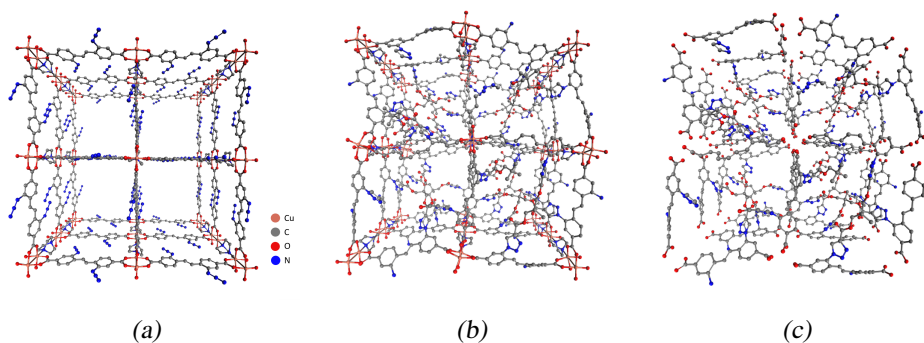


Fig. 4.15: This figure illustrates the full conversion process from a)SURMOF to b) the crosslinked SURMOF to c) the SURGEL where the metal nodes are removed. [All three figures adapted and reproduced from: [30]] The structures were kindly generated by Simon Spiegel at the Institute of Functional Interfaces, KIT.

As outlined in the introduction, the workflow for producing a SURGEL from a SURMOF begins with crosslinking the SURMOF structure. Following the publication of Tsotsalas et al., I carried out the crosslinking procedure by immersing the SURMOF in a heated toluene-crosslinker solution for seven days.[31] The exact procedure is described in Chapter 7.4. I confirmed the success of this postfunctional synthesis through XRD, to ensure the crystalline structure was preserved, and through IRRAS measurements, to verify the effective crosslinking.

The next step involved removing the metal nodes to leave behind a highly structured polymeric thin film membrane, templated by the SURMOF structure. For this process, I adhered to the procedure detailed in the aforementioned publication (See Chapter 7.4), immersing the crosslinked SURMOFs in a solution of EDTA in a mixture of water and ethanol. This step was again verified through XRD, this time to confirm the disappearance of the crystalline structure's signals, and through IRRAS, to investigate on the successful removal of metal ions.

#### 4. MACHINE LEARNING OPTIMIZATIONS

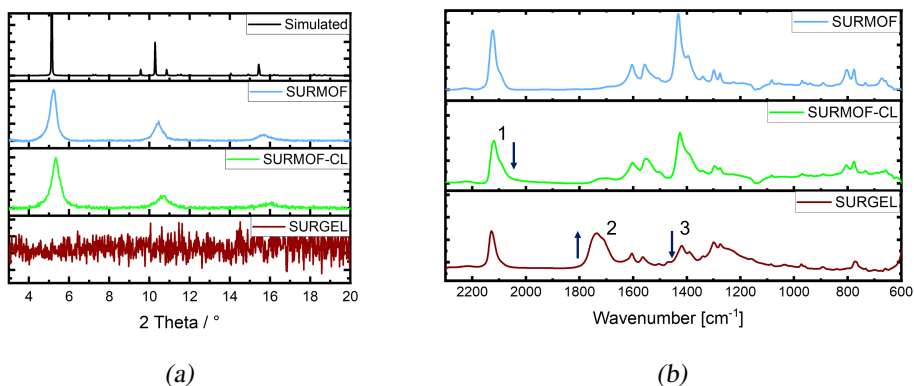


Fig. 4.16: a) Exemplary X-Ray diffractograms showcasing the simulated powder pattern (black), the signals of the SURMOF (blue), the cross-linked SURMOF (green) and the absence of signals in the SURGEL after the removal of the metal ions (brown). [Reproduced from: [30]] b) Exemplary IRRA-Spectra showcasing the signals of the SURMOF (blue), the crosslinked SURMOF (green) and the SURGEL after the removal of the metal ions (brown).

All steps of the process are depicted in Figure 4.15, with representative X-Ray diffractograms shown in Figure 4.16a. Furthermore, the verification of the crosslinking and subsequent transformation into a SURGEL by IRRAS are illustrated in Figure 4.16b, with band assignments provided in Table 4.5. Analysis of these spectra reveals a noticeable decrease in the azide band at  $2122\text{ cm}^{-1}$  during the crosslinking of the SURMOF, indicating the interaction of the previously free  $\text{N}_3$  groups with the crosslinker. The band does not vanish entirely, suggesting the presence of residual free azide groups. This observation aligns with the stoichiometry, considering the crosslinker has three binding sites, whereas the dasbdc linker, thus being present four times per pore layer, offers eight binding sites, leading to the possibility of unreacted  $\text{N}_3$  groups even if two crosslinkers per pore are assumed. Nonetheless, the reduction of the azide band confirms that crosslinking occurred. Upon metal removal, a pronounced band at  $1761\text{ cm}^{-1}$  emerges, attributed to  $\text{C}=\text{O}$  stretching vibrations from protonated carbonyl groups, indicating the liberation of carboxylic acid groups previously coordinated to copper. Concurrently, a decrease in the  $\text{C}=\text{O}$  stretching band of the corresponding deprotonated carbonyl group band around  $1432\text{ cm}^{-1}$  was observed. Together with the clear interpretation of the X-Ray diffractograms, these findings conclusively demonstrate the successful crosslinking and subsequent removal of metal ions.

Tab. 4.5: Peak assignment of the relevant signals in the IRRA-Spectra.

Number	Wavenumber [ $\text{cm}^{-1}$ ]	Peak Assignment
1	2122	$\nu_{as}\text{-N}_3$ Stretching vibration[61]
2	1761	$\nu\text{-C}=\text{O}$ Stretching vibration protonated[61]
3	1432	Sum band of $d\text{-CH}_2$ and $\nu_s\text{-C}=\text{O}$ Stretching vibration deprotonated[61]

### Gasseparation Test

In collaboration with Dr. Nicholas Prasetya from IFG, KIT, we evaluated the gas separation performances of H<sub>2</sub>/CH<sub>4</sub>, H<sub>2</sub>/N<sub>2</sub>, and H<sub>2</sub>/CO<sub>2</sub> across all materials: the SURMOF, crosslinked SURMOF (SURMOF-CL), and SURGEL. Initially, we focused on the evaluation of permeance: The SURMOF-CL consistently exhibited the lowest permeance, the SURMOF the highest, and the SURGEL slightly less than the SURMOF. Regarding individual gases, the membranes demonstrated permeance in order of increasing molecular weight as illustrated in Figure 4.17a: H<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>, CO<sub>2</sub>, indicating that gases with lower molecular weight permeate faster, aligning with the Knudsen Diffusion Model. Consequently, we adopted this model as the baseline for selectivity evaluation: The ideal gas selectivity of H<sub>2</sub> against CH<sub>4</sub>, N<sub>2</sub>, and CO<sub>2</sub> would be 2.82, 3.74, and 4.7, respectively. The observed selectivities for the SURMOF were 2.4, 2.6, and 3, respectively. For the SURMOF-CL, selectivities improved significantly by 21%, 50%, and 40%, to 2.9, 3.9, and 4.2, respectively. For the SURGEL, however, selectivities slightly decreased again by 10%, 15%, and 7%, to 2.6, 3.3, and 3.9.

This variation can be attributed to changes in molecular structure among the different membrane types (Refer to Figure 4.15): The SURMOF, with the largest and least flexible pores of approximately 1 nm in diameter, allows for straightforward gas flow with minimal hindrance, explaining both the high permeance and low selectivity. The SURMOF-CL, with azide groups intercrosslinked across the pores, forms narrower channels for molecular penetration, accounting for its lower permeance but improved selectivity. Additionally, crosslinking could also repair minor pinholes or defects in the SURMOF structure. The decrease in selectivity for the SURGEL, alongside an increase in permeance value (though not matching the SURMOF), follows logically: The conversion of SURMOF-CL to SURGEL removes metal ions, creating more space for gas flow but not as freely as in the SURMOF's pores. The resulting amorphous structure retains the framework of the SURMOF-CL but with greater freedom, particularly for the now unbound linker molecules, leading to slightly reduced selectivity yet substantially better than that of the original SURMOF.

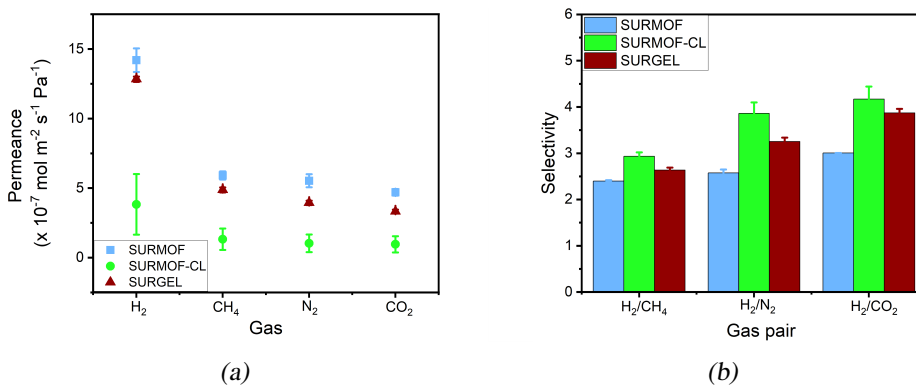


Fig. 4.17: a) Permeance of the SURMOF, SURMOF-CL and SURGEL. [Adapted and reproduced from: [30]] b) Respective selectivities of SURMOF, SURMOF-CL and SURGEL. [Adapted and reproduced from: [30]] The figures were kindly provided by Dr. Nicholas Prasetya.

## 4. MACHINE LEARNING OPTIMIZATIONS

### 4.2.4 Summary

In this chapter, I successfully optimized the crystallinity and stability of the complex, pillar-layered DASBDC SURMOF thin film by employing the SyCoFinder as a machine learning optimization tool. The crystallinity across all generations saw significant improvement, demonstrating the efficacy of the optimization strategy utilized. Remarkably, this achievement was realized within just 30 experiments, despite the simultaneous variation of five parameters.

I then transferred the optimal synthesis conditions containing 100% ethanol from gold-coated silicon wafers to porous alumina oxide substrates. This strategic move enabled the direct growth of the SURMOF on the intended substrate for gas separation applications, effectively circumventing the need for a membrane transfer that could potentially damage the material.

Further, I processed the SURMOFs into crosslinked SURMOFs and subsequently into SURGELS. All three materials were evaluated for their gas separation performance, with the crosslinked SURMOF emerging as the superior candidate in terms of selectivity, particularly for H<sub>2</sub>/CO<sub>2</sub> separation.

This demonstration highlights the efficiency of the SyCoFinder and ML methods at large, even when applied to chemically complex systems. It resulted in the creation of a promising product suited for use in gas separation technologies.

## 4.3 MOF thin film Optimization towards Roughness Reduction

The SyCoFinder has consistently demonstrated its value as a tool for fast and focused machine learning optimizations. In the first two optimization projects, I maintained a ratio of five variables to ten experiments per generation, despite increasing the chemical system's complexity. Now, I aimed to further test the machine learning approach by enhancing the complexity, adjusting the variables to experiment ratio to seven to ten. This adjustment meant that there would be fewer than two variations per variable in the Diverse Set and the subsequent algorithmic iterations. Additionally, the optimization objective was once again application-oriented, focusing on the integration of SURMOFs into electronic devices. In the next chapter, I will explore the effectiveness of the SyCoFinder under these highly complex optimization conditions, with multiple objectives to addressing the challenges associated with the device integration of SURMOFs.

A manuscript based on the findings of this work is currently under preparation. The data from this optimization have been published in the Chemotion repository and will also be referenced as part of the mentioned publication.[86]

### 4.3.1 Introduction

Among the many application areas for MOFs, one has captured significant attention over the last decade: their integration into electronic devices. Whether employed as luminescent and electrochemical sensors for detecting cancer biomarkers, as fiber optic sensors for monitoring water levels in industrial gases, or utilized in digital circuits, field-effect transistors, or mass-sensitive sensors, MOFs have

#### 4. MACHINE LEARNING OPTIMIZATIONS

emerged as a material of considerable interest.[22][87][88] A crucial attribute of MOFs that frequently requires consideration in these applications is surface roughness, which greatly influences the properties of thin films, particularly in optical applications. For instance, the work by Zhi-Ghang Gu et al. demonstrated that reducing surface roughness through ultrasonication cleaning enhanced the transparency of their SURMOF HKUST-1.[82] However, achieving exceptionally smooth surfaces with minimal surface roughness is challenging. We hypothesized that the initial cycles of epitaxially building a SURMOF are critical for attaining a smooth surface. This is because if these initial layers are not formed consistently, island formation is likely to occur, leading to a rough surface in the final product.

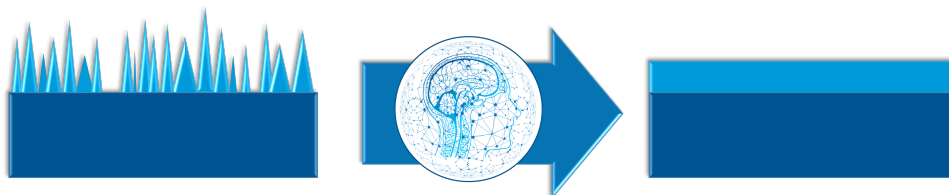


Fig. 4.18: Illustration of the goal to optimize the smoothness of a thin films' surface.

To validate my hypothesis, I embarked on a new machine learning endeavor with the SyCoFinder, which had already demonstrated its value in the previous optimizations. Recognizing that the complexity of this optimization surpassed the capabilities of the previously utilized five variables, I expanded the range to seven variables to thoroughly address potentially influencing factors, as detailed in Table 4.6. This expansion meant that the parameter space became immensely vast, rendering classical optimization methods adjusting one variable after another absolutely impractical.

Drawing from Gu's work, I was aware of the impact of ultrasonication as a cleaning step on SURMOF properties, suggesting that spray cleaning might also play a significant role. Given my intention to revisit HKUST-1, the choice of modulator became pivotal, alongside the concentrations of both the metal salt and the linker solution. With these considerations, five variables were already defined. But to specifically assess the effect of the initial synthesis cycles, I introduced two additional variables to vary the concentrations uniquely within the first three cycles. In an attempt to rigorously evaluate the SyCoFinder's capabilities, I decided to keep the limit of the number of experiments at ten per generation. The objective was to synthesize a pure phased, highly crystalline, and exceptionally smooth SURMOF, thereby undertaking an optimization that simultaneously varied seven variables to achieve three distinct goals. Furthermore, I aimed to deepen my understanding of the initial cycles' impact on SURMOF growth.

#### 4.3.2 ML Optimization

##### Set up

Working on the HKUST-1 SURMOF, constructed from the components depicted in Figure 4.3 and synthesized in an epitaxial manner, allowed me to utilize almost the

#### 4. MACHINE LEARNING OPTIMIZATIONS

identical robotic setup introduced in Chapter 4.1 (See Figure 4.1b), including ultrasonication and spray cleaning steps. Therefore, the possible program sequences also remained equivalent to the first optimizations as shown in Appendix 4. The introduction of two extra variables for the concentrations of metal and linker in the first three cycles necessitated a slight modification in the synthesis organization: The total number of cycles remained at 40, but the concentrations for the initial three cycles were varied independently from the subsequent 37 cycles. To achieve this, I swapped the metal and linker solutions after the first three cycles with the solutions prepared for cycles 4 – 40. To ensure reproducibility, this swap was always conducted immediately after the third cycle concluded, preventing the substrates from drying out between cycles. The complete synthesis procedure is described in Chapter 7.5 and all applied parameter combinations are to be found in the Appendix in Tables A.13, A.14 and A.15.

Tab. 4.6: The seven parameters chosen to be varied and their according ranges and cycles to which they are applied.[Adapted and reproduced from [52] - Manuscript accepted for publication]

Variable	Range	Applied on Cycle Number
Metal salt quantity	0.02 - 6.00 mmol	4 - 40
Linker quantity	0.02 - 10.00 mmol	4 - 40
Amount of water (modulator)	0.0 - 40.0 mL	1 - 40
Cleaning time via ultrasonication	0 - 100 s	1 - 40
Cleaning time via spray-unit	0 - 5 s	1 - 40
Metal salt quantity 2	0.02 - 6.00 mmol	1 - 3
Linker quantity 2	0.02 - 10.00 mmol	1 - 3

The optimization followed the SyCoFinder’s general sequence, as outlined in the introduction chapter 2.3.2 and elaborated upon in the previous optimization chapters 4.1 and 4.2. I set a goal of achieving an overall fitness of 80%, considering phase identity, crystallinity, and surface roughness together. Remarkably, this objective was still met after conducting 30 experiments, or, in other words, two generations of machine learning optimization.

#### Fitness Evaluation

The main fitness formula 4.12 this time included apart from the terms for phase identity, ensuring the origin of the correct compound HKUST-1 with pure phases and crystallinity, both emerging from XRD, another term for the quantification of roughness, which was measured by Ellipsometry.

$$fitness = fitness(phase\ identity) * fitness(crystallinity) * fitness(roughness) \quad (4.12)$$

$fitness(phase\ identity)$  as before is determined as a binary criterion to exclude samples, not or not purely forming HKUST-1. For deeper understanding please refer to Chapter 4.1.2.

#### 4. MACHINE LEARNING OPTIMIZATIONS

$$fitness(phase\ identity) = \begin{cases} 1 & \text{if } \frac{f_1+f_2}{2} = 1 \\ 0 & \text{else} \end{cases} \quad (4.13)$$

Also  $fitness(crystallinity)$  has been employed in the same way in both preceding optimizations and is explained in detail in Chapter 4.1.2.

$$fitness(crystallinity) = \frac{100 - \% \text{ Amorphous}}{100} \quad (4.14)$$

$$Amorphous[\%] = \frac{Global\ Area - Reduced\ Area}{Global\ Area} * 100 \quad (4.15)$$

The calculation of  $fitness(roughness)$  emerged from ellipsometric analysis, focusing on correlating the outcomes of an optical model with actual observed data. This method directly yielded the parameters of  $roughness$  and  $thickness$ , both quantified in nanometers. To emphasize the development towards minimal roughness in comparison to thickness, I computed the ratio of  $roughness$  to  $thickness$  and subtracted this value from one. Consequently, a lower roughness led to a  $fitness(roughness)$  value approaching one, ensuring that roughness was evaluated in proportion to thickness and recognizing the intrinsic link between these two parameters.

To ensure everything was uniform and comparable among all measurements, I adopted a consistent analysis method. This approach required setting a threshold to eliminate unreliable data. A key metric used in this context was the dimensionless Mean Squared Error (MSE), indicative of the congruence between model-generated data and actual measurements, as calculated by the evaluation software during curve fitting. A higher MSE suggests a greater deviation of the experimental surface from the idealized layer model, characterized by sharp interfaces and minimal roughness. Drawing on insights from prior research, I classified measurements with an MSE greater than 20 as unreliable. As a result, this method automatically excluded any measurement with an MSE higher than this value. Furthermore, based on practical insights, I established a cutoff for the roughness-to-thickness ratio at 60%. This criterion was intended to enable the machine learning algorithm to derive insights also from the imperfect results, but not being misdirected by them.

$$fitness(roughness) = \left(1 - \frac{roughness}{thickness}\right) * mse * R/T-ratio$$

$$mse = \begin{cases} 1 & \text{for } MSE < 20 \\ 0 & \text{for } MSE > 20 \end{cases} \quad (4.16)$$

$$R/T-ratio = \begin{cases} 1 & \text{for } \frac{R}{T} * 100 < 60\% \\ 0 & \text{for } \frac{R}{T} * 100 > 60\% \end{cases}$$

### 4.3.3 Results and Discussion

#### Evolution

The progression of overall fitness for all successful experiments, as illustrated in Figure 4.19a, indeed experienced a slight decline on average through the optimization process. However, this trend was counterbalanced by a consistent increase



#### 4. MACHINE LEARNING OPTIMIZATIONS

in the number of successful experiments from generation to generation. This dynamic led to a nuanced outcome: while the broader distribution of experiments contributed to a slight reduction in average values, the rise in experiments meeting the evaluation criteria underscored an improvement in sample quality. Furthermore, it's noteworthy that the best-performing experiment emerged as expected from the final generation.

To comprehend this trend, it is essential to investigate why the first generation yielded only three successful experiments, while Figure 4.19b reveals that, in reality, all thirty experiments produced crystalline HKUST-1. This discrepancy primarily stems from the stringent exclusion criteria based on roughness assessment. In conducting ellipsometry evaluations, a mathematical model is crafted and refined to align with the measurement curve. The more accurately this model fits, the more reliable the derived values for thickness and roughness are. However, roughness significantly impacts the dependability of results, as extensive scattering on rough surfaces can compromise the reliability of ellipsometric measurements for higher roughness values. This means, the higher the roughness, the more adjustment of the model is needed. But to maintain the comparability, required by the ML method, it was crucial to employ the same evaluation model and adjustment process for each sample, rather than tailoring them individually. Custom adjustments would have rendered the evaluations subjective and potentially unsuitable for machine learning optimizations. To navigate these challenges and ensure the reliability of the calculated values, I established two thresholds: one derived from the MSE, which gauges the model curve's alignment with the actual measurements, excluding any that surpassed an MSE of 20; and a second threshold for the roughness-to-thickness ratio at 60%, based on my experience with model reliability.

In the initial generation, as mentioned earlier, the SyCoFinder predominantly explored the edges of the defined parameter space. My selection of variable limits was sufficient to consistently produce HKUST-1 in principle. However, many samples were excluded due to excessive roughness, as indicated by their MSE values, and the unrealistic roughness-to-thickness ratios, resulting in only three experiments being rated successful in the Diverse Set.

Allowing the genetic algorithm to proceed led to five successful experiments in the next generation and six in the following, indicating an implicit improvement across generations as they now fell in the evaluation range.

Reviewing the roughness trends in Figure 4.19c, I observed an already high performing sample in the Diverse Set with a 76% fitness. Yet, the drive to continue stemmed from the fact that the crystallinity at this point was a modest 83%, and three data points were insufficient to conclusively assess the hypothesis regarding the initial three cycles' influence. Remarkably, the best experiment achieved an 81% fitness with 92% crystallinity and a mere 6.52 nm roughness. Given the complexity of optimizing seven parameters simultaneously towards multiple objectives, this outcome is exceptionally noteworthy.

All underlying diffractograms, of which all XRD related fitness values are derived, as well as the ellipsometric values are to be found in the Appendix in Chapter A.3.3 and Tables A.16, A.17 and A.18.

#### 4. MACHINE LEARNING OPTIMIZATIONS

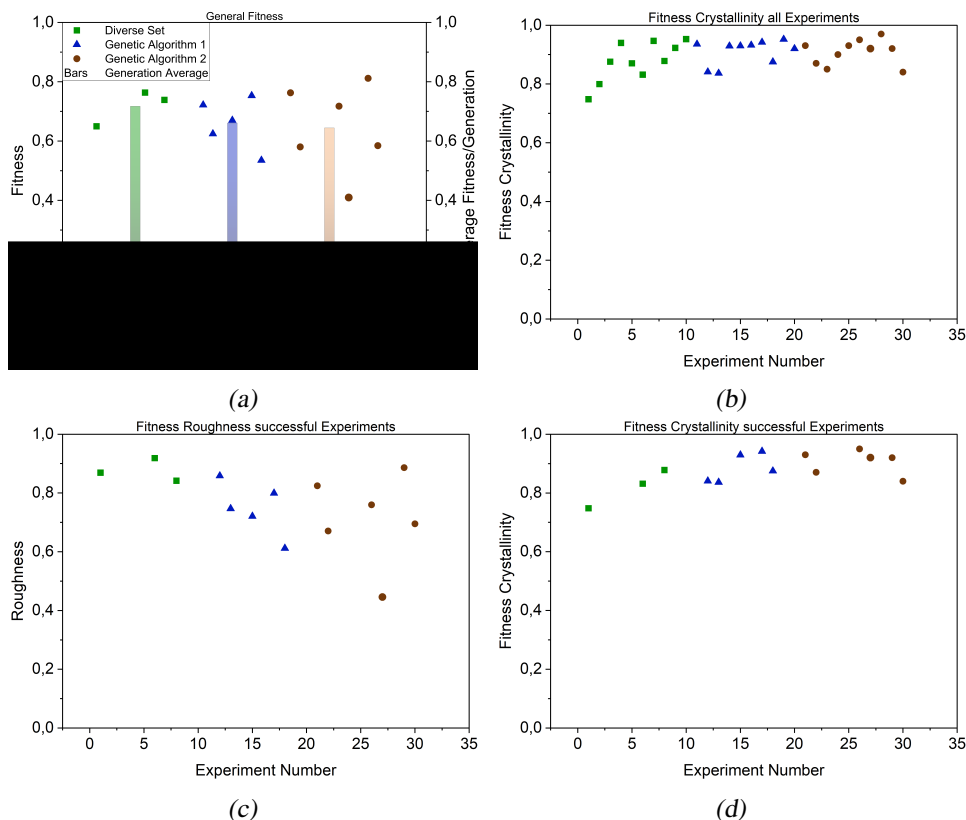


Fig. 4.19: a) Progression of Fitness Values Through Three Generations of Machine Learning Optimization: Diverse Set (Green Squares), First Genetic Algorithm (GA1) (Blue Triangles), and Second Genetic Algorithm (GA2) (Brown Dots). The same color coding applies to figures b), c), and d). Bars indicate the average fitness value of all experiments in each generation, using corresponding colors. b) Fitness Crystallinity for all experiments. c) Fitness for Roughness for all successful experiments. d) Fitness Crystallinity for all successful experiments. [All four Figures: Adapted and reproduced from [52] - Manuscript accepted for publication].

#### Importance of Variables

To assess the relative importance of the variables, the SyCoFinder analyzes data from all completed generations. This evaluation clearly identified, as shown in Figure 4.20a, that the modulator is the most crucial variable, a fact supported by literature for its role in accelerating the formation of crystalline HKUST-1.[80][83] However, this analysis did not deliver the desired insights into the growth mechanism. Given the significantly larger parameter space created this time, encompassing seven variables compared to previous efforts with only five variables, it's evident that the Diverse Set, focusing mainly on the edge cases of variable limits, suggests parameter sets unlikely to represent the optimal conditions. This observation is confirmed by examining the three parameter combinations from the Diverse Set that were rated at all; they are significantly distant from the conditions of the best experiment. However, the stringent exclusion criteria set by ellipsometry resulted in only 14 experiments being deemed successful over all generations. Consequently, the three Diverse Set combinations, representing 20% of these successful experiments, are likely to exhibit a high bias without being optimized. Taking

#### 4. MACHINE LEARNING OPTIMIZATIONS

these factors into account, I recalculated the importance of variables for all experiments from the two optimized generations (GA1 and GA2), as illustrated in Figure 4.20b.

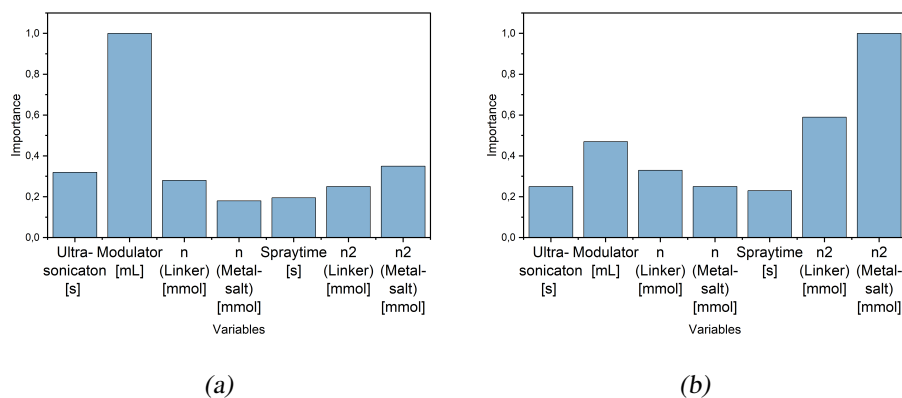


Fig. 4.20: a) Significance of selected variables (n-Linker, n-Metal, Modulator, Ultrasonication Time, Spray Cleaning Time, n2-Linker, n2-Metal) in all experiments, rated on a scale from 0 to 1. b) Significance of the same selected variables of all experiments in GA1 and GA2, rated on a scale from 0 to 1 [Adapted and reproduced from [52] - Manuscript accepted for publication].

Once more, the outcome is clear, identifying the concentration of the metal salt in the first three cycles as the most important variable, closely followed by the linker concentration in these initial cycles. This finding substantiates the hypothesis regarding the significant impact of the initial cycles. The previously dominant importance of the modulator has relatively decreased, yet it remains more significant than the other variables.

These developments can be explained upon examining the suggested parameters of the successful experiments in these optimized generations, which are presented in Table 4.7, arranged by decreasing fitness. Looking at the best three experiments a clear trend is visible: for both cycles 1-3 and cycles 4-40 the linker concentration is significantly higher than that of the metal salt. However, when comparing the initial cycles to the subsequent ones in the synthesis, the concentrations of both the metal salt and linker are higher in the initial cycles. This suggests that higher concentrations in the initial cycles aid in more thoroughly covering the surface with base layers, thus preventing island formation.

In conclusion, this analysis validated the hypothesis about the significant influence of the initial cycles, though it focused solely on concentration variables. For future optimizations, it would be equally important to explore the impact of other parameters during these early stages, such as variations in cleaning methods, to fully understand their potential effects on synthesis outcomes. In general, for an optimization as intricate in its evaluation criteria as this one, conducting more than 30 experiments could be beneficial to deepen our understanding of the underlying mechanisms. Nevertheless, these 30 experiments sufficed to achieve an optimized SURMOF product.

#### 4. MACHINE LEARNING OPTIMIZATIONS

Tab. 4.7: Parameter sets and according fitness values of all successful experiments in GA1 and GA2 ordered by decreasing fitness values.[Adapted and reproduced from [52] - Manuscript accepted for publication]

<b>Fitness</b>	<b>Ultra-sonication [s]</b>	<b>Modulator (Water) [mL]</b>	<b>n (Linker) [mmol] Cycle 4-40</b>	<b>n (Metal) [mmol] Cycle 4-40</b>	<b>Spray-cleaning [s]</b>	<b>n2 (Linker) [mmol] Cycle 1-3</b>	<b>n2 (Metal) [mmol] Cycle 1-3</b>
0.81	36.00	7.81	8.29	3.75	3.00	10.22	4.28
0.76	45.00	7.75	7.74	1.41	0.00	11.64	5.46
0.75	0.00	18.15	6.98	3.76	3.00	9.56	6.38
0.72	0.00	0.00	10.00	0.02	5.00	7.34	6.00
0.72	0.00	24.30	7.18	0.03	4.00	7.70	1.15
0.67	12.00	2.93	10.00	0.03	5.00	6.65	6.00
0.62	81.00	0.00	2.52	6.00	5.00	7.93	2.97
0.58	58.00	2.79	4.77	6.00	5.00	8.22	1.04
0.58	10.00	7.24	2.72	1.07	5.00	8.41	1.95
0.54	63.00	22.61	0.05	0.05	2.00	0.33	1.59
0.41	5.00	5.26	9.91	2.57	4.00	9.01	6.34

#### 4.3.4 Summary

In this chapter, I rigorously tested the capabilities of the SyCoFinder by simultaneously adjusting seven variables towards achieving three distinct goals, with only ten experiments in each generation. Despite the complexity of this task, the optimization was successful, resulting in an optimized HKUST-1 characterized by low surface roughness and high crystallinity, accomplished within just 30 experiments. This process confirmed the hypothesis that the initial synthesis cycles play a critical role in SURMOF growth. However, this exploration was limited to assessing the impact of concentration variations in the first three cycles. Future optimizations could benefit from examining other parameters during these early stages to gain a more comprehensive understanding. Overall, the SyCoFinder demonstrated its immense value, significantly expediting the optimization of specific properties in metal-organic frameworks. This suggests its potential applicability across various other research domains, highlighting its versatility and effectiveness in streamlining complex optimization challenges.

# 5

## Development of Software Toolsets for AI Integration in MOF Synthesis

### 5.1 Comparison of Genetic Algorithm and Bayesian Optimization

In this chapter, I delve into a comparative analysis between two prominent machine learning approaches: the Genetic Algorithm (GA) and Bayesian Optimization (BO), applied to the specific challenge of optimizing MOF nanoparticles towards a targeted size. This exploration showcases the distinct methodologies of GA and BO and also assesses their efficiency in addressing the same optimization problem within the field of porous material synthesis.

Some of the findings presented in this chapter were previously published as part of a Bachelor's thesis at the Technical University of Munich.[89]

#### 5.1.1 Introduction

The SyCoFinder, leveraging a genetic algorithm, has already established itself as an invaluable resource for optimizing specific properties in porous materials. However, machine learning encompasses a wide array of techniques including several alternatives for tackling optimization challenges. Among these, the Bayesian algorithm stands out for its proficiency in managing uncertainties or incomplete data sets, which closely aligns with the parameter spaces I derived from the influential variables in our study. In collaboration with Luisa Ortner from Dr. Felix Dietrich's group at the Department of Informatics at the Technical University of Munich (TUM), we initiated a comparative study to assess the effectiveness of Bayesian Optimization (BO) against the previously used Genetic Algorithm (GA). The primary distinction between these two algorithmic strategies lies in their approach to optimization tasks: the GA adopts an evolutionary strategy, initiating with a population of candidate solutions which it then recombines and mutates to foster improvement. Conversely, BO is grounded in Bayes' Theorem, which outlines the process of updating hypotheses with the incorporation of new information. Essentially, BO predicts outcomes and revises these predictions as new data becomes available, embodying its learning mechanism.

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

To practically compare the two algorithms, Luisa Ortner developed a Bayesian Optimization using a Gaussian Process, mirroring the approach I previously took with the GA. The Gaussian Process is particularly useful for identifying optima in situations where data or prior knowledge is limited. It not only predicts an objective function for a given point in the search space but also estimates the uncertainty associated with that point. This capability allows for a strategic balance in the optimization process, navigating between explored and unexplored regions to potentially uncover superior solutions. To determine the next point for evaluation, the BO employs acquisition functions that negotiate between the predictions and uncertainties produced by the Gaussian Process. Delving into the specifics of this implementation would exceed the scope of this thesis; hence, readers interested in a more detailed exploration are directed to Luisa Ortner's Bachelor's Thesis.[89]

For the BO to make its initial predictions, it requires data, which, in the context of the SyCoFinder, is derived from the Diverse Set. To facilitate this, I contributed data from ZIF-8 nanoparticles optimized for crystallinity and specific particle size, produced before I began my doctoral studies. This endeavor aimed at achieving a target particle size of 100 nm, setting the groundwork for further research into the application of MOFs in drug delivery.

Nonetheless, Luisa Ortner's group considered the typical ten experiments from the Diverse Set that I had utilized to be insufficient for the BO's initial training. Consequently, they opted to train the BO on all 30 experiments I had conducted before my doctoral studies, believing this larger dataset, attached in Appendix A.19, would provide a more robust foundation for the BO's optimization efforts. From this extended dataset, they generated ten new predictions for parameter combinations, which are to be found in Appendix A.20. I then carried out these predicted experiments and conducted practical measurements.

I performed the syntheses using an automated programmable microwave reactor. ZIF-8 was synthesized using zinc acetate as the metal source, 2-methylimidazole as the linker molecule, sodium hydroxide as the modulator, and methanol as the solvent. The selected variables and their respective limits, which were consistent for both GA and BO, are listed in Table 5.1. The evaluation method for the experiments remained uniform across both optimization approaches, employing XRD to assess phase purity and dynamic light scattering (DLS) to determine the particle size. The detailed procedure for the experimental synthesis can be found in Chapter 7.6, and the fitness calculation is based on Equation 5.1.

$$fitness = fitness(phase\ identity) * \left( 1 - \left( \frac{1}{\left| \frac{d(g)}{d(r)-d(g)} \right| + 1} \right) \right) * (1 - PDI) \quad (5.1)$$

Phase identity was once again employed as the binary exclusion criterion, as detailed in previous optimization chapters, with the sole difference being that only one XRD measurement per synthesis was carried out. This adjustment was made because I was dealing with nanoparticle syntheses, which produced only a single sample per synthesis. The term in parentheses pertains to the particle size and moves closer to a value of one as the nanoparticles near the target size of 100 nm. Additionally, the Polydispersity Index (PDI) serves as a measure for the size distribution of the particles. Being a value that ranges between zero and one, it approaches one as the particle size distribution widens, and is thus subtracted from

one to reflect this relationship.

Variable	Limits
Temperature [ $^{\circ}$ C]	40 – 140
Time [Min]	10 – 45
Metalsalt [mmol]	0.01 – 2.00
Reactants ratio [eq Linker of Metal]	0.5 – 50
Modulator ratio [eq Modulator of Metal]	0 - 10

Tab. 5.1: Definition of variables and their respective limits in which they were varied.

### 5.1.2 Results and Discussion

To evaluate the performance of BO against the GA, I synthesized and assessed ten samples from the BO and compared them directly with the outcomes from the second iteration of the Genetic Algorithm (GA2) within the SyCoFinder optimization. The X-Ray diffractograms forming the basis for the phase identity are attached in Appendix A.4.3 and the complete values from DLS measurements are included in Tables A.19 and Table A.20. Tables 5.2 and 5.3 present the comparative fitness values covering not only the individual evaluation criteria - phase identity, particle size, and PdI - but also the overall fitness. This comparison is crucial because the GA2 data had already been integrated into the training set for BO (See Appendix A.19). It's important to clarify that the values presented in Tables 5.2 and 5.3 reflect the fitness terms derived from the measurements as per Equation 5.1, rather than the direct measurement values themselves.

Tab. 5.2: Fitness terms of BO. [Adapted and reproduced from [89]]

Tab. 5.3: Fitness terms of GA2. [Adapted and reproduced from [89]]

Particle size	PdI	Phase identity	Fitness	Particle size	PdI	Phase identity	Fitness
0.13	0.00	1	0.00	0.17	0.81	1	0.14
0.06	0.64	1	0.04	0.26	0.33	1	0.09
0.33	0.24	1	0.08	0.36	0.92	1	0.33
0.66	0.94	1	0.62	0.72	0.92	1	0.66
0.16	0.15	1	0.02	0.71	0.92	1	0.65
0.33	0.50	1	0.16	-	-	0	0.00
0.14	0.29	1	0.04	0.53	0.88	1	0.47
0.10	0.00	1	0.00	0.16	0.54	1	0.09
0.10	0.34	1	0.03	0.20	0.03	1	0.01
0.17	0.49	1	0.08	-	-	0	0.00

A key comparison between these two optimization methods is phase purity, which indicates whether the synthesis resulted in the desired compound. In my analysis, all ten parameter combinations suggested by the BO produced pure ZIF-8 nanoparticles, whereas only eight out of ten attempts were successful with GA2.

Turning to the DLS measurements, I received insights into the fitness related to the interdependent particle size and PdI, as illustrated in Figure 5.1b. GA2 boasted

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

two exceptionally good experiments, followed closely by one notable result from the BO. The bulk of BO's suggestions fell into the lower performance range, while the rest of the GA2 outcomes were distributed more evenly, with half of them outperforming those suggested by BO.

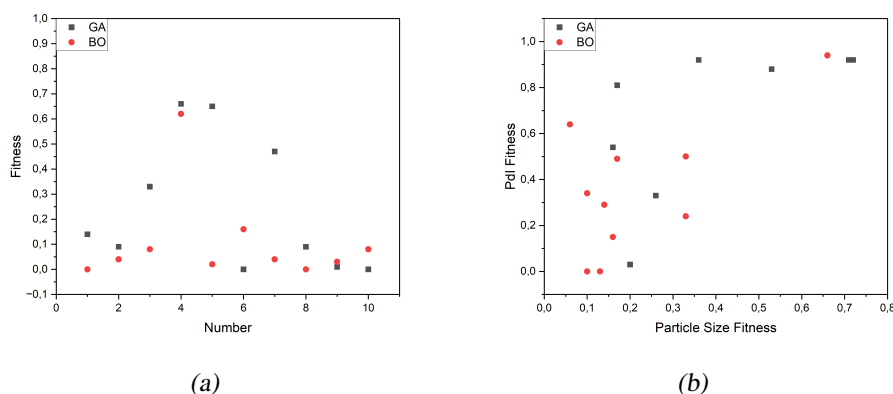


Fig. 5.1: a) Fitness values of BO (red dots) and GA2 (grey squares). b) Fitness of the particle size against the fitness of the PdI.[Adapted and reproduced from [89]]

Focusing on the overall fitness as illustrated in Figure 5.1a, I found that the top experiments from both BO and GA2 achieved similar fitness values - 0.66 for GA2 and 0.62 for BO. However, GA2 slightly outperformed BO, presenting two superior experiments to BO's one, with the remainder of BO's suggestions showing comparatively lower fitness levels. Intriguingly, when examining the parameters of the best experiments from both BO and GA2 in Table 5.4, and comparing them to the variable ranges in Table 5.1, I discovered the temperature, quantity of metal salt, and reactants ratio were closely aligned. The modulator ratio also showed similarities, with only the time variable displaying a significant discrepancy.

Tab. 5.4: Parameter sets of the two best syntheses of GA2 and BO.

Optimization	Temperature [°C]	Time [Min]	n(Metal) [mmol]	Reactants Ratio [eq Linker of Metal]	Modulator Ratio [eq Modulator of Metal]
BO	137.5	37.00	1.53	37.2	0.7
GA2	130.14	11.77	1.22	48.7	2.3

In drawing a conclusion, it's noteworthy that BO was trained on data generated from three generations of GA, which in contrast was evolved without any prior dataset. Ultimately, both methods produced one leading experiment that was closely matched in both parameters and fitness, yet the GA2's was slightly superior. Given that BO was trained on already optimized conditions and all prior data from the GA optimization - thus having access to two-thirds more data points than the GA, which only had ten data points from its diverse set - it was anticipated that BO might outperform GA2 significantly. However, this was not the case, even though the unexplored space for BO was much smaller than that for GA. This suggests



that, with further refinements, BO could potentially be more effective, or it may indicate that this specific optimization challenge is inherently more suited to a GA approach.

### 5.1.3 Summary

In this study, we compared optimizations techniques using a Genetic Algorithm (GA) and Bayesian Optimization (BO). The BO was applied by Luisa Ortner from Dr. Felix Dietrich's group at the Department of Informatics at the Technical University of Munich (TUM), specifically designed for our project aimed at optimizing ZIF-8 nanoparticles. Our goal was to achieve nanoparticles with a target size of 100 *nm* and low polydispersity.

The BO was trained on data generated by the GA, following the SyCoFinder protocol. Initially, the GA was informed by ten data points from the Diverse Set and subsequently refined through its first iteration to produce the GA2 experiments. We then compared the outcomes of all ten BO experiments against the ten GA2 experiments.

In this comparison, we discovered that the BO consistently recommended parameters that resulted in crystalline ZIF-8 nanoparticles for all experiments, whereas the GA2 yielded crystalline forms in 8 out of 10 experiments. However, when evaluating both methods based on the polydispersity criterion, particle size and overall fitness, the GA2 demonstrated clear superiority over the BO. Consequently, at least for the time being, the GA remains the preferred method for this particular optimization challenge.

## 5.2 Modification of SyCoFinder

### 5.2.1 Introduction

In my endeavor to optimize DASBDC as detailed in Chapter 4.2, I had already faced the challenge of selecting between two solvents: ethanol and methanol. However, the SyCoFinder was initially unable to process categorical variables. To circumvent this limitation, I had employed a strategy using a solvent ratio, where the SyCoFinder calculated one part of the ratio - effectively one solvent - while the counterpart was implicitly determined as the remainder of the ratio. In this particular project this approach was feasible since using a mixture of ethanol and methanol was also acceptable. But this raises questions for scenarios where a choice must be strictly made between one solvent or another, or when the decision involves choosing from more than two solvents, or even selecting among different linkers to determine which performs best in MOF synthesis. This challenge of incorporating categorical variables into optimization tasks applies broadly to any scenario requiring an exclusive selection from a set of distinct options, where intermediate values are not viable.

At its core, making choices from a set of defined options is straightforward when selections can be made randomly. However, integrating this process into a machine learning framework, like the SyCoFinder, introduces complexity. The SyCoFinder operates on evolutionary principles, combining and mutating parental values to

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

generate offspring, which to some extent always result in a mixture of their predecessors. This means evolutionary approaches are inherently not designed to choose exactly one option. Furthermore, marrying the need for continuous variables - such as time, temperature, or concentration - with categorical choices, like for example solvent type, significantly complicates the optimization task.

To develop a solution to this idea, I had the opportunity to collaborate with Professor Moosavi, the principal developer of the SyCoFinder, and his research group at the Department of Applied Sciences, University of Toronto (UofT).

As explained in previous chapters, the SyCoFinder operates through three sequential steps: The Diverse Set, which establishes a foundational, most diverse parameter set for training the GA; the GA itself, which refines these parameters towards a specific goal; and the Importance of Variables, which sheds light on the significance of the selected variables. Within the context of this project, we successfully adapted the Diverse Set to accommodate both categorical and continuous variables.

### 5.2.2 Modification

Before calculating the Diverse Set in the SyCoFinder tool, it is necessary to select the variables, define their lower and upper limits, and possibly assign weights to these variables. The specification of lower and upper limits indicates we are working with continuous variables, which means the variable can assume any value within the defined range. For instance, if the temperature variable has limits set from 25 to 150 degrees, it can adopt values like 25, 26, 27, and so forth, up to 150. In contrast, categorical variables operate differently: they consist of a specific set of values without any intermediate options. For example, if the solvent variable includes options such as "DMF," "Ethanol," and "Hexane," only one of these values can be chosen at a time. This distinction poses a challenge for the SyCoFinder, which is primarily designed to work with continuous variables.

The purpose of the Diverse Set is to encompass as broad a range of the experimental space as possible, constructed from all selected variables. This involves identifying the most dissimilar point from a randomly chosen starting point within the parameter space, which in case of the Diverse Set is generated as a grid of points for each variable. Such a strategy aims to assemble the most varied set of parameter combinations, providing the subsequently following GA with a comprehensive overview of the search space to start its optimization process. To achieve this, a technique known as MaxMin is employed, demonstrated in Figure 5.2 using a straightforward two-dimensional example involving Temperature and Time. For simplicity, in this illustration, only points at the intersections of the grid lines are considered as potential selections. The first point, Point 1, is the randomly selected starting point within the parameter space. The next step involves finding Point 2, identified by calculating the minimal distance from Point 1, and then determining Point 3, which is the furthest away from Point 2. This method guarantees the selection of the most dissimilar parameter values relative to previously chosen landmark points, thereby maximizing coverage of the experimental parameter space. This is crucial for effectively exploring the space, especially in scenarios with higher-dimensional parameter spaces, which as in all my previous optimization scenarios, is usually the

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

case. The procedure ends, when the amount of landmark points matches the defined amount of experiments by the user.

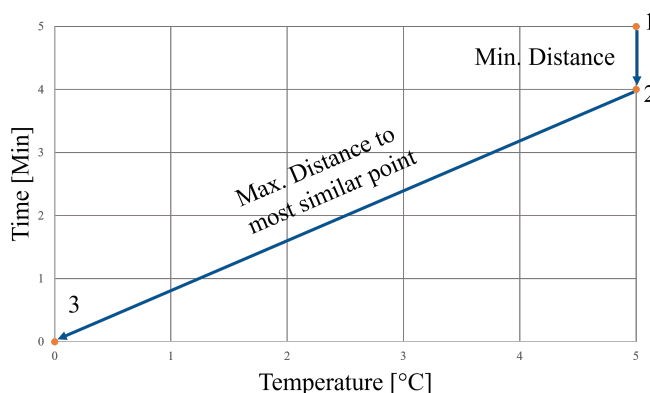


Fig. 5.2: Illustration of the selection process for the points chosen in the Diverse Set on a simple two-dimensional example.

The output resulting from this procedure typically appears as shown in Table 5.5, exemplary featuring five variables across ten experiments. Our aim was to introduce an option for generating output similar to the example provided in Table 5.6. Clearly, the datatype for categorical variables, represented by strings, is distinct from the data in Table 5.5, which comprises integer or floating-point values. Defining a parameter space with these string variables poses a challenge, as their lack of numerical value makes placement within the space uncertain. More critically, the inability to compute vectors from strings means that determining minimal and maximal distances, a key aspect of our optimization process, becomes infeasible.

Tab. 5.5

Temperature [°C]	Time [Min]	Water [mL]	Metal [mmol]	Linker [mmol]	Solvent [category]
25	1	1	0.01	0.01	Ethanol
150	45	6	1	2	DMF
25	1	3.5	1	2	Hexane
150	1	6	0.01	1	DMF
150	23	1	1	0.01	Hexane
25	45	6	0.51	0.01	Hexane
87.5	45	1	0.01	2	Ethanol
87.5	1	6	1	0.01	DMF
150	1	1	0.51	2	Ethanol
25	23	6	0.01	2	Hexane

Tab. 5.6

To manage categorical data, I required a numerical representation, and various methods are available, as shown in Table 5.7. Integer encoding, for instance, would assign sequential values to each category. However, this approach was not suitable for my needs because it could introduce bias due to the numerical hierarchy, af-

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

fecting the equal probability of selection for each solvent. An alternative was to use descriptors, which involve assigning numerical values based on solvent properties, such as the boiling points in my example. While this is demonstrated in Table 5.7, it similarly risks unequal bias. Moreover, while boiling points can describe solvents, the goal was to accommodate any category, potentially requiring different descriptors and adding complexity by necessitating a system for users to define these descriptors. Ultimately, I opted for one-hot encoding, a method that translates categorical data into a binary format (0's and 1's), ensuring each category is uniquely represented without inherent bias, as also illustrated in Table 5.7. The term 'one-hot' derives from the characteristic that in this encoded sequence, only one '1' appears per encoded term.

Having selected this method as the most suitable for my problem, the categorical variables could then be integrated into the Diverse Set process.

Tab. 5.7: Encoding options to assign numerical values to strings.

<b>Strings</b>	<b>Integer Encoding</b>	<b>Descriptors</b>	<b>One-Hot Encoding</b>
DMF	1	153°C	[1 0 0]
Ethanol	2	78°C	[0 1 0]
Hexane	3	69°C	[0 0 1]

To create a Diverse Set that includes both continuous and categorical variables, I had to modify the main execution function, specifically the compute function. Previously, generating a Diverse Set with only continuous variables began by establishing the continuous parameter space through grid dimension calculations. This grid, constructed from all variables (dimensions), determines the number of samples per dimension by dividing each dimension into specified sections. For instance, if we set each dimension to range from zero to one and divide it into three grid points per dimension, we would have three samples per dimension. Consequently, for three variables, the total number of samples the Diverse Set would need to explore is calculated as  $3 * 3 * 3$ , resulting in 27 samples, indicating that the continuous space comprises 27 distinct samples. These options are calculated and stored in an array.

Now, defining the search space for categorical variables was necessary. Unlike continuous variables, applying a grid doesn't work for categorical variables because they don't occupy a continuous space that can be divided into sections. Instead, the number of categorical options directly represents the categorical search space. However, to integrate both categorical and continuous spaces, their respective options needed to be multiplied to determine the total number of samples within the search space. For instance, if there were four categorical options to combine with the 27 options from the continuous space, the total search space would then encompass 108 possible options. This highlights a critical aspect of the process: the addition of even a single variable can exponentially increase the size of the search space, significantly extending the calculation time as the number of dimensions grows.

To go beyond merely calculating the number of samples in the search space and actually compute the 108 possible options in my example, it was essential to inte-

grate the combination of continuous and categorical data. This required utilizing the previously mentioned one-hot encoded data, which I then combined with the array containing the continuous options. This combination was possible because categorical variables had been transformed from strings into unique binary representations. It's important to ensure that the formatting is consistent throughout the array, especially since one-hot encoded categories are represented by three digits, whereas each continuous variable is represented by a single value.

After incorporating all 108 options into the array, it was processed using the MaxMin method. This allowed for the handling of all inputs and the generation of the Diverse Set by selecting options based on the MaxMin principle and the number of samples specified by the user. However, the resulting Diverse Set was still in its normalized form, as exemplary illustrated in Table 5.8. To prepare the final output for the user, I first isolated the continuous portion of the normalized Diverse Set and re-encoded it following an already established procedure. Next, I addressed the categorical section of the array, applied a one-hot re-encoding function, and merged both re-encoded parts. This process yielded the complete Diverse Set, ready to be presented to the user.

Tab. 5.8: Processing the returned Diverse Set to its final output form.

Status of DS	Example
Normalized DS	[0., 0., 0., 1., 0., 0.]
Selection of the two parts of the DS	[0., 0., 0.] [1., 0., 0.]
Reencoding continuous Part	[100. 0. 0.8 1. 0. 0. ]
Reencoding categorical Part	['100.0' '0.0' '0.8' 'Ethanol']

### Testing the Modifications

Listing 5.1: Example for mixed variables - numerical and categorical

```
# Example weights for all four variables
var_importance = [1.0, 1.0, 1.0, 1.0]
# Lower bounds for numerical variables
var_LB = [100, 0, 0.8]
# Upper bounds for numerical variables
var_UB = [200, 40, 2.0]
# The actual categories
categories = ['DMF', 'Ethanol', 'Hexan', 'Propanol']
# Calculated number of categories
number_of_categories_list = [len(categories)]
# Providing Input parameter for computing the diverse set
diverse_set = compute(
    var_importance=var_importance,
    var_LB=var_LB,
    var_UB=var_UB,
    number_of_categorical_variables=1,
    number_of_categories_list=number_of_categories_list,
    num_samples=10,
    ngrids_per_dim=3
)
```

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

In Listing 5.1, I present a test example to demonstrate the modifications, with the resulting output displayed in Listing 5.2. This confirms that the modifications are effective, showcasing three continuous or numerical variables - such as 'Temperature [°C]', 'Water [mL]', and 'Concentration [mmol]' - alongside one categorical variable 'Solvents', which includes the options 'DMF', 'Ethanol', 'Hexane', and 'Propanol'. The full final adapted code of the Diverse Set is attached in Appendix A.5.

Generated Diverse Set:

Listing 5.2: Generated output from example for mixed variables

```
[ ['100.0', '0.0', '0.8', 'Ethanol']  
[ '200.0', '40.0', '2.0', 'Ethanol']  
[ '100.0', '20.0', '2.0', 'Hexan']  
[ '150.0', '40.0', '0.8', 'Propanol']  
[ '200.0', '0.0', '0.8', 'Ethanol']  
[ '200.0', '0.0', '2.0', 'DMF']  
[ '100.0', '40.0', '1.4', 'DMF']  
[ '150.0', '0.0', '1.4', 'Propanol']  
[ '200.0', '20.0', '1.4', 'Hexan']  
[ '100.0', '0.0', '1.4', 'Ethanol']]
```

### Limitations

While the Diverse Set can now accommodate both numerical and categorical variables, it is currently limited to handling just one set of categorical variables. This limitation is generally adequate for many applications. However, to increase the capacity for multiple sets of categorical variables, further modifications in array management will be necessary. Additionally, for a comprehensive machine learning optimization process, adjustments to the GA, which typically follows the Diverse Set, are also needed.

### 5.2.3 Summary

In this project, we have successfully adapted the code of the Diverse Set by implementing one-hot encoding, which now allows for the combination of categorical values and continuous variables. Besides, I modified the existing code to enable calculation of the search space, incorporating categorical options. I adapted the program for making the data compatible with the MaxMin method, allowing for the selection of suitable options for the final Diverse Set. Subsequently, I re-encoded the normalized data to ensure for user convenience and easy interpretation. This advancement marks a significant improvement in the system's ability to handle diverse data types. While significant progress has been made, the task to modify the Genetic Algorithm (GA) code accordingly and enhance the Importance of Variables feature remain open. Addressing these components will be a focus of future projects, aiming to further develop and enhance the system's overall capabilities.

### 5.3 Development of Chemotion input mask

The field of chemistry boasts a rich tradition of conducting experiments, advancing syntheses, analyzing data, and deriving insights from these activities. Historically, chemists have meticulously documented their experimental procedures and outcomes. However, the utility of handwritten notes is limited to those who possess them, and such records are prone to being misplaced or lost. Similarly, data stored on personal computers face challenges regarding safety, accessibility, and searchability; data confined to a single hard drive requires prior knowledge of its existence for it to be retrieved, rather than the convenience of a search function. These limitations hinder the pace of material development and optimization, as building upon prior research necessitates easy access to past findings. Without such access, experiments may be unnecessarily duplicated by different researchers, sometimes leading to redundant or unproductive approaches. Conversely, being able to seamlessly continue from where previous research left off, rather than retracing already covered ground, can significantly expedite scientific progress. This underscores the critical need for creating databases that are universally accessible. The upcoming chapter explores the creation of enhancements and adjustments for porous materials within such a database system, Chemotion. This platform is designed to facilitate the storage and publication of data in alignment with the FAIR principles discussed in Chapter 2.5, making scientific information readily available and usable for the broader community.

#### 5.3.1 Introduction

In recent years, it has become increasingly recognized as good scientific practice to publish raw data alongside publications. This shift is largely due to the efforts of the GO-FAIR initiative, which has been addressing the much-needed improvements in the findability, accessibility, interoperability, and re-usability of scientific data. This approach is not just a convenience for today's researchers but also a valuable contribution to the scientific community, as it reduces the likelihood of data loss and makes previously collected data available for reuse or reinterpretation. In the context of this thesis, which focuses on fully autonomous machine learning optimizations, having accessible data is an essential component. It is crucial to have a system for storing synthesis conditions when experiments are conducted, and for managing the measurement files generated from sample evaluations, ideally linking them with their corresponding synthesis data. For these data to be useful in machine learning optimizations and to allow for automatic evaluation, they must be in a machine-readable format. Chemotion addresses these needs by offering an electronic lab journal that serves as a modern replacement for the traditional handwritten lab notebook, alongside a connected repository for the long-term storage and publication of data following the FAIR principles.

The challenge we face is that research across various fields is often documented on an individual basis. While certain standardizations have been established in areas of chemistry, such as the conventions for drawing reaction schemes, the diversity becomes more pronounced when delving into subfields like organic chemistry or materials science. Here, the types of information that need to be stored diverge significantly, beginning with different setups for synthesis and methods for mea-

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

surements. To ensure that all necessary information is captured without loss, our approach involves working towards standardizations specific to each research field. However, it is crucial to also accommodate the unique needs of each field by offering tailored options.

Throughout my doctoral studies, I had the privilege of collaborating with Dr. Nicole Jung, Pei-Chi Huang, and Chia-Lin Lin from Prof. Bräse's group at the Institute of Biological and Chemical Systems, KIT. This collaboration began with the first machine learning optimization discussed in Chapter 4.1 and continued throughout my doctoral research. Together, we developed standardized input masks for Chemotion tailored to the field of materials science, which were then implemented at the Institute of Functional Interfaces (IFG), KIT. Specifically, we integrated various synthesis methods pertinent to porous material synthesis, delving into the intricacies of the required input fields. In the following sections, I will present the latest developments that are currently active in the Chemotion repository. These updates have been utilized to publish the data from my most recent projects.

### 5.3.2 Results and Discussion

#### Reaction Schema

One significant challenge was creating a consistent notation for MOF or SURMOF reaction schemas. Previously, the practice involved merely illustrating the components from which the MOF or SURMOF was constructed, which failed to accurately represent the porous structure. To address this, we introduced a simplified version that not only includes the molecular components but also illustrates how they are interconnected, thereby reflecting the framework's structure. For SURMOFs, this notation additionally shows how the components are attached to a surface. An example of this new approach, applied to the previously discussed DASBDC pillar-layered SURMOF, is depicted in Figure 5.3.

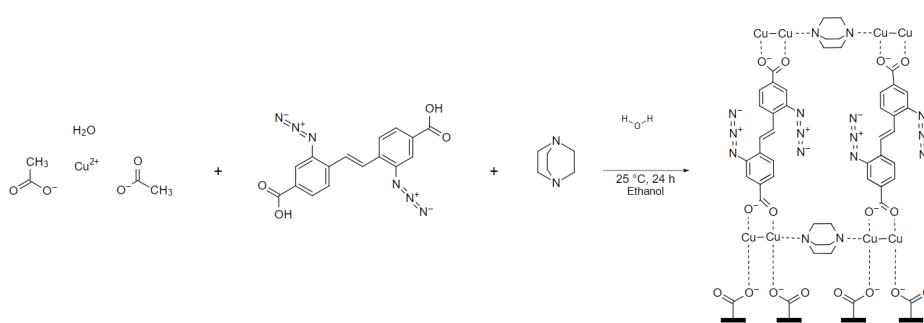


Fig. 5.3: Reaction schema of the synthesis of the DASBDC pillar layered SURMOF.

#### (SUR)MOF Segment

After experimenting with separate segments for various synthesis methods related to MOFs, SURMOFs, and nanoparticles - which often led to confusion about where to place specific information - we opted for a unified MOF segment. This comprehensive section encompasses all possibilities and guides users through the initial



## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

steps by highlighting the main aspects, then revealing only the relevant options to be completed. Initially, users are prompted to select the Type of Metal-Organic Framework, with choices including MOFs, SURMOFs, and nanoparticles. This choice as presented in Figure 5.4 then narrows down the available reaction techniques; for instance, a dipping method would not be applicable for nanoparticle production. Based on the selected reaction technique, tailored input fields are presented to collect specific information.

Type of Metal Organic Framework  
SURMOF

Reaction techniques  
Select...  
Dipping  
Pump and Flow  
Spin coating  
Spraying  
Chemical vapor assisted

Processing type  
Select...  
manual  
automated  
half-automated  
others

Fig. 5.4: Unified MOF segment starting with queries to guide the user to the correct input mask.[Screenshot from Chemotion Repository [90]]

My primary role in this process in addition to idea generation, was to collate the necessary input fields from experts who specialize in the various synthesis techniques and to devise those for the reaction techniques I personally utilized. A significant part of my contribution involved developing the segment for reaction techniques using dipping robots. This included not just gathering details on applied concentrations and solvents, but also documenting program sequences, the exact synthesis set-up and systematically recording the unique composition in each vessel. To ensure utmost precision and comprehensively include all relevant details for the dipping robots, we introduced sections dedicated to materials preparation and solvent mixture definition illustrated in Figure 5.5.

Preparation of the used chemicals/materials in detail

Table 1: Preparation of materials

	Vessel No	Vessel Part	Function	Molecule	Method	Solvent	Volume - Part	Volume - Vessel	Duration	Temperature
+	V	A	m		s.	EtOH	210 ml	210 ml	20 m	25 °C
+	V	A	s.		n.	EtOH	210 ml	210 ml	20 m	25 °C
+	V	A	I.		s.	EtOH	174 ml	210 ml	20 m	25 °C

Table 2: Definition of solvents

	Vessel No	Solvent A	Volume A	Solvent B	Volume B	Solvent C	Volume C
+	V1	EtOH	210 ml	Select...	ml	Select...	l
+	V2	EtOH	210 ml	Select...	ml	Select...	l
+	V5	EtOH	210 ml	Select...	ml	Select...	l

Fig. 5.5: Section for materials preparation and solvent definition.[Screenshot from Chemotion Repository [90]]

## 5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS

Additionally, to specifically address the synthesis process using robots, we defined what constitutes one cycle, with an example provided in Figure 5.6. Moreover, for SURMOFs in particular, we incorporated a section detailing substrates and substrate preparation, as depicted in Figure 5.7. It's important to acknowledge that while I played a key role in developing the principles and required fields for the input masks, the actual implementation was carried out by Dr. Nicole Jung, Pei-Chi Huang, and Chia-Lin Lin.

**Definition of synthesis cycle**

Number of cycles  
40

Table 3: Setting referring to materials and their conditions

	Step number	Vessel number	Step type	Duration	Temperature
+ x	1	V1	metal insta...	10 m	25 °C
+ x	2	V2	rinsing	2 m	25 °C
+ x	3	V3	rinsing	2 m	25 °C
+ x	4	V4	rinsing	2 m	25 °C
+ x	5	V5	linker insta...	15 m	25 °C
+ x	6	V6	rinsing	2 m	25 °C
+ x	7	V7	rinsing	2 m	25 °C
+ x	8	V8	rinsing	2 m	25 °C

Humidity [% rel.]  
12 %

Atmosphere requirements  
protective N2

Other conditions

Termination type  
storage in solution

Solvent of termination  
EtOH

Duration of termination step  
3 h

Fig. 5.6: Defining a synthesis cycle for the dipping robot synthesis in Chemotion.[Screenshot from Chemotion Repository [90]]

**Preparation of substrate**

Substrate  
Silicon wafer

Coating  
gold

Dimensions  
1x3cm

Activation  
SAM

Sam type  
MHDA (16- Mercaptohexa...

Duration  
72 h

Fig. 5.7: Section for substrate choice and substrate preparation for SURMOFs.[Screenshot from Chemotion Repository [90]]

Building on the already comprehensive features of Chemotion, which include sections for documenting reactions, chemical properties, and measurement data storage, our newly developed segment serves as a significant enhancement, seamlessly integrating with the platform's commitment to long-term and complete information preservation. Furthermore, coupling this expansion to Chemotion's overar-

## *5. DEVELOPMENT OF SOFTWARE TOOLSETS FOR AI INTEGRATION IN MOF SYNTHESIS*

ching capabilities like assigning DOIs to data in the repository and offering data in machine-readable JSON formats, aligns with the FAIR principles and brings Chemotion features to the material science community.

### **5.3.3 Summary**

Creating a balance between specificity and generality poses a significant challenge, especially when developing an input mask that needs to be broad enough to accommodate numerous cases while also being detailed enough to encompass all necessary information. We successfully navigated this challenge by enhancing features in Chemotion, our well-established research data management system. Among these enhancements was the development of an input mask tailored for robotic layer-by-layer synthesis. This new feature is designed to capture all relevant details and is adaptable to a wide range of synthesis setups, paving the road to integrate this system in autonomous optimization processes. Looking ahead, this development is poised to support future robotic synthesis endeavors, all in alignment with the FAIR principles. This ensures our advancements contribute positively to the scientific community by facilitating more efficient and accessible research data management.

## 6

# Summary and Outlook

This study was guided by the vision of fully autonomous laboratories, delving into several relevant aspects of accelerating materials optimization to enhance and tailor porous materials for diverse applications. The use of automated synthesis systems, the evaluation of machine learning methods' efficiency, and the management of research data in this context played vital roles and were addressed each according to their inherent challenges.

In the beginning I conceptualized an illustration of meshing gears (See Figure 1.1) each representing essential components for autonomous optimization processes interlocking to symbolize their interdependence. One prominent gear in this schematic depicted the employment of automated synthesis procedures. I argued that transitioning from manual to automated synthesis is not just beneficial but critical for machine learning tasks, primarily due to the gains in reproducibility and the capacity to execute tasks in parallel. Reproducibility is of paramount importance in any machine learning-based optimization, since the learning process is purely based on the data generated by experiments. Therefore, the integrity of this data is crucial, as undocumented, user-dependent variables can mislead the learning process, affecting the outcomes and efficiency of the optimization efforts.

In the first half of my thesis, I undertook studies on three distinct practical machine learning optimizations, each varying in complexity, whether due to the chemical system involved or the intricacies of the machine learning task itself (See the upper three subfigures in Figure 6.1). These studies were conducted utilizing an automated robotic synthesis setup, integrating advanced automation to streamline the optimization process. The machine learning method employed was the SyCoFinder, developed by Moosavi et al., specifically designed to enable materials optimizations without any prior data. It achieves this by generating its own empirical dataset, from which a genetic algorithm can learn and subsequently suggest optimized synthesis parameters. I applied it for three distinct optimization tasks on metal-organic framework thin films, each with different objectives. This involved emphasizing certain parameters in SURMOFs and also evaluating the general applicability and effectivity of the evolutionary driven SyCoFinder.

In the first project, the objective was to optimize a specific orientation, the 111-orientation, in HKUST-1 SURMOF. Chemically, producing the HKUST-1 system is not particularly challenging. However, given the selected SAM and substrate,

## 6. SUMMARY AND OUTLOOK

achieving the 111-orientation was unlikely. Therefore, my challenge lay in identifying synthesis conditions that could facilitate the growth of this specific orientation. In this project, I simultaneously varied five variables across ten experiments per generation. In thirty experiments, or three generations - with the first being the empirically diverse set - a best experiment was identified with an 84% fitness, exhibiting a 100% pure 111-orientation. With the last step of the SyCoFinder, the evaluation of the importance of variables, I could also identify the critical variable to achieve this goal - in fact the absence of water as a modulator favors the 111-orientation due to kinetics in bond formation. Thus, the SyCoFinder proved its worth by quickly yielding the desired optimized product and also provided valuable insights into the growth mechanism of HKUST-1 SURMOF.

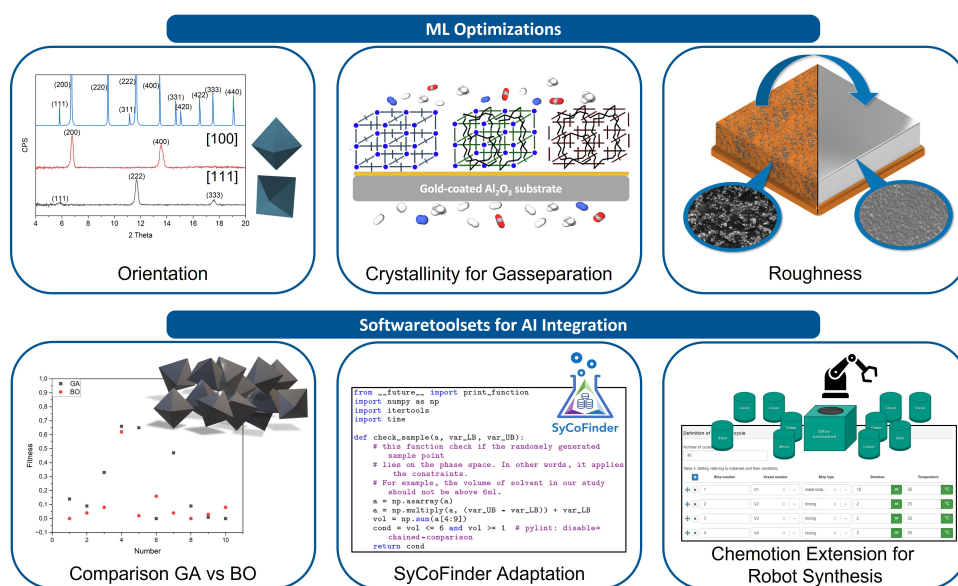


Fig. 6.1: Illustrative summary of all projects undertaken throughout this thesis. ML Optimizations (upper) from left to right: Orientation Optimization [Adapted and reproduced from [28]], Crystallinity Optimization for Gasseparation [Adapted and reproduced from: [30]], Roughness Optimization [Adapted and reproduced from [52] - Manuscript accepted for publication]; Softwaretoolsets for AI Integration (lower) from left to right: Comparison of GA and BO towards nanoparticle size, Adaptation of the SyCoFinder to handle categorical variables [SyCoFinder Logo reproduced from [42]], Extension of Chemotion by developing an input mask for automated robotic synthesis [Robotic schema reproduced from [30], Screenshot from Chemotion Repository [90]].

The second optimization task was more complex in terms of the chemical system, focusing on the DASBDC-pillar-layered SURMOF system, interconnected through both DASBDC and DABCO. DASBDC features two azide groups available for post-synthesis modification. After synthesizing the SURMOF, I crosslinked the structures and then removed the metal atoms, resulting in a highly structured polymer, referred to as a SURGEL. The final objective this time was also more application oriented, aiming to test the SURMOF, crosslinked SURMOF and SURGEL as gas separation membranes. To achieve this, the goal of the optimization was to produce a highly crystalline and stable SURMOF, as a well-defined template structure was crucial for all substructures derived from it. Again I varied five vari-

## 6. SUMMARY AND OUTLOOK

ables and received a best sample in thirty experiments. I then applied the optimized conditions, initially developed on gold-coated silicon wafers, to porous alumina-oxide substrates and achieved the same quality in the SURMOF, demonstrating the transferability of the optimized conditions. The switch to a different substrate was specifically aimed at the gas separation experiments, to avoid a potentially damaging membrane transfer. The subsequent tests for gas separation revealed that the crosslinked SURMOF exhibited superior selectivity, particularly in separating  $H_2/CO_2$ .

Now that the SURMOF synthesis is optimized and the end product has demonstrated promising performance in gas separation experiments, it would be prudent to also optimize the crosslinking process. Currently, this process is time-intensive and could potentially be shortened while aiming for the highest crosslinking ratio to enhance the membrane's stability and resistance to defects. In the long term, integrating this membrane into a real-world application for gas separation would be an interesting avenue to explore.

After two successful optimizations, the third project was designed to thoroughly test the capabilities of the SyCoFinder. I varied seven variables across ten experiments per generation, targeting multiple goals simultaneously. The primary goal was to reduce the roughness of the HKUST-1 SURMOF surfaces to make them suitable for use in electronic devices, such as sensors. Additionally, the SURMOF needed to be pure-phased and exhibit high crystallinity. Despite challenges in evaluating roughness, I successfully received a best sample again within thirty experiments. Furthermore, I gained valuable insights into the influence of variables, particularly the importance of the initial few cycles in forming the SURMOF's foundation. These initial cycles may benefit from a different concentration than the rest of the synthesis, usually lasting for forty cycles.

The future direction of this optimization is clearly indicated by the goal, which was to demonstrate that roughness in SURMOFs can be reduced, aiming to facilitate their integration into electronic devices. Additionally, a key aspect of this project was to identify the critical variables involved, with the aim of applying this knowledge to other SURMOFs as well. However, when optimizing other SURMOF systems for smoothness, the evaluation process could potentially be accelerated by employing a measurement method better suited to automation than ellipsometry. Moreover, given the limitations of obtaining reliable results from ellipsometry with very rough samples, switching to a high-resolution microscopy technique could be a beneficial alternative. In ellipsometry, many samples that in principle yielded HKUST-1 were disregarded because their high roughness could not be properly evaluated, leading to a significant loss of information on the parameter space. This data would have been valuable to the machine learning algorithm, despite the samples in question being very rough. Only samples with roughness below a certain threshold could be evaluated reliably with ellipsometry. Therefore, changing the measurement method to utilizing all available data will accelerate future optimization efforts and therefore bring the integration of SURMOFs into various electronic devices closer to realization.

Addressing the broader topic, the second half of my thesis concentrated more on aspects related to the machine learning toolset, briefly illustrated in the lower three subfigures of Figure 6.1. Having thoroughly tested the application of a genetic

## 6. SUMMARY AND OUTLOOK

algorithm, I began to consider what other machine learning methods might be suitable for my optimization problems and whether there might be one that could outperform the evolutionary strategy. Therefore, I initiated a comparative study between the genetic algorithm (GA) and Bayesian Optimization (BO), developed in collaboration with a research group at the Technical University of Munich. I evaluated these algorithmic approaches based on the task of optimizing ZIF-8 nanoparticles to achieve a specific particle size. Both algorithms performed similarly in their respective best experiment, though the GA was closer to the optimization goals. The determining factor in giving the GA preference over BO of this comparison was that the BO had been trained on all experiments that emerged from the GA's optimization, including the GA's optimal one, yet did not suggest more optimal parameters than those identified by the GA. Therefore, in this investigation the GA of the SyCoFinder remained the preferred algorithm for subsequent optimization tasks.

However, considering that BO offers many adjustable parameters, further tailoring the BO could still enhance its effectiveness to be applied as a valuable tool for future optimizations.

During the second optimization I had encountered a problem concerning my selected variables: I had used a mixture of methanol and ethanol as the solvent because it was uncertain which solvent or solvent ratio would yield the best synthesis results. For two solvents, I could easily use the ratio of one as a variable and complete the total volume with the other solvent. However, this approach made me realize that it would not be feasible in the SyCoFinder with three or more solvents, particularly if the experiment required the exclusive choice of just one solvent. This realization led to the idea of integrating categorical variables. In collaboration with Professor Moosavi, I modified the SyCoFinder's Diverse Set to accommodate both categorical and continuous variables, such as time, temperature, and concentrations. This adaptation has enhanced the SyCoFinder's capability to address a wider range of optimization challenges.

This adaptation also holds high potential for future enhancements: Initially, the changes applied to the Diverse Set must also be integrated into the GA to facilitate a comprehensive optimization process. Then subsequently, the scientific community could utilize the SyCoFinder for a wider array of optimization challenges, such as varying different linker molecules in a MOF structure along with other synthesis parameters.

After addressing the intricate challenges of machine learning and employing automated synthesis and measurement techniques, the final section of my thesis tackles the issues related to management of the recorded data. Although the goal of machine learning optimization is to identify the optimal parameter combination for the best sample, it generates a significant amount of valuable data throughout the process. During my optimization projects, I stored data in the Chemotion ELN and, upon publication, shared the raw measurement data and synthesis conditions with the scientific community via the Chemotion repository. Especially by publishing all attempts, including failed ones, and leveraging Chemotion's capability to provide data in a machine-readable format, anyone can use my data for example to train machine learning models. However, using automated robotic synthesis setups, which differ from the traditional chemical synthesis setups already stan-

## 6. SUMMARY AND OUTLOOK

standardized in Chemotion, made it challenging to comprehensively cover all synthesis conditions, preparation methods, substrate specifications, and more. To address this issue, I developed a new input mask for robotic syntheses in Chemotion, in collaboration with the software developers.

Returning to the concept of autonomous materials optimization, this thesis has played a pivotal role in assessing and improving various gears to enhance the transmission of experimental work into advances in porous material science. This research can now act as a reference endeavor demonstrating strategies and insights for integrating the explored components and accelerating application oriented optimizations procedures in the future. Envision a future where one simply instructs a computer with requirements and returns later to discover a product that has been meticulously optimized.



# 7

## Experimental

### 7.1 Substrate Preparation

For all SURMOF syntheses, both the gold-coated silicon wafer substrates and the gold-coated alumina oxide substrates were pre-treated with 16-mercaptohexadecanoic acid (MHDA) 72 *h* prior to the synthesis. The MHDA solution was prepared by dissolving 1.4 *mg* of MHDA in 25 *mL* of glacial acetic acid and 225 *mL* of ethanol (99.8%). This alkane-thiolate-based SAM solution underwent ultrasonication for 30 *min*, under shielding from light. The SAM solution was utilized for up to two weeks and was freshly prepared thereafter.

### 7.2 Solvents

All solvents for all experiments in this thesis were purchased commercially and used without further purification.

### 7.3 MOF thin film Optimization towards Orientation

#### 7.3.1 SURMOF synthesis

The solutions containing metal salt and organic linker to yield HKUST-1 were prepared separately. For the metal salt solution a defined amount of copper(II) acetate was dissolved in 210 *mL* of absolute ethanol (99.8%) and treated with ultrasonication for 15 *min*. The solution of the organic linker contained both the linker molecule 1,3,5-benzenetricarboxylic acid (BTC) and water as the modulator. The total volume of this solution again was set to 210 *mL*, therefore subtracting the volume of the modulator from the solvents volume, which again was ethanol (abs.). BTC was dissolved in the respective mixture of water and ethanol and treated with ultrasonication for 15 *min*. The respective amounts for each experiment in the machine learning procedure can be found in the appendix (See Appendix A.1). The freshly prepared metal salt and linker solutions as well as the required cleaning solutions, each containing 210 *mL* ethanol were then transferred to the robotic set up in a glovebox in closed vessels. Here the solutions were transferred to their respective vessels at the respective position corresponding to the robotic program

## 7. EXPERIMENTAL

sequence as described in Chapter 4.1.2. The only exception in terms of volume is the filling of the ultrasonication bath, which contains 420 mL of ethanol (abs.).

The starting position of the synthesis robot, a TX-60 robot from Stäubli, is occupied with an empty vessel carrying the sample holder lid, to which the formerly prepared four substrates are fixed upside down. In a starting sequence the robot grips the lid from above, elevates it and starts the main program by immersing the substrates for a defined time in the first solution, always defined as the metal salt solution. This is followed by one or more cleaning steps with the options of spray cleaning, ultrasonication or dip cleaning, followed by the linker solution and again one or more cleaning steps, before the next cycle starting with the metal solution. If not stated differently the amount of cycles was always fixed to 40. The immersion times for metal salt solution was 10 min, 5 min in case of any dip cleaning and 15 min respectively for the linker solution. The times for spray cleaning and ultrasonication were varied according to the ML-suggestions in appendix A.1 between 0 – 5 s and 0 – 100 s respectively.

As final step of the synthesis, the robot released the sampleholder at the finishing position, containing fresh ethanol.

### 7.3.2 Measurements

**XRD** I measured all x-ray diffractograms within a  $2\Theta$  range of  $3^\circ$  to  $20^\circ$ , utilizing a Bruker D8 Advance diffractometer configured in  $\Theta - \Theta$  geometry and outfitted with a LYNXEYE position-sensitive detector that features 192 active stripes. To identify the distinctive gold diffraction peak of the substrate, which served as a reference, I additionally recorded the range from  $37^\circ$  to  $40^\circ$   $2\Theta$ . All measurements were taken at a rate of two seconds per step, amounting to a total measurement time of 384 seconds for each step. Afterwards I performed height error correction with respect to the gold diffraction peak and background subtraction on each measurement file, using the DIFFRAC.EVA software version 5.2.0.3 from Bruker AXS.

## 7.4 MOF thin film Optimization for Gasseparation Application

### 7.4.1 Sample holder 3D-printing

The selected material for the sample holder is a viscous fluid composed of a fluorinated methacrylate monomer, which solidifies through UV curing in the printing process. The particular printing method used is known as Digital Light Processing (DLP), carried out with an Asiga Max UV DLP Printer.

### 7.4.2 SURMOF synthesis

The solutions containing metal salt and organic linker were prepared separately, thus in case of the metal salt containing copper(II) acetate in a respective mixture of methanol and ethanol according to the parameters provided by the SyCoFinder in the appendix (See Appendix A.2). The linker solution is composed of defined

## 7. EXPERIMENTAL

amounts of 2,2'-Diazido-4,4'-stilbenedicarboxylic acid (DASBDC) as linker, 1,4-Diazabicyclo-[2.2.2]octane (DABCO) as second linker and water as modulator all dissolved in the same mixture ratio of methanol and ethanol as the metal salt solution. Copper(II) acetate and DABCO were purchased commercially and used without further purification, but the DASBDC linker was synthesized by Dr. Simon Oßwald of the group of Prof. Bräse at the Institute of Organic Chemistry, KIT, according to the published synthesis conditions from Tsotsalas et al. in 2014.[31] The total volume when synthesizing the DASBDC-pillar-layered SURMOF on gold coated silicon wafers is fixed to 210 mL, in case of gold coated alumina oxide substrates the volume amounts to 260 mL. It is to be noted, that the volume of the modulator water needs to be subtracted from the total solvent volume in case of the linker solution before calculating the respective amounts of methanol and ethanol according to the provided ratio. Both solutions are treated with ultrasonication for 15 min.

The prepared solutions are transferred to the robotic setup as described in Chapter 7.3 and placed at their respective positions. The robot is again a TX-60 six-axis industrial robot from Stäubli. For the synthesis set up apart from metal salt and linker solutions two times three cleaning solutions are prepared containing pure solvent mixture of methanol and ethanol according to the respective ML parameters. All vessels are placed around the robot at their respective program sequence position and the synthesis is carried out for 40 cycles. The sequence of one cycle behaves according to: 1. Metal solution (10 minutes), 2.-4. Dip cleaning in the respective Ethanol/Methanol mixture (each 2 minutes), 5. Linker solution (15 minutes), and 6.-8. Dip cleaning in the respective Ethanol/Methanol mixture, but in separate vessels (each 2 minutes). As a last step the substrates are placed in a final vessel containing fresh solvent, with the same ratio of methanol to ethanol as all other solutions.

### 7.4.3 Crosslinking

The crosslinker was synthesized in accordance with literature specifications and kindly provided by Dr. Simon Oßwald from Prof. Bräse's group at the Institute of Organic Chemistry, KIT.[31]

For crosslinking a solution of 1 mg/ml trimethylolethane tripropiolate in toluene was prepared. The substrates carrying the produced SURMOFs were immersed in this solution and constantly heated at 80°C for seven days. Afterwards the substrates were rinsed several times with ethanol and acetone and dried under nitrogen flow.

### 7.4.4 Transformation to SURGEL

Following the procedure detailed in the publication of Tsotsalas et al. the crosslinked SURMOFs were immersed in a 1 : 1 water:ethanol solution with 1 mg/10ml EDTA for 90 min.[31]

In case of the SURMOFs on gold coated silicon wafer substrates, this was sufficient to find the crystalline structure gone in XRD. For the porous gold coated alumina oxide substrates I adjusted the amount of EDTA to 2 mg/10ml and increased the immersion time to four days, to yield an amorphous X-Ray diffractogram.

## 7. EXPERIMENTAL

For purification the SURGEL was rinsed with ethanol and water and dried under nitrogen flow.

### 7.4.5 Measurements

**XRD** The main XRD signals appear similar to the HKUST-1 between  $3^\circ$  and  $20^\circ 2\theta$ . Therefore the same measurement conditions were applied:

I utilized a Bruker D8 Advance diffractometer, set up in  $\Theta - \Theta$  geometry and equipped with a LYNXEYE position-sensitive detector boasting 192 active stripes. To pinpoint the characteristic diffraction peak of gold from the substrate, which acted as a reference, I additionally recorded the diffraction angle range from  $37^\circ$  to  $40^\circ 2\theta$ . All measurements were recorded at a consistent pace of two seconds per step and for subsequent data acquisition, I undertook the task of applying height error correction relative to the gold diffraction peak and executing background noise elimination on the collected data files, employing the DIFFRAC.EVA software, version 5.2.0.3, developed by Bruker AXS.

**IRRAS** All measurements were conducted using a Vertex 80 FT-IR spectrometer from Bruker. The spectra were acquired through infrared reflection-absorption spectroscopy (IRRAS) with a grazing incidence angle of  $80^\circ$ . For reference measurements, deuterated  $C_{16}-SH$  was applied on a gold-coated silicon wafer substrate. The absorption measurements spanned the mid-infrared range from  $4000\text{ cm}^{-1}$  to  $600\text{ cm}^{-1}$ .

**Gas separation performance measurement** The gas separation efficiency of the SURMOF, SURMOF-CL, and SURGEL samples was assessed by Dr. Nicholas Prasetya from the Institute of Functional Interfaces, KIT, using a Wicke-Kallenbach apparatus. Initially, to precondition the samples, they were installed in the membrane permeation cell. Here, nitrogen and helium gases were introduced at flow rates of  $50\text{ mL min}^{-1}$  on the feed and permeate sides, respectively, and maintained for a minimum of 18 h. This procedure aimed to dry the samples under controlled conditions while limiting their exposure to ambient air. Following this preconditioning phase, the permeation of pure gases -  $H_2$ ,  $CH_4$ ,  $N_2$ , and  $CO_2$  - through the membrane was measured at a consistent flow rate of  $50\text{ mL min}^{-1}$ . Argon, serving as the sweep gas, was introduced to the retentate side at a flow rate of  $10\text{ mL min}^{-1}$  to assist in the separation process. The permeance of the gases through the membranes was determined using a Varian micro gas chromatograph (GC) CP-4900, which utilized argon as the carrier gas and was fitted with molsieve and PPQ columns. To ensure accuracy, the gas permeance measurements were repeated at least five times until consistent readings were achieved.

## 7. EXPERIMENTAL

### 7.5 MOF thin film Optimization towards Roughness Reduction

#### 7.5.1 SURMOF synthesis

The utilized chemicals, the preparation procedure, as well as the robotic set-up is equal to the one applied in the optimization towards orientation (Refer to Chapter 7.3). As substrates again the gold coated silicon wafers were utilized. However, the synthesis procedure varied slightly: The total amount of 40 cycles per experiment is divided into three cycles plus 37 cycles. In the first three cycles different concentrations of metal salt and linker solution are applied than in the subsequently following 37 cycles. Therefore the content of the vessels at the positions for metal salt and linker in the robotic set up were exchanged after cycle number 3. All applied concentrations as well as the employed values for spray cleaning and ultrasonication are to be found in the appendix (See Appendix A.3).

#### 7.5.2 Measurements

**XRD** Once more, I made use of a Bruker D8 Advance diffractometer, configured in a  $\Theta - \Theta$  geometry and outfitted with a LYNXEYE position-sensitive detector. To identify the distinct diffraction peak of gold from the substrate, serving as a reference point, I specifically captured the diffraction angle range between  $37^\circ$  and  $40^\circ 2\Theta$ . Measurements were systematically taken at a rate of two seconds per step. For the analysis phase, I performed height error adjustments in relation to the gold diffraction peak and carried out background noise reduction on the dataset, utilizing the DIFFRAC.EVA software, version 5.2.0.3, provided by Bruker AXS.

**Ellipsometry** I performed the spectroscopic ellipsometry measurements with an M2000 instrument from J.A. Woollam Co. Inc., located in Lincoln, NE, USA. These measurements took place at a  $70^\circ$  angle of incidence, spanning a spectral range from 370 to 1000 nm, under ambient conditions. For the analysis of the SURMOF thin film's experimental data, I applied a Cauchy-fit model through the CompleteEase software (V5.19) provided with the instrument. This model accounts for surface roughness by introducing an additional effective medium layer, which is modeled as a 50 – 50 mix of the ambient environment and the layer material. I established the optical properties of the gold-coated silicon substrate by conducting a reference measurement on the uncoated substrate before beginning the synthesis.

### 7.6 Comparison GA vs BO

#### 7.6.1 ZIF-8 Nanoparticle synthesis

I performed the synthesis of ZIF-8 nanoparticles as a microwave assisted synthesis, employing a Monowave 450 Microwave Reactor from Anton Paar, utilizing C30 wide neck reaction vessels with a total volume of 30 mL and a maximum filling volume of 20 mL.

First I prepared metal salt solutions of zinc acetate dihydrate in methanol also con-

## 7. EXPERIMENTAL

taining the respective amount of modulator sodium hydroxide under stirring for 30 *min*. In parallel but in a separate vial I prepared the linker solution containing 2-methylimidazole in methanol treated with ultrasonication for 20 *min*. The respective concentrations of the metal salt and corresponding equivalents of linker and modulator are noted in the appendix (See Appendix A.4).

Directly prior to the synthesis the corresponding metal salt and linker solution were combined in the reaction vessel and placed in the autosampler of the microwave. The syntheses were conducted with time and temperature set according to the suggestions of the Bayesian optimization prediction and are to be found in the appendix (See Appendix A.4).

For purification all samples were centrifuged and washed three times with methanol and subsequently three times with ethanol. During the last washing step the samples were divided: I kept 2 *mL* of each sample in suspension for further analysis via DLS and dried the rest under reduced pressure for characterization via XRD.

### 7.6.2 Measurements

**XRD** For ZIF-8 I adapted the measurement range to 4° to 40° 2 $\Theta$  again employing the Bruker D8 Advance diffractometer in  $\Theta - \Theta$  geometry. All measurements were conducted at a consistent speed of two seconds per step.

In the process of evaluating the acquired data applied height error correction with respect to the gold diffraction peak to all files and performed a background subtraction applying built in routines from the DIFFRAC.EVA software, version 5.2.0.3, developed by Bruker AXS.

**DLS** I conducted all DLS measurement in ethanol, thus first inserting 50  $\mu\text{L}$  of a sample into a disposable sizing cuvette and filling it up with 2.5 *mL* of ethanol. In the measurement chamber the samples were equilibrated to 20°C for 3 *min* and then measured twice with 12 inherent measurement intervals. The acquired values for particle size and polydispersity index were directly used for calculation without further data refinement.

# 8

## References

- (1) O. M. Yaghi, G. Li and H. Li, *Nature*, 1995, **378**, 703–706.
- (2) S. Kitagawa, R. Kitaura and S.-i. Noro, *Angewandte Chemie International Edition*, 2004, **43**, 2334–2375.
- (3) K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T.-H. Bae and J. R. Long, *Chemical Reviews*, 2012, **112**, Publisher: American Chemical Society, 724–781.
- (4) L. J. Murray, M. Dincă and J. R. Long, *Chemical Society Reviews*, 2009, **38**, Publisher: Royal Society of Chemistry, 1294–1314.
- (5) C. Jiang, X. Wang, Y. Ouyang, K. Lu, W. Jiang, H. Xu, X. Wei, Z. Wang, F. Dai and D. Sun, *Nanoscale Advances*, 2022, **4**, Publisher: RSC, 2077–2089.
- (6) W. Liang, P. M. Bhatt, A. Shkurenko, K. Adil, G. Mouchaham, H. Aggarwal, A. Mallick, A. Jamal, Y. Belmabkhout and M. Eddaoudi, *Chem*, 2019, **5**, 950–963.
- (7) S. Wuttke, M. Lismont, A. Escudero, B. Rungtaweevoranit and W. J. Parak, *Biomaterials*, 2017, **123**, 172–183.
- (8) P. Horcajada, R. Gref, T. Baati, P. K. Allan, G. Maurin, P. Couvreur, G. Férey, R. E. Morris and C. Serre, *Chemical Reviews*, 2012, **112**, Publisher: American Chemical Society, 1232–1268.
- (9) S. Begum, Z. Hassan, S. Bräse, C. Wöll and M. Tsotsalas, *Accounts of Chemical Research*, 2019, **52**, 1598–1610.
- (10) M. Eddaoudi, D. B. Moler, H. Li, B. Chen, T. M. Reineke, M. O’Keeffe and O. M. Yaghi, *Accounts of Chemical Research*, 2001, **34**, 319–330.
- (11) H. Furukawa, K. E. Cordova, M. O’Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
- (12) J. L. C. Rowsell and O. M. Yaghi, *Microporous and Mesoporous Materials*, 2004, **73**, 3–14.
- (13) K. Müller, K. Fink, L. Schöttner, M. Koenig, L. Heinke and C. Wöll, *ACS Applied Materials & Interfaces*, 2017, **9**, 37463–37467.
- (14) S. S.-Y. Chui, S. M.-F. Lo, J. P. H. Charmant, A. G. Orpen and I. D. Williams, *Science*, 1999, **283**, Publisher: American Association for the Advancement of Science, 1148–1150.

## 8. REFERENCES

- (15) V. Gargiulo, A. Policicchio, L. Lisi and M. Alfe, *Energy & Fuels*, 2023, **37**, Publisher: American Chemical Society, 5291–5302.
- (16) M. Alfe, A. Policicchio, L. Lisi and V. Gargiulo, *Renewable and Sustainable Energy Reviews*, 2021, **141**, 110816.
- (17) R. J. Kuppler, D. J. Timmons, Q.-R. Fang, J.-R. Li, T. A. Makal, M. D. Young, D. Yuan, D. Zhao, W. Zhuang and H.-C. Zhou, *Coordination Chemistry Reviews*, 2009, **253**, 3042–3066.
- (18) M. Schlesinger, S. Schulze, M. Hietschold and M. Mehring, *Microporous and Mesoporous Materials*, 2010, **132**, 121–127.
- (19) J. Klinowski, F. A. A. Paz, P. Silva and J. Rocha, *Dalton Transactions*, 2011, **40**, Publisher: Royal Society of Chemistry, 321–330.
- (20) Z. Wang and C. Wöll, *Advanced Materials Technologies*, 2019, **4**, 1800413.
- (21) P. Lindemann, M. Tsotsalas, S. Shishatskiy, V. Abetz, P. Krolla-Sidenstein, C. Azucena, L. Monnereau, A. Beyer, A. Gölzhäuser, V. Mugnaini, H. Gliemann, S. Bräse and C. Wöll, *Chemistry of Materials*, 2014, **26**, Number: 24 Publisher: American Chemical Society, 7189–7193.
- (22) I. Stassen, N. Burtch, A. Talin, P. Falcaro, M. Allendorf and R. Ameloot, *Chemical Society Reviews*, 2017, **46**, Number: 11 Publisher: The Royal Society of Chemistry, 3185–3241.
- (23) D.-H. Chen, H. Gliemann and C. Wöll, *Chemical Physics Reviews*, 2023, **4**, 011305.
- (24) J. Liu, O. Shekhah, X. Stammer, H. K. Arslan, B. Liu, B. Schüpbach, A. Terfort and C. Wöll, *Materials*, 2012, **5**, Number: 9 Publisher: Molecular Diversity Preservation International, 1581–1592.
- (25) T. P. Vello, M. Strauss, C. A. R. Costa, C. C. Corrêa and C. C. Bof Bufon, *Physical Chemistry Chemical Physics*, 2020, **22**, Number: 10, 5839–5846.
- (26) O. Shekhah, H. Wang, S. Kowarik, F. Schreiber, M. Paulus, M. Tolan, C. Sternemann, F. Evers, D. Zacher, R. A. Fischer and C. Wöll, *Journal of the American Chemical Society*, 2007, **129**, Number: 49, 15118–15119.
- (27) H. K. Arslan, O. Shekhah, J. Wohlgemuth, M. Franzreb, R. A. Fischer and C. Wöll, *Advanced Functional Materials*, 2011, **21**, 4228–4231.
- (28) L. Pilz, C. Natzeck, J. Wohlgemuth, N. Scheuermann, P. G. Weidler, I. Wagner, C. Wöll and M. Tsotsalas, *Advanced Materials Interfaces*, 2023, **10**, 2201771.
- (29) P. Thissen, J. Wohlgemuth, P. Weidler, D. Smilgies, L. Heinke, N. Schewe, M. Koenig, P. Krolla and C. Wöll, *Advanced Functional Materials*, **n/a**, 2301535.
- (30) L. Pilz, C. Natzeck, J. Wohlgemuth, N. Scheuermann, S. Spiegel, S. Oßwald, A. Knebel, S. Bräse, C. Wöll, M. Tsotsalas and N. Prasetya, *Journal of Materials Chemistry A*, 2023, **11**, Publisher: The Royal Society of Chemistry, 24724–24737.
- (31) M. Tsotsalas, J. Liu, B. Tettmann, S. Grosjean, A. Shahnas, Z. Wang, C. Azucena, M. Addicoat, T. Heine, J. Lahann, J. Overhage, S. Bräse, H. Gliemann and C. Wöll, *Journal of the American Chemical Society*, 2014, **136**, Number: 1, 8–11.



## 8. REFERENCES

- (32) Q. Li, J. F. Quinn and F. Caruso, *Advanced Materials*, 2005, **17**, 2058–2062.
- (33) T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997, 414 pp.
- (34) Y. Gavrilova, *AI vs. ML vs. DL: What's the Difference*, AI vs. ML vs. DL: What's the Difference, 2020, <https://serokell.io/blog/ai-ml-dl-difference> (visited on 02/04/2024).
- (35) S. Raschka and V. Mirjalili, *Machine Learning mit Python und Scikit-Learn und TensorFlow : Das umfassende Praxis-Handbuch für Data Science, Predictive Analytics und Deep Learning*, mitb, 2nd edn., 2018.
- (36) D. Sharma and N. Kumar, *International Journal of Advanced Research in Computer Engineering & Technology*, 2017, **6**, 2278–1323.
- (37) M. R. Karim, *Java Deep Learning Projects: Implement 10 real-world deep learning applications using Deeplearning4j and open source APIs*, Google-Books-ID: OsNiDwAAQBAJ, Packt Publishing Ltd, 2018, 428 pp.
- (38) Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich and M. Tsotsalis, *Angewandte Chemie International Edition*, 2022, **61**, Number: 19, e202200242.
- (39) C. W. Jones, *JACS Au*, 2022, **2**, Publisher: American Chemical Society, 1504–1505.
- (40) S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit and H. J. Kulik, *Nature Communications*, 2020, **11**, Publisher: Nature Publishing Group, 4068.
- (41) Y. J. Colón and R. Q. Snurr, *Chemical Society Reviews*, 2014, **43**, Number: 16 Publisher: The Royal Society of Chemistry, 5735–5749.
- (42) S. M. Moosavi, L. Talirz and B. Smit, *Synthesis condition finder*, materialscloud.org, 2022, <https://www.materialscloud.org/work/tools/sycofinder> (visited on 09/24/2022).
- (43) S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou and B. Smit, *Nature Communications*, 2019, **10**, Number: 1, 539.
- (44) *Genetic Algorithms*, ed. S. Ventura, J. María Luna and J. María Moyano, Accepted: 2023-02-15T14:44:34Z, IntechOpen, 2022.
- (45) B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams and A. G. Doyle, *Nature*, 2021, **590**, Number: 7844 Publisher: Nature Publishing Group, 89–96.
- (46) P. I. Frazier, *A Tutorial on Bayesian Optimization*, 2018.
- (47) E. Lee, D. Eriksson, D. Bindel, B. Cheng and M. Mccourt, Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), Conference on Uncertainty in Artificial Intelligence, ISSN: 2640-3498, PMLR, 2020, pp. 260–269.
- (48) M. Linares-Moreau, L. A. Brandner, T. Kamencek, S. Klokic, F. Carraro, K. Okada, M. Takahashi, E. Zojer, C. J. Doonan and P. Falcaro, *Advanced Materials Interfaces*, 2021, **8**, Number: 21, 2101039.
- (49) J. Liu and J. E. Hein, *Nature Synthesis*, 2023, **2**, Number: 6 Publisher: Nature Publishing Group, 464–466.

## 8. REFERENCES

- (50) M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. Aspuru-Guzik, *Accounts of Chemical Research*, 2022, **55**, Publisher: American Chemical Society, 2454–2466.
- (51) A. L. Dias and T. Rodrigues, *Nature*, 2023, **624**, Bandiera\_abtest: a Cg\_type: News And Views Number: 7992 Publisher: Nature Publishing Group Subject\_term: Chemistry, Computer science, Machine learning, 530–531.
- (52) L. Pilz, M. Koenig, M. Schwotzer, H. Gliemann, C. Wöll and M. Tsotsalas, *Advanced Functional Materials*, Accepted for Publication.
- (53) P. Tremouilhac, P.-C. Huang, C.-L. Lin, Y.-C. Huang, A. Nguyen, N. Jung, F. Bach and S. Bräse, *Chemistry-Methods*, 2021, **1**, 8–11.
- (54) C.-L. Lin, P.-C. Huang, S. Graessle, C. Grathwol, P. Tremouilhac, S. Vanderheiden, P. Hodapp, S. Herres-Pawlis, A. Hoffmann, F. Fink, G. Manolikakes, T. Opatz, A. Link, M. M. B. Marques, L. J. Daumann, M. Tsotsalas, F. Biedermann, H. Mutlu, E. Täuscher, F. Bach, T. Drees, S. Neumann, N. Jung and S. Bräse, *Supporting Sustainability of Chemistry by Linking Research Data with Physically Preserved Research Materials*, preprint, Chemistry, 2023.
- (55) P. Kalhor, N. Jung, S. Bräse, C. Wöll, M. Tsotsalas and P. Friederich, *Advanced Functional Materials*, 2023, **n/a**, 2302630.
- (56) C. R. Brundle and C. A. Evans, *Encyclopedia of Materials Characterization*, Elsevier, 1992, vol. 14.
- (57) In *The Basics of Crystallography and Diffraction*, ed. C. Hammond, Oxford University Press, 2015.
- (58) L. Spieß, G. Teichert, R. Schwarzer, H. Behnken and C. Genzel, *Moderne Röntgenbeugung: Röntgendiffraktometrie für Materialwissenschaftler, Physiker und Chemiker*, Springer Fachmedien, Wiesbaden, 2019.
- (59) P. J. Larkin, *Infrared and Raman Spectroscopy : Principles and Spectral Interpretation*, Elsevier, Second Edition, 2018.
- (60) W. Bechmann and I. Bald, *Einstieg in die Physikalische Chemie für Naturwissenschaftler*, Springer, Berlin, Heidelberg, 2020.
- (61) H. Günzler and H.-U. Gremlich, *IR-Spektroskopie: eine Einführung*, Wiley-VCH, Weinheim, 4., vollst. überarb. und aktualisierte Aufl, 2003, 352 pp.
- (62) H.-U. Gremlich, in *Ullmann's Encyclopedia of Industrial Chemistry*, ISSN: 1435-6007, John Wiley & Sons, Ltd, 2000.
- (63) J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. Ritchie, J. H. J. Scott and D. C. Joy, *Scanning Electron Microscopy and X-Ray Microanalysis*, Springer, New York, NY, 2018.
- (64) R. Egerton, *Physical Principles of Electron Microscopy*, Springer International Publishing, Cham, 2016.
- (65) A. Ul-Hamid, *A Beginners' Guide to Scanning Electron Microscopy*, Springer International Publishing, Cham, 2018.
- (66) T. Yoshizawa, *Handbook of optical metrology: principles and applications / edited by Toru Yoshizawa*, Taylor & Francis Inc, 2015, 919 pp.

## 8. REFERENCES

- (67) J. Company, Woollam, *Polarized Light*, J.A. Woollam, <https://www.jawoollam.com/resources/ellipsometry-tutorial/polarized-light> (visited on 02/17/2024).
- (68) J. A. W. Company, *CompleteEASE Software Manual*.
- (69) J. Stetefeld, S. A. McKenna and T. R. Patel, *Biophysical Reviews*, 2016, **8**, 409–427.
- (70) W. Schärfl, *Light Scattering from Polymer Solutions and Nanoparticle Dispersions*, Springer, Berlin, Heidelberg, 2007.
- (71) L. Pilz, *Fully Automated Optimization of Robot-Based MOF Thin Film Growth via Machine Learning Approaches*, Data of the Diverse Set, 2022.
- (72) L. Pilz, *Fully Automated Optimization of Robot-Based MOF Thin Film Growth via Machine Learning Approaches*, Data of Genetic Algorithm 1, 2022.
- (73) L. Pilz, *Fully Automated Optimization of Robot-Based MOF Thin Film Growth via Machine Learning Approaches*, Data of Genetic Algorithm 2, 2022.
- (74) Q. Min Wang, D. Shen, M. Bülow, M. Ling Lau, S. Deng, F. R. Fitch, N. O. Lemcoff and J. Semanscin, *Microporous and Mesoporous Materials*, 2002, **55**, Number: 2, 217–230.
- (75) B. D. McCarthy, T. Liseev, A. M. Beiler, K. L. Materna and S. Ott, *ACS Applied Materials & Interfaces*, 2019, **11**, Publisher: American Chemical Society, 38294–38302.
- (76) J. Liu and C. Wöll, *Chemical Society Reviews*, 2017, **46**, Number: 19 Publisher: The Royal Society of Chemistry, 5730–5770.
- (77) X. Liu, M. Kozłowska, T. Okkali, D. Wagner, T. Higashino, G. Brenner-Weiß, S. M. Marschner, Z. Fu, Q. Zhang, H. Imahori, S. Bräse, W. Wenzel, C. Wöll and L. Heinke, *Angewandte Chemie International Edition*, 2019, **58**, 9590–9595.
- (78) R. Haldar, M. Jakoby, A. Mazel, Q. Zhang, A. Welle, T. Mohamed, P. Krolla, W. Wenzel, S. Diring, F. Odobel, B. S. Richards, I. A. Howard and C. Wöll, *Nature Communications*, 2018, **9**, Number: 1 Publisher: Nature Publishing Group, 4332.
- (79) J.-L. Zhuang, M. Kind, C. M. Grytz, F. Farr, M. Diefenbach, S. Tussupbayev, M. C. Holthausen and A. Terfort, *Journal of the American Chemical Society*, 2015, **137**, Number: 25, 8237–8243.
- (80) K. Müller, J. Singh Malhi, J. Wohlgemuth, R. A. Fischer, C. Wöll, H. Gliemann and L. Heinke, *Dalton Transactions*, 2018, **47**, Number: 46, 16474–16479.
- (81) A. Liaw and M. Wiener, *R News*, 2002, **2**, 5.
- (82) Z.-G. Gu, A. Pfriem, S. Hamsch, H. Breitwieser, J. Wohlgemuth, L. Heinke, H. Gliemann and C. Wöll, *Microporous and Mesoporous Materials*, 2015, **211**, 82–87.
- (83) B. Zhang, J. Zhang, C. Liu, X. Sang, L. Peng, X. Ma, T. Wu, B. Han and G. Yang, *RSC Advances*, 2015, **5**, Number: 47, 37691–37696.
- (84) N. Nijem, K. Fürsich, S. T. Kelly, C. Swain, S. R. Leone and M. K. Gilles, *Crystal Growth & Design*, 2015, **15**, Number: 6, 2948–2957.

## 8. REFERENCES

- (85) L. Pilz, *Utilizing machine learning to optimize metalorganic framework-derived polymer membranes for gas separation*, Data of the complete Optimization, 2023.
- (86) L. Pilz, *Enhancing the Quality of MOF Thin Films for Device Integration through Machine Learning: A Case Study on HKUST-1 SURMOF Optimization*, Data of the complete Optimization, 2024.
- (87) B. Mohan, S. Kumar, H. Xi, S. Ma, Z. Tao, T. Xing, H. You, Y. Zhang and P. Ren, *Biosensors and Bioelectronics*, 2022, **197**, 113738.
- (88) S.-I. Ohira, Y. Miki, T. Matsuzaki, N. Nakamura, Y.-k. Sato, Y. Hirose and K. Toda, *Analytica Chimica Acta*, 2015, **886**, 188–193.
- (89) L. Ortner, Bachelor's thesis, Technical University of Munich, 2023, 63 pp.
- (90) S. Bräse, N. Jung, P. Tremouilhac, P.-C. Huang, C.-L. Lin, Y.-C. Huang, A. Nguyen and F. Bach, *Chemotion Repository*, Chemotion Repository, <https://www.chemotion-repository.net/home/welcome> (visited on 03/07/2024).
- (91) S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou and B. Smit, *Capturing chemical intuition in synthesis of metal-organic frameworks*, 2018.
- (92) L. Talirz and S. M. Moosavi, *GitHub - materialscloud-org/sycofinder at v0.1.0*, materialscloud-org/sycofinder, <https://github.com/materialscloud-org/sycofinder/tree/v0.1.0> (visited on 03/07/2024).

# List of Figures

1.1	Components of autonomous optimization. . . . .	3
2.1	Schematic representation of the construction of a metal-organic framework; the metal nodes (blue spheres) connect with the organic linker molecules (yellow struts) to form a three-dimensional porous network. . . . .	5
2.2	a) Schematic illustration of the copper paddlewheel, a highly popular metal-oxo cluster, featuring a di-copper core coordinated by four carboxylic groups. b) Structural representation of an HKUST-1 MOF, composed of copper (blue) and trimesic acid (brown: carbon, red: oxygen, white: hydrogen). . . . .	6
2.3	This figure illustrates the formation of surface anchored metal-organic frameworks. Metal nodes (blue spheres) interact with organic linker molecules (yellow bars) to produce highly oriented SURMOFs in a layer-by-layer fashion. This method commences with a self-assembled monolayer (SAM, green bars), which forms the initial layer for SURMOF growth. [Adapted and reproduced from: [28]] . . . . .	7
2.4	a) Classification of the terms artificial intelligence, machine learning, neural networks and deep learning. [Graph inspired by Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials, Yulia Gavrilova, April 08, 2020 [34]] b) Illustration of the subtopics of machine learning: supervised learning, unsupervised learning and reinforcement learning. They are differentiated by their respective tasks. . . . .	9
2.5	This figure provides an overview of the three main tools of the synthesis condition finder (SyCoFinder). [SyCoFinder Logo reproduced from [42]] . . . . .	11
2.6	This figure provides an overview of the actual optimization process provided by the synthesis condition finder (SyCoFinder). [Adapted and reproduced from: [30]] . . . . .	12
2.7	This diagram depicts the sequence of steps processed by a genetic algorithm to predict and subsequently optimize parameter sets. . .	13
2.8	a) Photograph of the robotic set-up at the Institute of Functional Interfaces at KIT.[Adapted and reproduced from [52] - Manuscript accepted for publication] b) Exemplary sampleholder for the robotic set-up, capable of carrying four samples per synthesis. . . . .	14
2.9	This figure illustrates the principles of FAIR Data. . . . .	15

List of Figures

3.1	a) This figure illustrates the principles of the origin of X-rays, with K, L, M representing the electron shells around the nucleus N. When electrons from shells near the nucleus are displaced by the impact of sufficiently energetic electrons, electrons from outer shells transition to these vacated positions, emitting X-rays in the process (eg. $K_{\alpha}$ ). b) Illustration of Bragg's law and the diffraction of X-Rays respectively, with the blue spheres representing atoms in lattice planes. . . . .	17
3.2	a) Illustration of various atomic layers and the spacing between them within a cubic crystal structure. b) Depiction of Miller indices of different lattice planes. [Reprinted from Encyclopedia of Materials Characterization, Vol 14, Brundle, C. Richard, Evans, Charles A., XRD - X-Ray Diffraction, Page No. 201, Copyright (1992), with permission from Elsevier][56] . . . . .	18
3.3	Schematic XRD measurement set-up. . . . .	19
3.4	Depiction of a two-atomic molecule simplified as a mechanical model comprising two masses linked by a spring. . . . .	20
3.5	a) Model of the harmonic oscillator. b) Model of the anharmonic oscillator. . . . .	21
3.6	Illustration of transverse light propagating in the z-direction, decomposed into components along the x- and y-axes for three scenarios: a) linearly polarized light, b) circularly polarized light, and c) elliptically polarized light. [The three Subfigures are inspired by [67].] . . . . .	23
3.7	Schematic illustration of measurement set-up in Ellipsometry. . .	24
3.8	Description of the fundamental sequence of modelling and fitting in ellipsometry for evaluating measurement data. [This Figure is inspired by J.A. Woollam, Software Manual for Spectroscopic Ellipsometers, Figure 3-1, SE Data Analysis Flowchart.[68]] . . . .	25
3.9	Schematic illustration of measurement set-up in DLS. . . . .	26
4.1	a) Illustration of the parameterspace built from several parameters in which the algorithm is to search the optimum, while effects on the synthesis outcome are both unknown and potentially intersecting when varying parameters simultaneously. [Reproduced from [28]] b) Schematic illustration of the robotic set-up for the dipping layer-by-layer synthesis under inert conditions. [Reproduced from [28]] . . . . .	30
4.2	a) Comparative X-ray diffractograms showcasing simulated bulk-MOF (in blue) versus HKUST-I SURMOFs grown with orientation in the [100] direction (in red) and the [111] direction (in black). [Reproduced from: [28]] b) top: a representative SEM image of an HKUST-I SURMOF oriented in the [100] direction; bottom: a representative SEM image of an HKUST-I SURMOF oriented in the [111] direction. [Adapted and reproduced from: [28]] . . . . .	31

List of Figures

4.3 This figure depicts the structural components of HKUST-1. a) The metal-salt providing the nodes; copper(II) acetate. b) The linker molecule forming the connection bonds between the metal centers; trimesic acid (BTC). c) The SAM which is the connection compound between surface substrate and first SURMOF layer; MHDA. 31

4.4 This figure is annotated with a consistent legend across all parts: Green squares represent data from the first generation (Diverse Set), blue triangles indicate the second generation (Genetic Algorithm 1), and brown dots correspond to the third generation (Genetic Algorithm 2). a) This panel displays the fitness values of successful trials within the machine learning-guided optimization process, with bars indicating the average fitness of successful trials per generation, color-coded to match the legend. [Reproduced from [28]] b) This section depicts the crystallinity trends among all trials producing HKUST-1, regardless of achieving the desired [111]-orientation, over the generations. [Reproduced from [28]] c) Illustrated here is the evolution of crystallinity in successful trials—those achieving a preferred 111-orientation and thus a fitness value greater than zero—across the three generations. [Reproduced from [28]] d) The progression of achieving the desired [111]-orientation throughout the three generations is detailed here. [Reproduced from [28]] . . . . . 34

4.5 SEM recordings for the best three HKUST-1-thin films in order of evolution over the optimization process. a) Sample 4 (04-DS); Fitness: 0,78. b) Sample 18 (08-GA1); Fitness: 0,71. c) Sample 27 (07-GA2); Fitness: 0,84. [Adapted and reproduced from [28]] . 35

4.6 The significance of each variable employed in the machine learning-guided optimization process is depicted for every variable. [Reproduced from [28]] . . . . . 36

4.7 X-Ray diffractograms of four samples from a synthesis conducted for 80 cycles with the best performing parameter combinations from the optimization; This proves that the orientation in [111] direction is preserved also for higher cycle numbers. [Reproduced from [28]] . . . . . 38

4.8 This figure depicts the structural components of the DASBDC-SURMOF and SURGEL. a) The metal-salt providing the nodes of the SURMOF; copper(II) acetate. b) The linker molecule forming the connection bonds between the metal centers; 2,2'-Diazido-4,4'-stilbenedicarboxylic acid (DASBDC). c) The second linker molecule forming the perpendicular connection bonds between the metal centers to form a pillar layered structure; 1,4-Diazabicyclo-[2.2.2]octane (DABCO). d) The crosslinker to interconnect the structure in a postsynthetic procedure; trimethylolethane tripropiolate. . 40

4.9 Visualization of the DASBDC-DABCO pillar layered SURMOF structure. [Adapted and reproduced from: [30]] a) top view; b) side view. The graphs were kindly created by Simon Spiegel at the Institute of Functional Interfaces, KIT. . . . . 40

List of Figures

4.10	a) Schematic illustration of the applied robotic set-up with three dip-cleaning vessel for each reaction solution. The ultrasonication station is not used in this case. [Reproduced from: [30]] b) Exemplary X-Ray Diffractogram of a measured sample compared to a simulated powder diffractogram of the pillar layered DASBDC SURMOF. [Reproduced from: [30]] . . . . .	42
4.11	Evolution of the three performed generations throughout the optimization process of the DASBDC SURMOF. [Reproduced from: [30]] . . . . .	43
4.12	Importance of Variables depicting the significance of the chosen variables in reference to a defined goal. [Reproduced from: [30]] .	44
4.13	a) 3D printed sampleholder for porous Al <sub>2</sub> O <sub>3</sub> -substrates. [Reproduced from: [30]] b) Photograph of the gold coated aluminaoxide substrates. c) SEM recording of the porous Al <sub>2</sub> O <sub>3</sub> -substrates. [Reproduced from: [30]] . . . . .	47
4.14	X-Ray diffractograms of the SURMOFs with the best synthesis conditions in pure ethanol applied on a) gold-coated silicon wafer substrate and b) gold coated aluminaoxide substrate. . . . .	48
4.15	This figure illustrates the full conversion process from a)SURMOF to b) the crosslinked SURMOF to c) the SURGEL where the metal nodes are removed. [All three figures adapted and reproduced from: [30]] The structures were kindly generated by Simon Spiegel at the Institute of Functional Interfaces, KIT. . . . .	48
4.16	a) Exemplary X-Ray diffractograms showcasing the simulated powder pattern (black), the signals of the SURMOF (blue), the cross-linked SURMOF (green) and the absence of signals in the SURGEL after the removal of the metal ions (brown). [Reproduced from: [30]] b) Exemplary IRR-Spectra showcasing the signals of the SURMOF (blue), the crosslinked SURMOF (green) and the SURGEL after the removal of the metal ions (brown). . . . .	49
4.17	a) Permeance of the SURMOF, SURMOF-CL and SURGEL. [Adapted and reproduced from: [30]] b) Respective selectivities of SURMOF, SURMOF-CL and SURGEL. [Adapted and reproduced from: [30]] The figures were kindly provided by Dr. Nicholas Prasetya.	50
4.18	Illustration of the goal to optimize the smoothness of a thin films' surface. . . . .	52
4.19	a) Progression of Fitness Values Through Three Generations of Machine Learning Optimization: Diverse Set (Green Squares), First Genetic Algorithm (GA1) (Blue Triangles), and Second Genetic Algorithm (GA2) (Brown Dots). The same color coding applies to figures b), c), and d). Bars indicate the average fitness value of all experiments in each generation, using corresponding colors. b) Fitness Crystallinity for all experiments. c) Fitness for Roughness for all successful experiments. d) Fitness Crystallinity for all successful experiments. [All four Figures: Adapted and reproduced from [52] - Manuscript accepted for publication]. . . . .	56



## List of Figures

4.20	a) Significance of selected variables (n-Linker, n-Metal, Modulator, Ultrasonication Time, Spray Cleaning Time, n2-Linker, n2-Metal) in all experiments, rated on a scale from 0 to 1. b) Significance of the same selected variables of all experiments in GA1 and GA2, rated on a scale from 0 to 1 [Adapted and reproduced from [52] - Manuscript accepted for publication]. . . . .	57
5.1	a) Fitness values of BO (red dots) and GA2 (grey squares). b) Fitness of the particle size against the fitness of the PdI.[Adapted and reproduced from [89]] . . . . .	62
5.2	Illustration of the selection process for the points chosen in the Diverse Set on a simple two-dimensional example. . . . .	65
5.3	Reaction schema of the synthesis of the DASBDC pillar layered SURMOF. . . . .	70
5.4	Unified MOF segment starting with queries to guide the user to the correct input mask.[Screenshot from Chemotion Repository [90]] .	71
5.5	Section for materials preparation and solvent definition.[Screenshot from Chemotion Repository [90]] . . . . .	71
5.6	Defining a synthesis cycle for the dipping robot synthesis in Chemotion.[Screenshot from Chemotion Repository [90]] . . . . .	72
5.7	Section for substrate choice and substrate preparation for SURMOFs.[Screenshot from Chemotion Repository [90]] . . . . .	72
6.1	Illustrative summary of all projects undertaken throughout this thesis. ML Optimizations (upper) from left to right: Orientation Optimization [Adapted and reproduced from [28]], Crystallinity Optimization for Gasseparation [Adapted and reproduced from: [30]], Roughness Optimization [Adapted and reproduced from [52] - Manuscript accepted for publication]; Softwaretoolsets for AI Integration (lower) from left to right: Comparison of GA and BO towards nanoparticle size, Adaptation of the SyCoFinder to handle categorical variables [SyCoFinder Logo reproduced from [42]], Extension of Chemotion by developing an input mask for automated robotic synthesis [Robotic schema reproduced from [30], Screenshot from Chemotion Repository [90]]. . . . .	75
A.1	X-Ray Diffractograms of the Diverse Set for Parametercombinations 1 to 3.[Adapted and reproduced from [28]] . . . . .	109
A.2	X-Ray Diffractograms of the Diverse Set for Parametercombinations 4 to 6.[Adapted and reproduced from [28]] . . . . .	110
A.3	X-Ray Diffractograms of the Diverse Set for Parametercombinations 7 to 9.[Adapted and reproduced from [28]] . . . . .	111
A.4	X-Ray Diffractograms of the Diverse Set for Parametercombinations 10.[Adapted and reproduced from [28]] . . . . .	112
A.5	X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 1 to 3.[Adapted and reproduced from [28]] . . . . .	113
A.6	X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 4 to 6.[Adapted and reproduced from [28]] . . . . .	114
A.7	X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 7 to 9.[Adapted and reproduced from [28]] . . . . .	115

List of Figures

A.8	X-Ray Diffractograms of the Genetic Algorithm 1 for Parameter-combinations 10.[Adapted and reproduced from [28]] . . . . .	116
A.9	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 1 to 3.[Adapted and reproduced from [28]] . . . . .	117
A.10	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 4 to 6.[Adapted and reproduced from [28]] . . . . .	118
A.11	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 7 to 9.[Adapted and reproduced from [28]] . . . . .	119
A.12	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 10.[Adapted and reproduced from [28]] . . . . .	120
A.13	X-Ray Diffractograms of the Diverse Set for Parametercombinations 1 to 5.[Reproduced from: [30]] . . . . .	124
A.14	X-Ray Diffractograms of the Diverse Set for Parametercombinations 6 to 10.[Reproduced from: [30]] . . . . .	125
A.15	X-Ray Diffractograms of the Genetic Algorithm 1 for Parameter-combinations 1 to 5.[Reproduced from: [30]] . . . . .	126
A.16	X-Ray Diffractograms of the Genetic Algorithm 1 for Parameter-combinations 6 to 10. It is to be noted, that experiment number 8 is missing, since the amount of metal-salt was suggested to zero by the genetic algorithm.[Reproduced from: [30]] . . . . .	127
A.17	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 1 to 10.[Reproduced from: [30]] . . . . .	128
A.18	X-Ray Diffractograms of the Diverse Set for Parametercombinations 1 to 5.[Reproduced from [52] - Manuscript accepted for publication] . . . . .	132
A.19	X-Ray Diffractograms of the Diverse Set for Parametercombinations 6 to 10.[Reproduced from [52] - Manuscript accepted for publication] . . . . .	133
A.20	X-Ray Diffractograms of the Genetic Algorithm 1 for Parameter-combinations 1 to 5.[Reproduced from [52] - Manuscript accepted for publication] . . . . .	134
A.21	X-Ray Diffractograms of the Genetic Algorithm 1 for Parameter-combinations 6 to 10.[Reproduced from [52] - Manuscript accepted for publication] . . . . .	135
A.22	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 1 to 5.[Reproduced from [52] - Manuscript accepted for publication] . . . . .	136
A.23	X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter-combinations 6 to 10.[Reproduced from [52] - Manuscript accepted for publication] . . . . .	137
A.24	X-Ray Diffractograms of the second generation of GA Optimization (GA2) for ZIF-8. . . . .	140
A.25	X-Ray Diffractograms of the Bayesian Optimization for ZIF-8. . . . .	141

# List of Tables

4.1	Definition of variables and their respective limits in which they were varied.[Reproduced from [28]] . . . . .	32
4.2	All successful synthesis parameter sets with their respective fitness values ordered by decreasing fitness.[Reproduced from [28]] . . .	37
4.3	Definition of variables and their respective ranges in which they are to be varied.[Reproduced from: [30]] . . . . .	41
4.4	Parameter sets and according fitness values of all successfully rated experiments ordered by decreasing fitness values.[Reproduced from: [30]] . . . . .	46
4.5	Peak assignment of the relevant signals in the IRRA-Spectra. . . .	49
4.6	The seven parameters chosen to be varied and their according ranges and cycles to which they are applied.[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	53
4.7	Parameter sets and according fitness values of all successful experiments in GA1 and GA2 ordered by decreasing fitness values.[Adapted and reproduced from [52] - Manuscript accepted for publication] .	58
5.1	Definition of variables and their respective limits in which they were varied. . . . .	61
5.2	Fitness terms of BO. [Adapted and reproduced from [89]] . . . . .	61
5.3	Fitness terms of GA2. [Adapted and reproduced from [89]] . . . . .	61
5.4	Parameter sets of the two best syntheses of GA2 and BO. . . . .	62
5.5	. . . . .	65
5.6	. . . . .	65
5.7	Encoding options to assign numerical values to strings. . . . .	66
5.8	Processing the returned Diverse Set to its final output form. . . . .	67
A.1	All parameter combinations and their corresponding fitness value of the Diverse Set (DS).[Reproduced from [28]] . . . . .	105
A.2	All parameter combinations and their corresponding fitness value of the first genetic algorithm (GA1).[Reproduced from [28]] . . .	105
A.3	All parameter combinations and their corresponding fitness value of the second genetic algorithm (GA2).[Reproduced from [28]] . .	106
A.4	Fitness values of single fitness criteria and overall fitness for the Diverse Set.[Reproduced from [28]] . . . . .	106
A.5	Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 1.[Reproduced from [28]] . . . . .	107
A.6	Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 2.[Reproduced from [28]] . . . . .	107

*List of Tables*

A.7	All parameter combinations and their corresponding fitness value of the Diverse Set (DS).[Reproduced from: [30]] . . . . .	120
A.8	All parameter combinations and their corresponding fitness value of the first genetic algorithm (GA1).[Reproduced from: [30]] . . .	121
A.9	All parameter combinations and their corresponding fitness value of the second genetic algorithm (GA2).[Reproduced from: [30]] .	121
A.10	Fitness values of single fitness criteria and overall fitness for the Diverse Set.[Reproduced from: [30]] . . . . .	122
A.11	Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 1.[Reproduced from: [30]] . . . . .	122
A.12	Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 2.[Reproduced from: [30]] . . . . .	123
A.13	All parameter combinations and their corresponding fitness value of the Diverse Set (DS).[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	129
A.14	All parameter combinations and their corresponding fitness value of the Genetic Algorithm (GA1).[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	129
A.15	All parameter combinations and their corresponding fitness value of the Genetic Algorithm (GA2).[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	130
A.16	Fitness values of single fitness criteria and overall fitness for the Diverse Set.[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	130
A.17	Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 1.[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	131
A.18	Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 2.[Adapted and reproduced from [52] - Manuscript accepted for publication] . . . . .	131
A.19	Provided training dataset from GA Optimization for BO.[Adapted and reproduced from [89]] . . . . .	138
A.20	Suggested Parametersets and Fitness Evaluation Values of BO.[Adapted and reproduced from [89]] . . . . .	139

## 9

# Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
BO	Bayesian Optimization
BTC	1,3,5-benzenetricarboxylate
CL	Crosslinked
DABCO	1,4-Diazabicyclo[2.2.2]octane
DASBDC	2,2'-Diazido-4,4'-stilbenedicarboxylic acid
DLS	Dynamic Light Scattering
DMF	Dimethylformamid
ELN	Electronic Lab Notebook
Eq	Equivalent
EtOH	Ethanol
FAIR	Findable, Accessible, Interoperable, Reusable
GA	Genetic Algorithm
GC	Gas Chromatograph
HKUST-1	Hong-Kong University of Science and Technology [Cu <sub>3</sub> (btc)]
IR	Infrared
IRRAS	Infrared Reflection-Absorption Spectroscopy
LBL	Layer-by-Layer
MeOH	Methanol
ML	Machine Learning
MOF	Metal-Organic Framework
MSE	Mean Squared Error
NP	Nanoparticle
PCP	Porous Coordination Polymer
RDM	Research Data Management
RT	Room Temperature
SAM	Self-Assembled Monolayer
SE	Secondary Electron

## 9. ABBREVIATIONS

<b>Abbreviation</b>	<b>Meaning</b>
SEM	Scanning Electron Microscopy
SURMOF	Surface Anchored Metal-Organic Framework
SyCoFinder	SynthesisConditionFinder
US	Ultrasonication
XRD	X-Ray Diffraction
ZIF	Zeolitic Imidazole Framework

# Acknowledgements

## **Institutional and Resource Acknowledgments**

I would like to acknowledge the assistance of ChatGPT by OpenAI, which provided valuable support in refining the language and clarity of some parts of the text within this thesis. I also extend my gratitude to the graduate school BIF-IGS for offering engaging courses in Python programming, project management, and more, as well as for organizing highly interesting retreats during my PhD time. Furthermore, my sincere appreciation goes to the KHYS (Karlsruhe House of Young Scientists) for awarding me the Networking Grant. This grant afforded me the invaluable opportunity to undertake a highly enriching stay abroad at UC San Diego.

## **Personal Acknowledgments**

I would first of all like to thank my supervisor PD Manuel Tsotsalas for the motivation and incentive over three amazing years of doctoral studies, for the exceptional support and the opportunity to participate in several conferences and supporting and encouraging my projects abroad, as well as the right advice at the right time. Further thanks go to Prof. Stefan Bräse for taking over the co-referral of my doctoral thesis and for a fruitful collaboration. I would also like to thank Prof. Christof Wöll for his advice, motivating input and interesting discussions as well as for the friendly welcome at the Institute of Functional Interfaces.

A special thanks to my colleagues in the Tsotalas research group: Ilona (Everyone needs an Ilona), Simon (ARRRRRRR), Roy, Yixuan and Amirali for mutual support, friendship, sharing and help, amusing lab clean-ups, endless discussions, coffee, amazing conference visits and most of all for making hard working days bearable and fun.

I would like to thank all the members of the IFG for their help, support and advice. Special thanks go to Hartmut Gliemann for always being patient and supportive, always keeping a cool head and for taking care of every little ache and pain, to Peter Weidler for making fun of me, for fixing my computer (sooooo often), for always fixing the XRD, which of course always fails when there is absolutely no time, and for sharing his overwhelming knowledge, Carsten Natzeck and Jonas Wohlgemuth for friendship, honesty and always cheering me up and for a great atmosphere around the robot projects as well as good organization, Angela Weiss, without whom the institute would collapse, Alex Welle for his honest attitude, his absolutely awesome sarcasm (I really appreciate this!) and for advice throughout my whole PhD, Matthias Schwotzer for his infinite patience in explaining SEM, his endurance in supporting my projects during the recording of the most difficult

## 10. ACKNOWLEDGEMENTS

samples and fun around the topic of ducks, Nicole Jung for advice throughout my whole doctoral studies, a lot of knowledge and understanding and above all for inspiration, Stefan Heißler for his open ear, his support not only in matters of IR and his valuable suggestions and advice, to Ilona Wagner for really taking care of so many organizational things and that she always shares her knowledge so selflessly, for cozy library sessions, for sharing dreams of vacations and our lively exchanges about everything under the sun in the office, to Simon Spiegel for coffee, talks, coffee, funny conference trips, coffee, killing that unbearable red wine, coffee, dancing bananas and coffee, to Meike König for her patience and her experience in ellipsometry, which she explained to me again and again, to Nina Scheuermann, who taught me everything around the robots and was always available when something needed to be repaired 'quickly', to Anemar Bruno Kanj for lab company, silent hoods, friendship, gossip, whiskey, beer, coffee and lots of fun, to Katharina Bleher for the fastest hike on an excursion and super interesting discussions about repairing mountain bikes and motorcycles, to Nicholaus Praysetya for the smooth collaboration, the super nice exchange and his openness, to Erik Koch and Luca Oswald for listening and taking my advice to heart and the very funny lab sessions - I really appreciated putting the robots in your hands - and finally to Simon Oßwald from IOC for providing the necessary substances for one of my projects, his patience for the publication and our exchange.

During my PhD, I had the opportunity to collaborate with several groups, where I enjoyed outstanding experiences and received exceptional support. I want to express my gratitude to Dr. Nicole Jung, Pei-Chi Huang (Peggy), and Chia-Lin Lin (Claire) from IBCS-FMS who were always immediately available for support regarding Chemotion and for the fantastic collaboration. Moreover, I want to thank Dr. Felix Dietrich, Luisa Ortner, and Josef Mayr at the TUM for their amazing work and efforts on the BO project. Thanks also to Prof. Seth M. Cohen for welcoming me at UCSD, and to his group members Prantik, Conor, Alysia, Kyle, Sharon, Ryjul, Grant, and Jeewon for the heartwarming welcome and integration into the group, as well as the best margaritas on the beach ever. Furthermore, special thanks to Prof. Mohamad Moosavi for hosting me at UofT, getting me really into coding, and our super interesting collaboration. Additional thanks go to his group members Mehrad, Sartaaaj, and Rooz for very honest but also very funny office talks and culinary adventures - I hope the 'Lena-bowl' is refilled frequently.

I would also like to thank Jasper, Hella, Max, Ilona, Peter, Carsten and Stefan for carefully proofreading my thesis. Everyone knows how absolutely indispensable that is.

But a PhD wouldn't be possible without the people who are there when you leave the institute, need to clear your head, and have supported you in all possible ways over the years.

Jasper, Mart, Julia, Alex and Fabi - you are the best flatmates ever. Thank you for being such an unusual and amazing home, for sharing motorcycle rides and holidays, mechanic's sessions, cooking, barbecues and everything else we have done and will do together over the last few years and in future and for bringing in kilotons of tomatoes to keep me writing.



## 10. ACKNOWLEDGEMENTS

Special thanks go to Mart for his deep friendship and always logical approach to finding solutions, for discussions and stolen time. Nom.

Moreover, I want to deeply thank my parents, who are always a warm, cozy home for me, a home where I can always come and hide from the world, where I can relax and where I am pampered and loved, for their endless support and for always having good advice at hand for all situations in life. Thanks for the late night martinis and wine, the great food, no matter how late it gets, tomatoes, chocolate raisins and the 'TamTam' for my inner master of ceremonies.

Deep thanks also go to Sofia, for being the best little sister in the world, for the last experiment, for staying up with me and walking the streets together, when I was abroad, for sharing my experiences and calming me down, for trust and love. HDIDSLWDMUDGFATADWDNNGH! :-\*

My heartfelt thanks go to Hella, for being a refuge place, for keys, a bed I can always crash on, free garages in the middle of an expensive city, cat pets (awww), great girls' nights and for now correcting the third and hopefully last thesis in my life - I can't believe you really went through with this three times and are still with me. But that doesn't begin to describe how important you are to me - thank you for everything - above all for your friendship.

Lastly I want to thank Jasper. You are the best thing that ever happened to me, and I'm pretty sure I wouldn't be at this point in my life without you. Thank you for always being by my side, for breaking free, for coming back, for all the adventures we have had and still have ahead of us, and above all for your endless love. I could use many more words, and yet they wouldn't be adequate to express how I feel about you, so to simply sum it up: AWGUNVM!

## APPENDIX

## A.1 MOF thin film Optimization towards Orientation

### A.1.1 Synthesis Parameter

#### Diverse Set

Tab. A.1: All parameter combinations and their corresponding fitness value of the Diverse Set (DS).[Reproduced from [28]]

Number	Ultra-sonication [s]	Modulator (Water) [mL]	n (Linker) [mmol]	n (Metal-salt) [mmol]	Spray [s]	Fitness
1	0	0.0	0.02	0.02	0	0.00
2	100	40.0	10.00	10.00	5	0.11
3	0	0.0	5.01	10.00	5	0.18
4	100	0.0	10.00	0.02	3	0.78
5	100	20.0	0.02	10.00	0	0.00
6	0	40.0	10.00	5.01	0	0.00
7	60	40.0	0.02	0.02	5	0.00
8	60	0.0	10.00	10.00	0	0.48
9	100	0.0	0.02	5.01	5	0.00
10	0	20.0	10.00	0.02	5	0.00

#### Genetic Algorithm 1

Tab. A.2: All parameter combinations and their corresponding fitness value of the first genetic algorithm (GA1).[Reproduced from [28]]

Number	Ultra-sonication [s]	Modulator (Water) [mL]	n (Linker) [mmol]	n (Metal-salt) [mmol]	Spray [s]	Fitness
11	100	12.4	0.02	8.96	4	0.00
12	0	1.7	6.53	0.95	0	0.00
13	25	0.0	6.59	8.92	4	0.14
14	37	0.0	0.38	1.67	0	0.00
15	60	0.0	10.00	10.00	0	0.00
16	89	41.8	4.39	3.40	6	0.00
17	56	0.1	0.02	0.02	3	0.00
18	100	0.0	9.32	2.77	5	0.71
19	60	40.0	0.02	0.02	5	0.00
20	95	30.4	8.87	5.01	5	0.00

**Genetic Algorithm 2**

Tab. A.3: All parameter combinations and their corresponding fitness value of the second genetic algorithm (GA2).[Reproduced from [28]]

Number	Ultra-sonication [s]	Modulator (Water) [mL]	n (Linker) [mmol]	n (Metal-salt) [mmol]	Spray [s]	Fitness
21	92	39.8	7.74	3.52	5	0.00
22	82	0.0	9.74	4.95	2	0.17
23	25	0.0	6.59	8.92	4	0.54
24	77	0.0	3.50	0.94	4	0.00
25	73	40.3	4.19	2.65	5	0.00
26	96	0.1	3.45	2.08	4	0.00
27	100	0.0	7.25	0.02	3	0.84
28	85	8.6	1.00	3.40	5	0.00
29	94	30.0	7.27	5.66	5	0.00
30	37	0.0	0.38	1.67	0	0.00

**A.1.2 Fitness****Diverse Set**

Tab. A.4: Fitness values of single fitness criteria and overall fitness for the Diverse Set.[Reproduced from [28]]

Number	Phase Identity	Crystallinity [%]	111-Preference [%]	Fitness
1	1	40.0	0.0	0.00
2	1	23.0	49.0	0.11
3	1	60.0	30.0	0.18
4	1	78.0	100.0	0.78
5	1	65.0	0.0	0.00
6	1	42.0	0.0	0.00
7	1	59.0	0.0	0.00
8	1	63.0	76.0	0.48
9	1	50.0	0.0	0.00
10	1	72.0	0.0	0.00

**Genetic Algorithm 1**

Tab. A.5: Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 1.[Reproduced from [28]]

Number	Phase Identity	Crystallinity [%]	111-Preference [%]	Fitness
11	1	93.0	0.0	0.00
12	0	86.0	0.0	0.00
13	1	89.0	16.0	0.14
14	1	85.0	0.0	0.00
15	1	86.0	0.0	0.00
16	1	89.0	0.0	0.00
17	1	77.0	0.0	0.00
18	1	90.0	79.0	0.71
19	1	96.0	0.0	0.00
20	1	96.0	0.0	0.00

**Genetic Algorithm 2**

Tab. A.6: Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 2.[Reproduced from [28]]

Number	Phase Identity	Crystallinity [%]	111-Preference [%]	Fitness
21	1	95.0	0.0	0.00
22	1	88.0	20.0	0.17
23	1	89.0	61.0	0.54
24	1	89.0	0.0	0.00
25	1	92.0	0.0	0.00
26	1	88.0	0.0	0.00
27	1	84.0	100.0	0.84
28	1	91.0	0.0	0.00
29	1	94.0	0.0	0.00
30	1	77.0	0.0	0.00

**A.1.3 Possible program sequences emerging from the robotic set-up and the suggested parameters.**

[Reproduced from [52] - Manuscript accepted for publication]

1. Case: No Spray-parameter, no Ultrasonication-parameter:  
Metalsalt-solution - Dip-cleaning - Linker-solution - Dipcleaning
2. Case: Spray-parameter and Ultrasonication-parameter:  
Metalsalt-solution - Spray-cleaning - Ultrasonication-cleaning - Linker-solution - Spray-cleaning - Dipcleaning

## *Appendix*

3. Case: No Spray-parameter, but Ultrasonication-parameter:  
Metalsalt-solution - Ultrasonication-cleaning - Dip-cleaning - Linker-solution  
- Ultrasonication-cleaning - Dip-cleaning
4. Case: Only Spray-parameter, no Ultrasonication-parameter:  
Metalsalt-solution - Spray-cleaning - Dip-cleaning - Linker-solution - Spray-  
cleaning - Dip-cleaning

### A.1.4 X-Ray Diffraction

#### Diverse Set

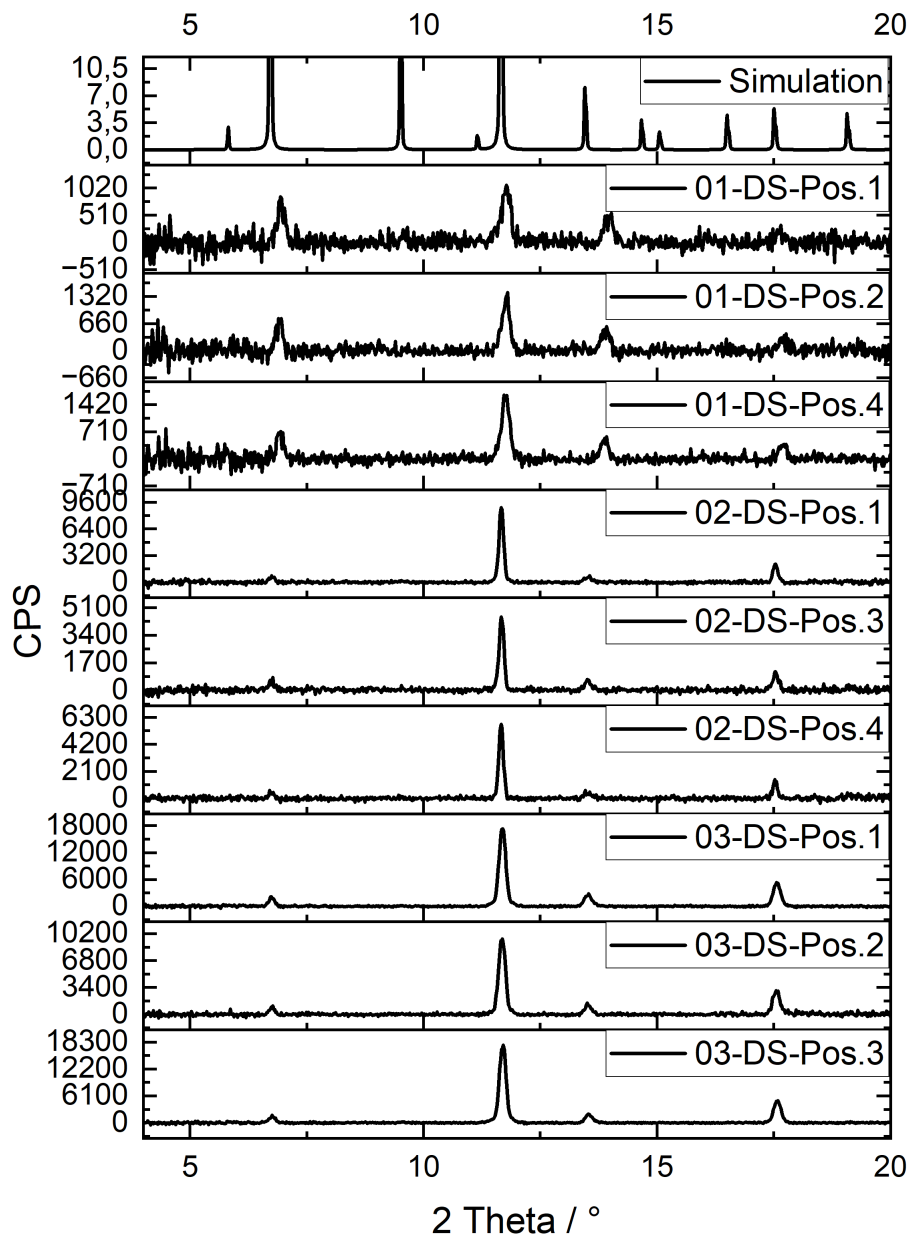


Fig. A.1: X-Ray Diffractograms of the Diverse Set for Parametercombinations 1 to 3.[Adapted and reproduced from [28]]

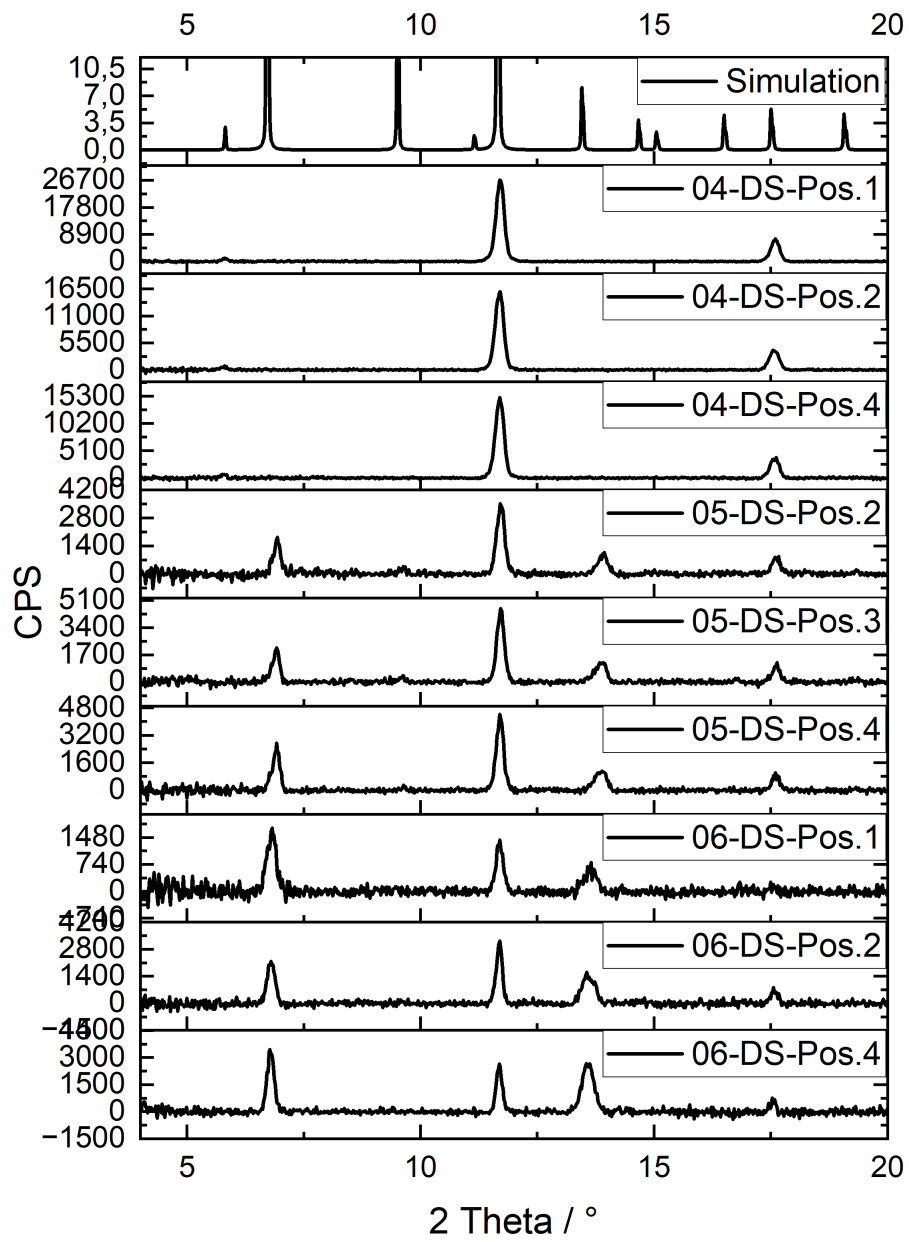


Fig. A.2: X-Ray Diffractograms of the Diverse Set for Parametercombinations 4 to 6.[Adapted and reproduced from [28]]



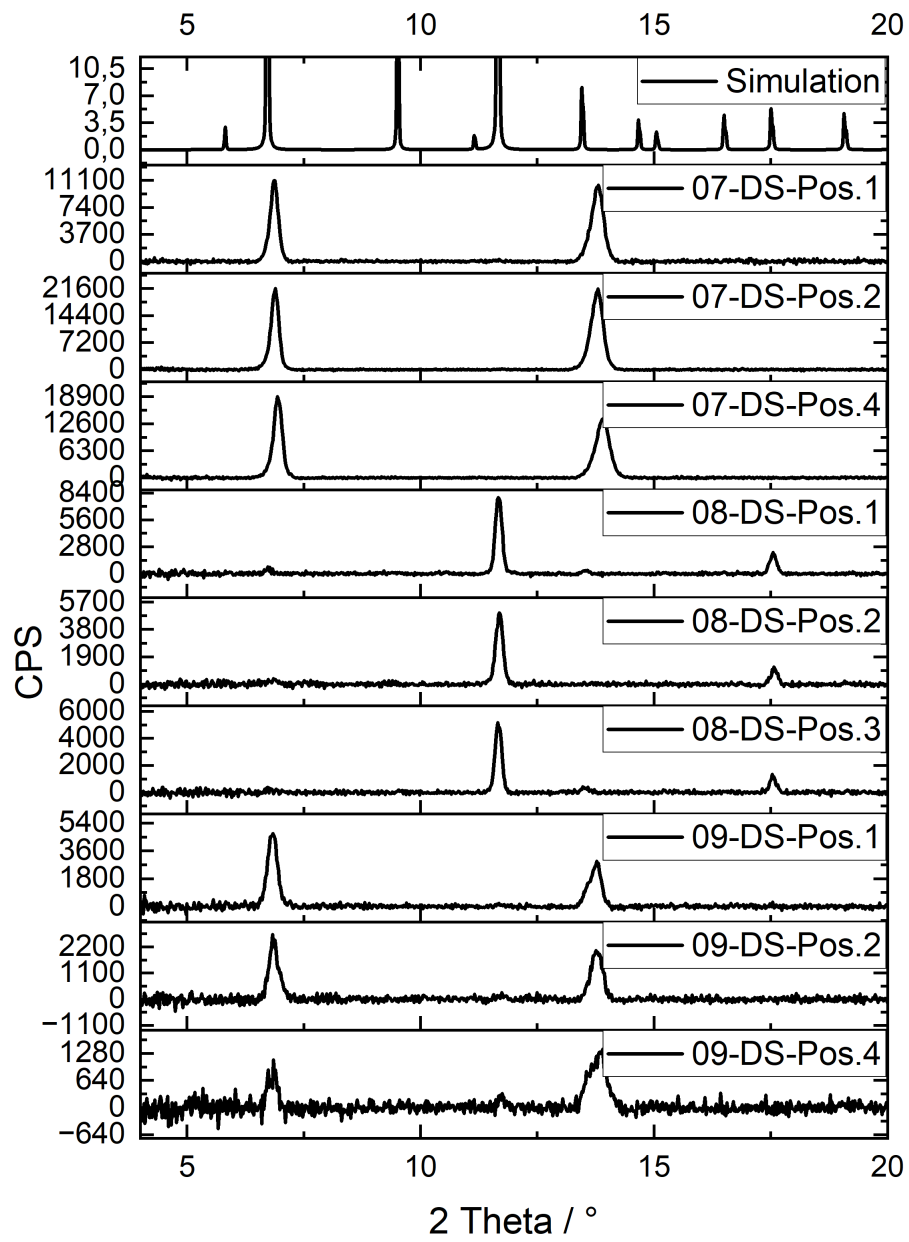


Fig. A.3: X-Ray Diffractograms of the Diverse Set for Parametercombinations 7 to 9.[Adapted and reproduced from [28]]

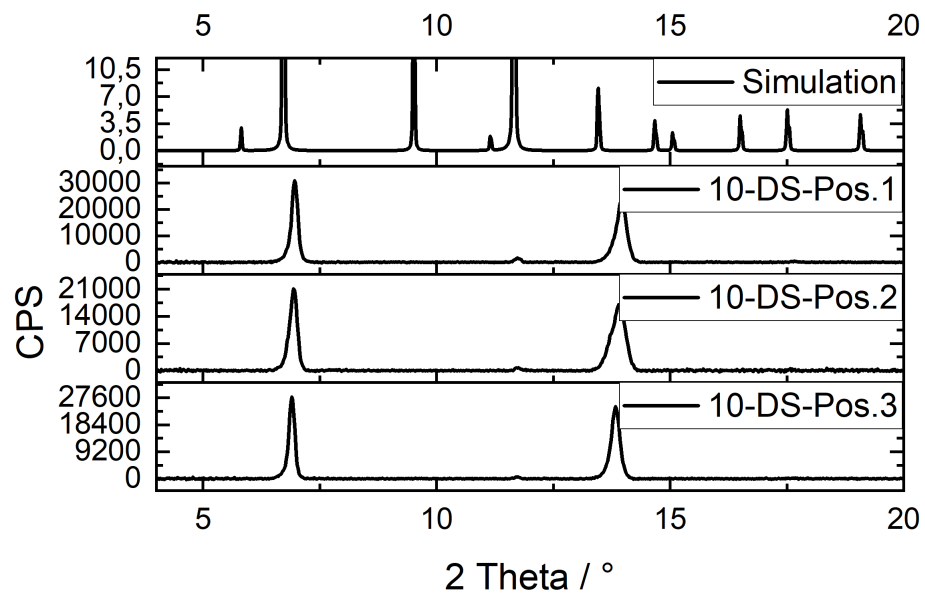


Fig. A.4: X-Ray Diffractograms of the Diverse Set for Parametercombinations 10.[Adapted and reproduced from [28]]

Genetic Algorithm 1

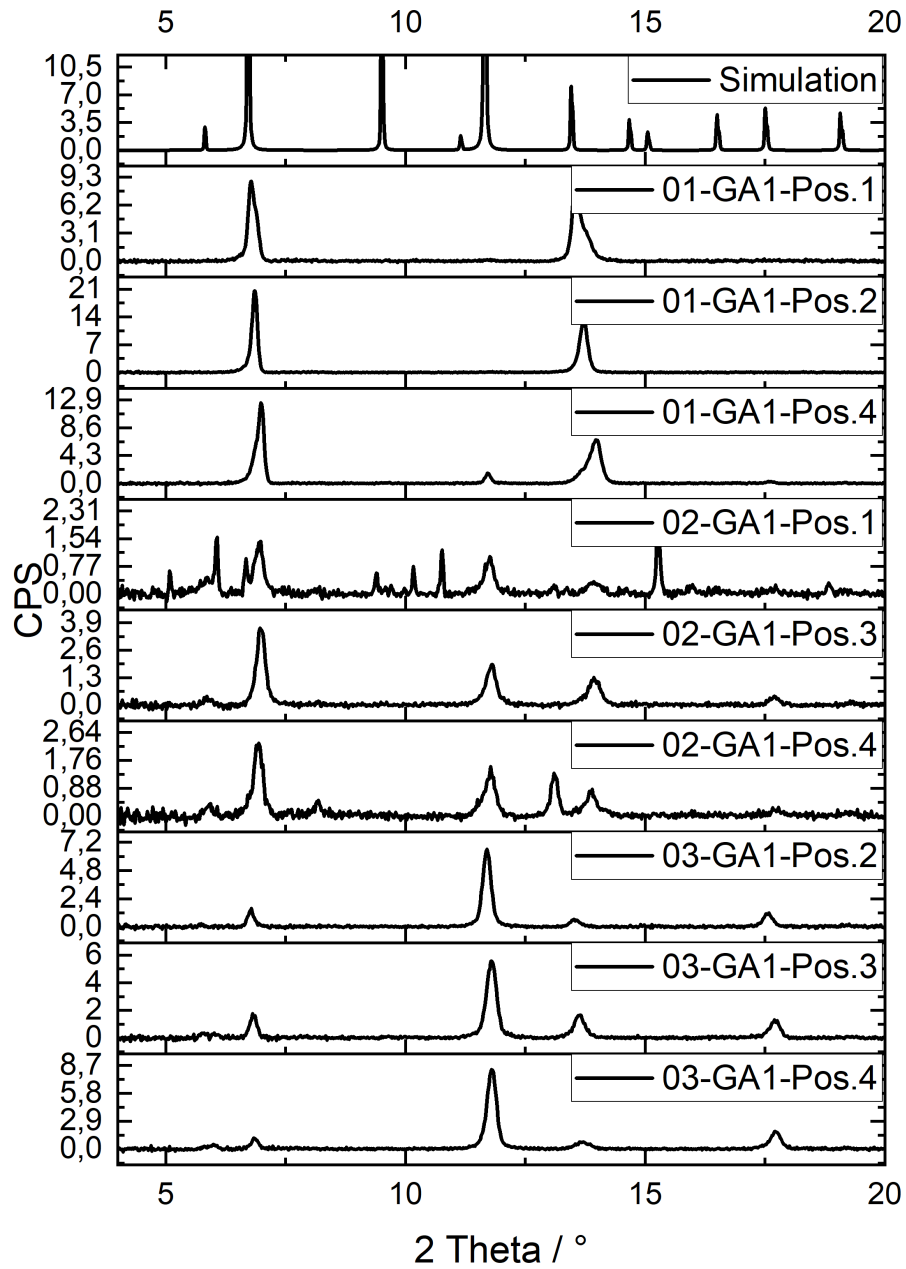


Fig. A.5: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 1 to 3.[Adapted and reproduced from [28]]

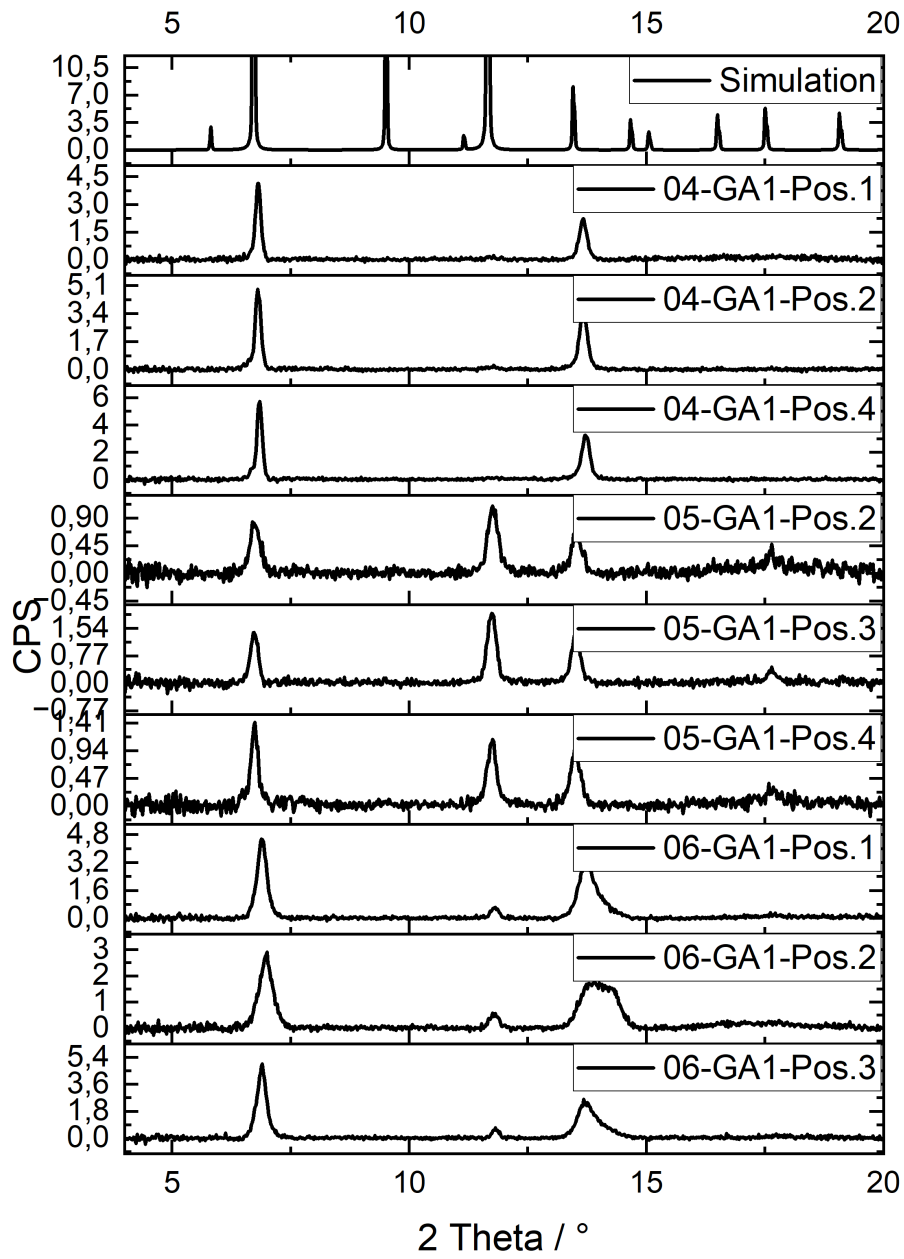


Fig. A.6: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 4 to 6.[Adapted and reproduced from [28]]

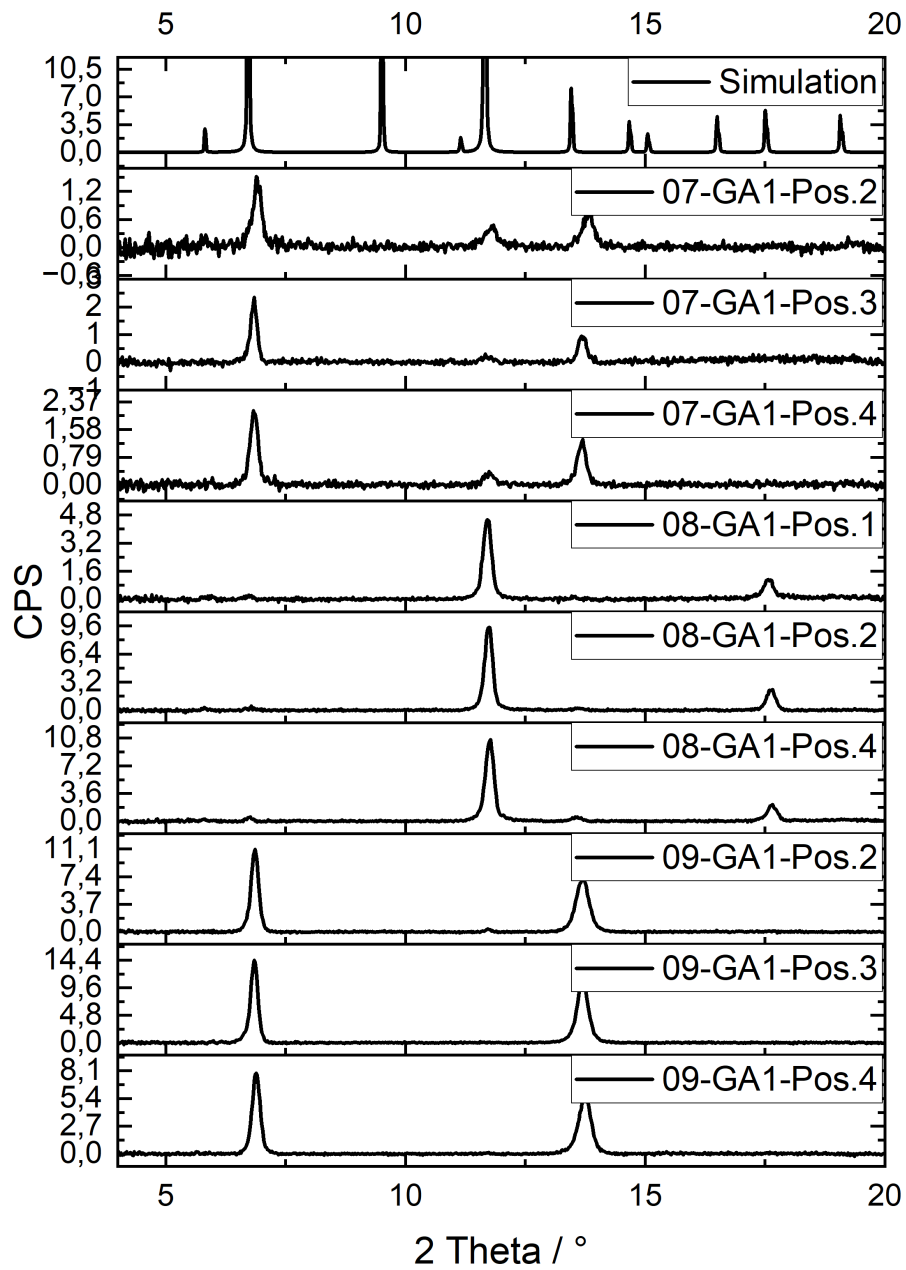


Fig. A.7: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 7 to 9.[Adapted and reproduced from [28]]

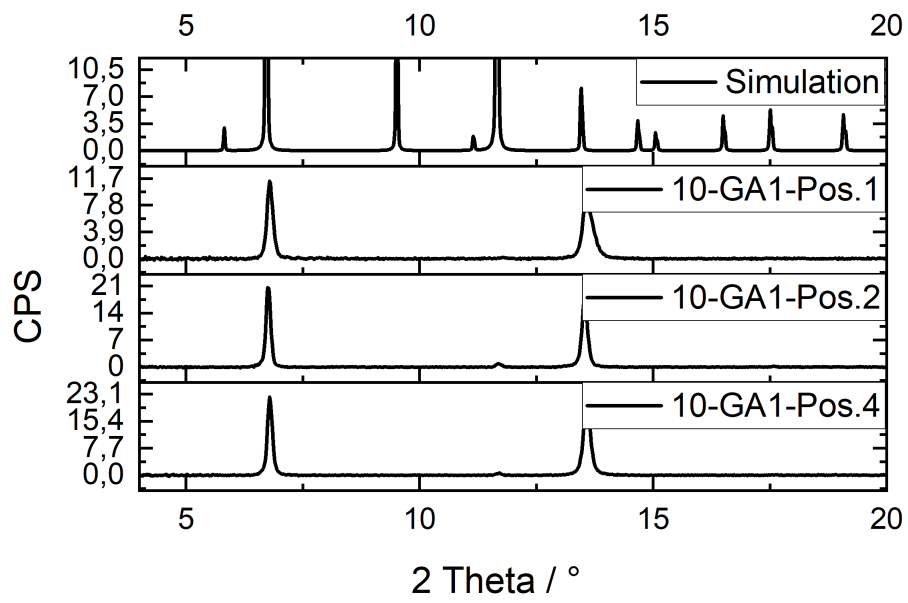


Fig. A.8: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 10.[Adapted and reproduced from [28]]

Genetic Algorithm 2

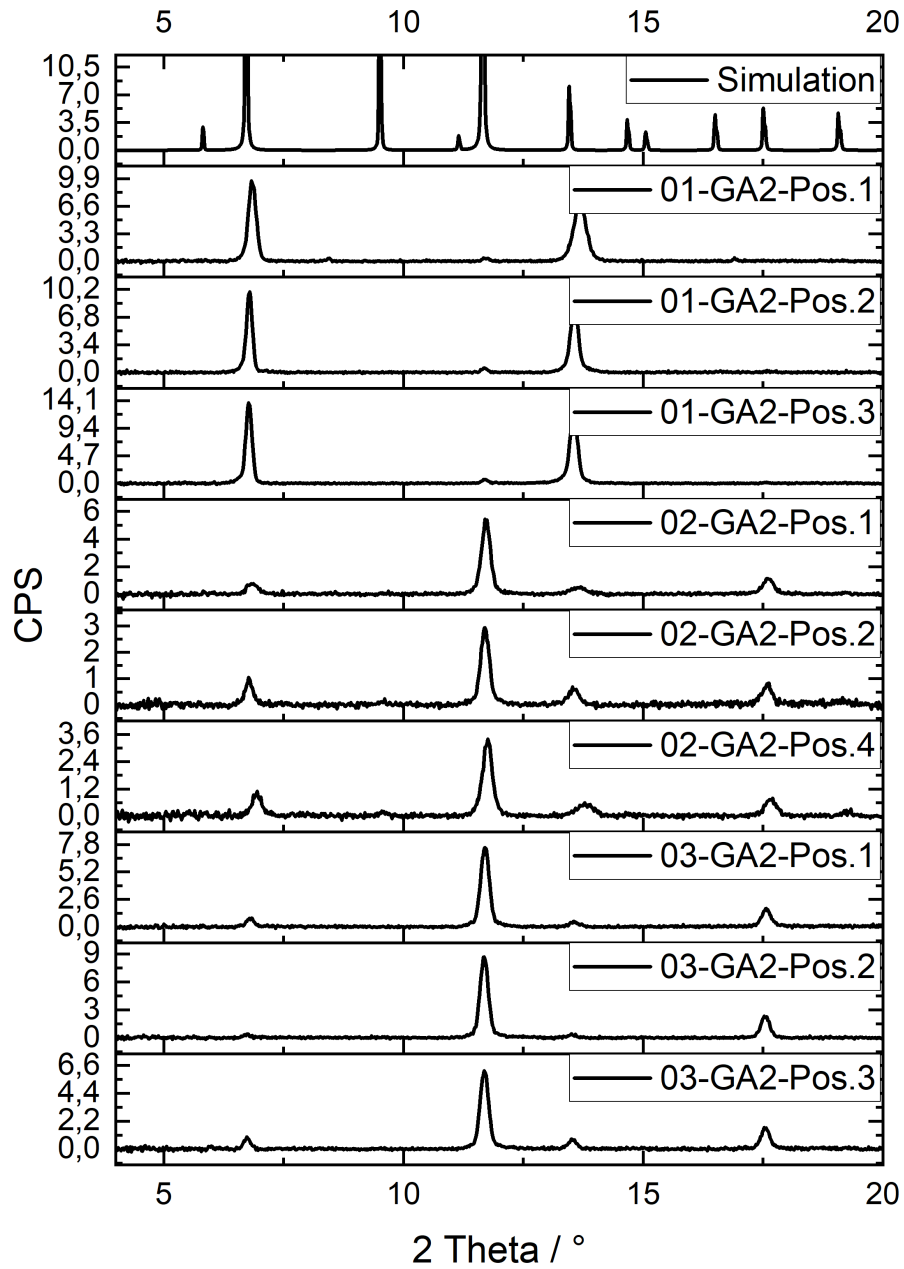


Fig. A.9: X-Ray Diffractograms of the Genetic Algorithm 2 for Parametercombinations 1 to 3.[Adapted and reproduced from [28]]

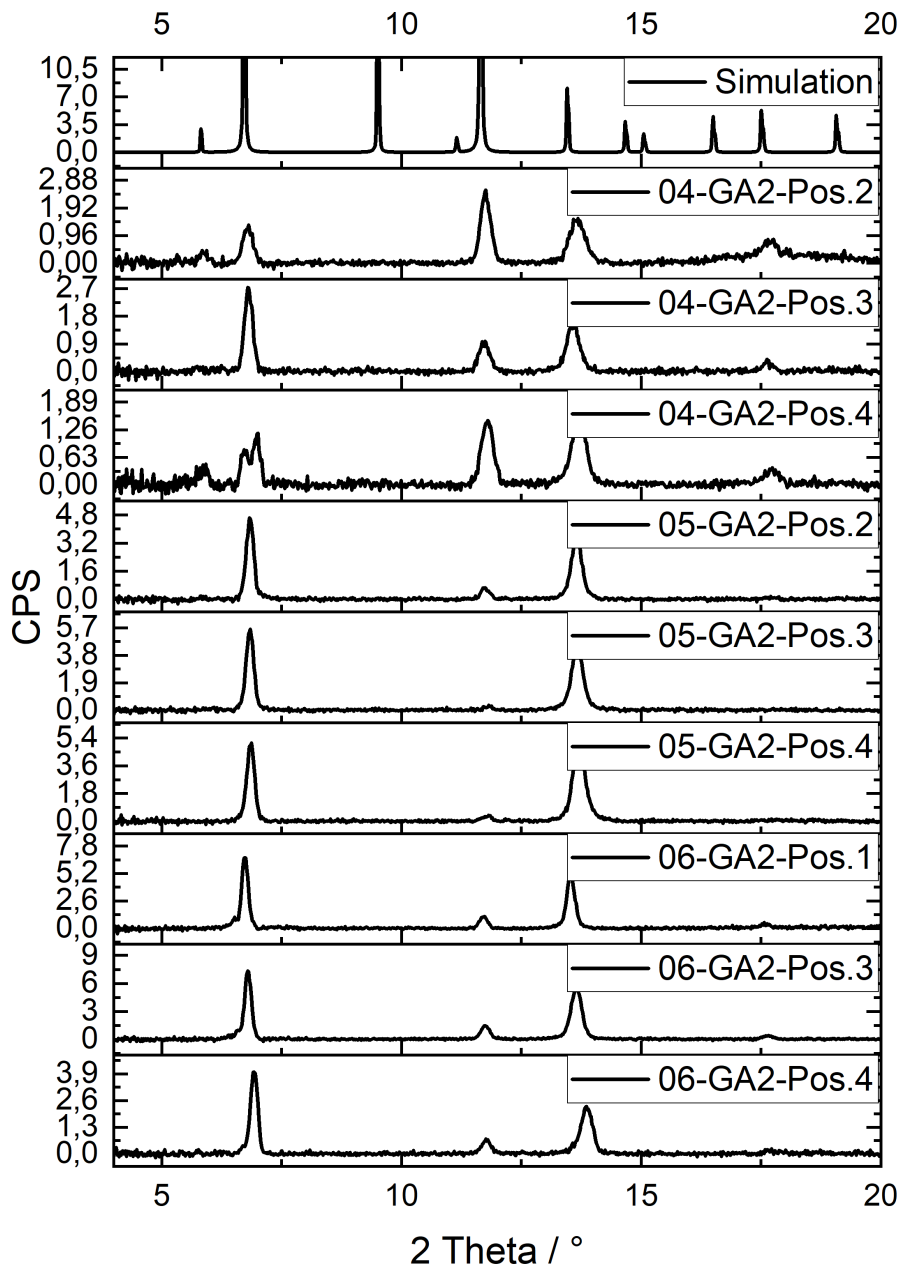


Fig. A.10: X-Ray Diffractograms of the Genetic Algorithm 2 for Parametercombinations 4 to 6.[Adapted and reproduced from [28]]



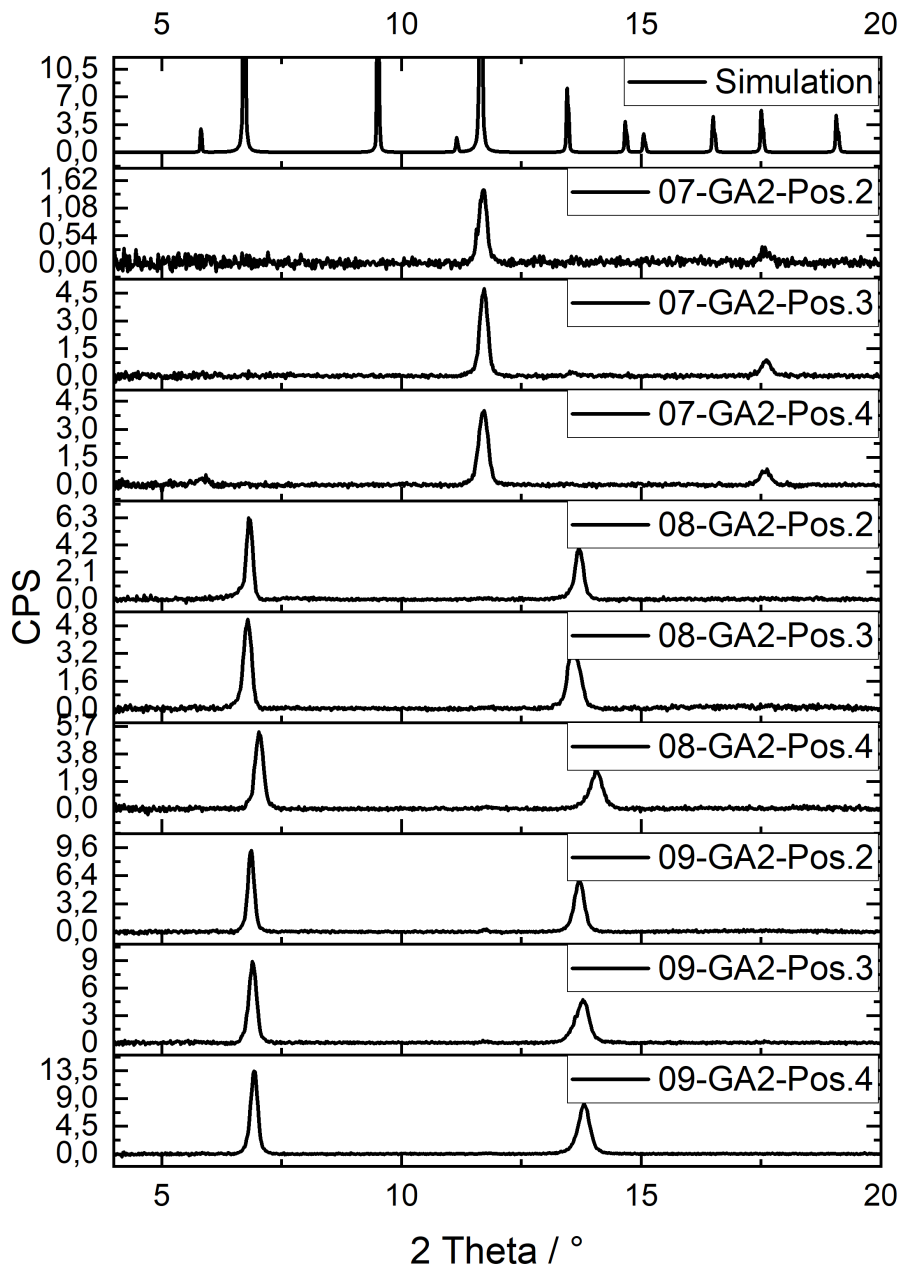


Fig. A.11: X-Ray Diffractograms of the Genetic Algorithm 2 for Parametercombinations 7 to 9.[Adapted and reproduced from [28]]

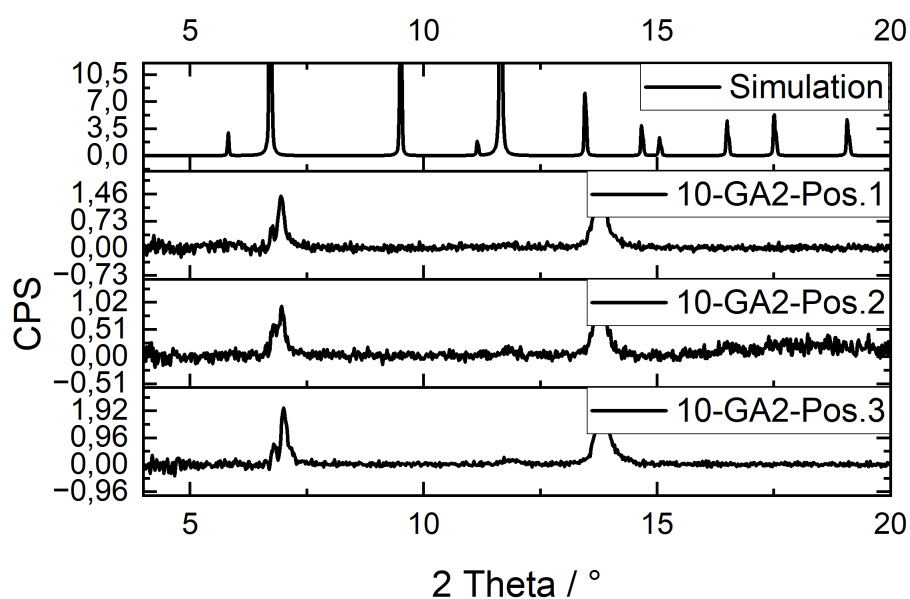


Fig. A.12: X-Ray Diffractograms of the Genetic Algorithm 2 for Parametercombinations 10.[Adapted and reproduced from [28]]

## A.2 MOF thin film Optimization for Gasseparation Application

### A.2.1 Synthesis Parameter

#### Diverse Set

Tab. A.7: All parameter combinations and their corresponding fitness value of the Diverse Set (DS).[Reproduced from: [30]]

Number	c(Linker) [mM/l]	c(Metal) [mM/l]	Modulator (Water) [ml]	DABCO [eq of Linker]	EtOH [%]	Fitness
1	0.01	0.05	0	0.10	0	0.33
2	1.00	5.00	40	2.00	100	0.48
3	0.01	0.05	20	2.00	100	0.37
4	1.00	0.05	40	0.10	50	0.63
5	1.00	2.52	0	2.00	0	0.00
6	0.01	5.00	40	1.05	0	0.63
7	0.51	5.00	0	0.10	100	0.00
8	0.51	0.05	40	2.00	0	0.00
9	1.00	0.05	0	1.05	100	0.00
10	0.01	2.52	40	0.10	100	0.36

**Genetic Algorithm 1**

Tab. A.8: All parameter combinations and their corresponding fitness value of the first genetic algorithm (GA1).[Reproduced from: [30]]

Number	c(Linker) [mM/l]	c(Metal) [mM/l]	Modulator (Water) [ml]	DABCO [eq of Linker]	EtOH [%]	Fitness
11	0.01	0.05	0	0.25	87	0.00
12	0.85	0.05	2	0.91	16	0.00
13	0.49	1.20	40	0.85	23	0.00
14	1.00	5.00	40	2.00	100	0.89
15	0.97	4.39	36	0.61	66	0.00
16	0.85	4.70	40	0.99	100	0.00
17	0.84	0.05	40	1.43	26	0.93
18	0.23	0.00	12	0.85	48	0.00
19	0.33	1.64	40	1.78	0	0.00
20	0.78	5.00	36	1.87	100	0.72

**Genetic Algorithm 2**

Tab. A.9: All parameter combinations and their corresponding fitness value of the second genetic algorithm (GA2).[Reproduced from: [30]]

Number	c(Linker) [mM/l]	c(Metal) [mM/l]	Modulator (Water) [ml]	DABCO [eq of Linker]	EtOH [%]	Fitness
21	0.51	5.00	19	1.26	100	0.88
22	0.31	0.33	0	1.36	25	0.00
23	0.70	0.56	40	1.36	24	0.94
24	0.37	0.83	38	1.23	52	0.00
25	0.60	2.35	22	0.74	49	0.92
26	0.01	0.05	0	0.25	87	0.00
27	0.87	5.00	40	2.00	100	0.83
28	0.78	5.00	36	1.87	100	0.91
29	0.58	0.05	7	0.32	78	0.00
30	0.92	4.89	15	1.47	91	0.87

**A.2.2 Fitness****Diverse Set**

Tab. A.10: Fitness values of single fitness criteria and overall fitness for the Diverse Set.[Reproduced from: [30]]

Number	Phase Identity	Crystallinity [%]	Fitness
1	1	32.7	0.33
2	1	47.6	0.48
3	1	37.3	0.37
4	1	63.4	0.63
5	0	0.0	0.00
6	1	62.7	0.63
7	0	0.0	0.00
8	0	0.0	0.00
9	0	0.0	0.00
10	1	35.8	0.36

**Genetic Algorithm 1**

Tab. A.11: Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 1.[Reproduced from: [30]]

Number	Phase Identity	Crystallinity [%]	Fitness
11	0	0.0	0.00
12	0	0.0	0.00
13	0	0.0	0.00
14	1	88.9	0.89
15	0	0.0	0.00
16	0	0.0	0.00
17	1	93.0	0.93
18	0	0.0	0.00
19	0	0.0	0.00
20	1	72.1	0.72

**Genetic Algorithm 2**

Tab. A.12: Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 2.[Reproduced from: [30]]

Number	Phase Identity	Crystallinity [%]	Fitness
21	1	87.6	0.88
22	0	0.0	0.00
23	1	93.8	0.94
24	0	0.0	0.00
25	1	91.6	0.92
26	0	0.0	0.00
27	1	82.7	0.83
28	1	91.0	0.91
29	0	0.0	0.00
30	1	86.5	0.87

### A.2.3 X-Ray Diffraction

#### Diverse Set

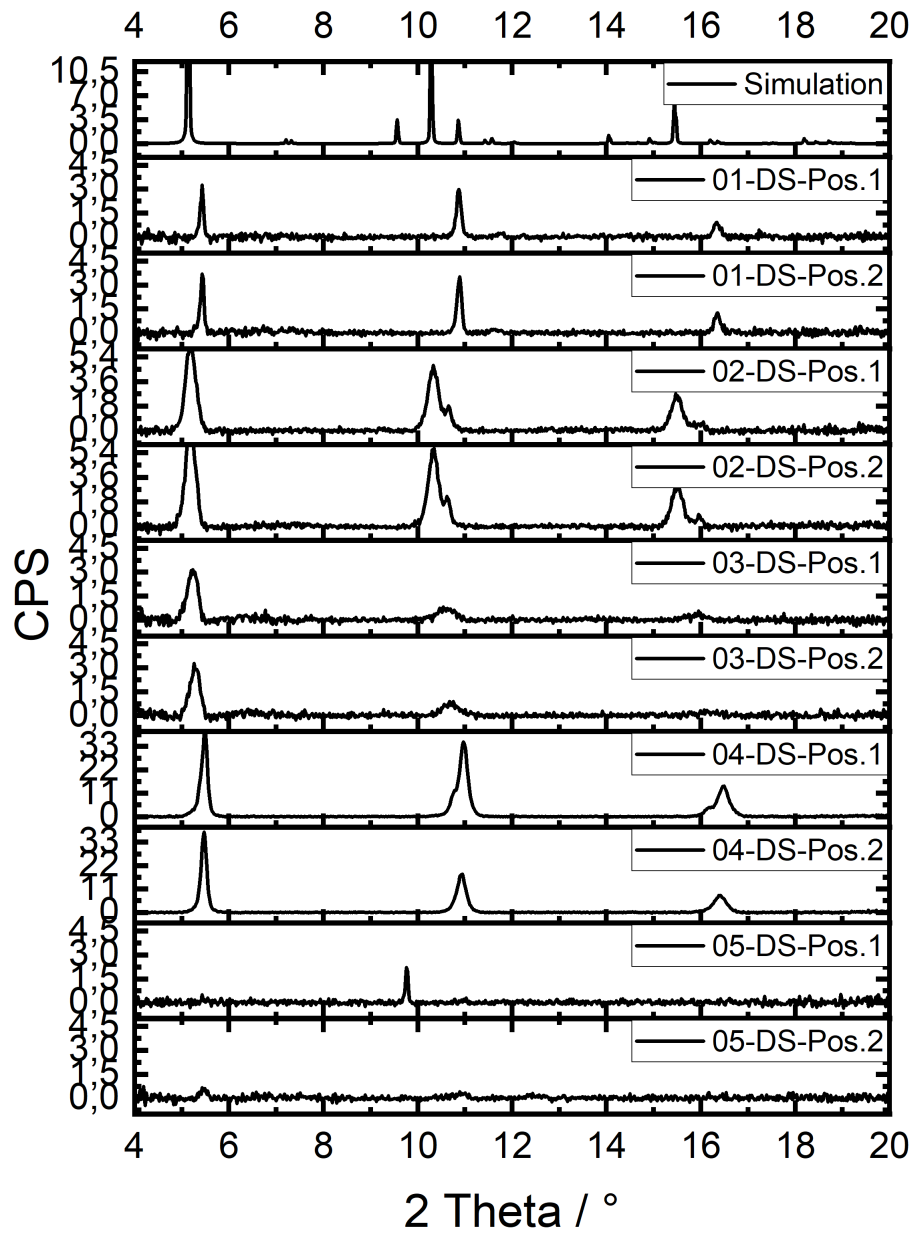


Fig. A.13: X-Ray Diffractograms of the Diverse Set for Parametercombinations 1 to 5.[Re-produced from: [30]]

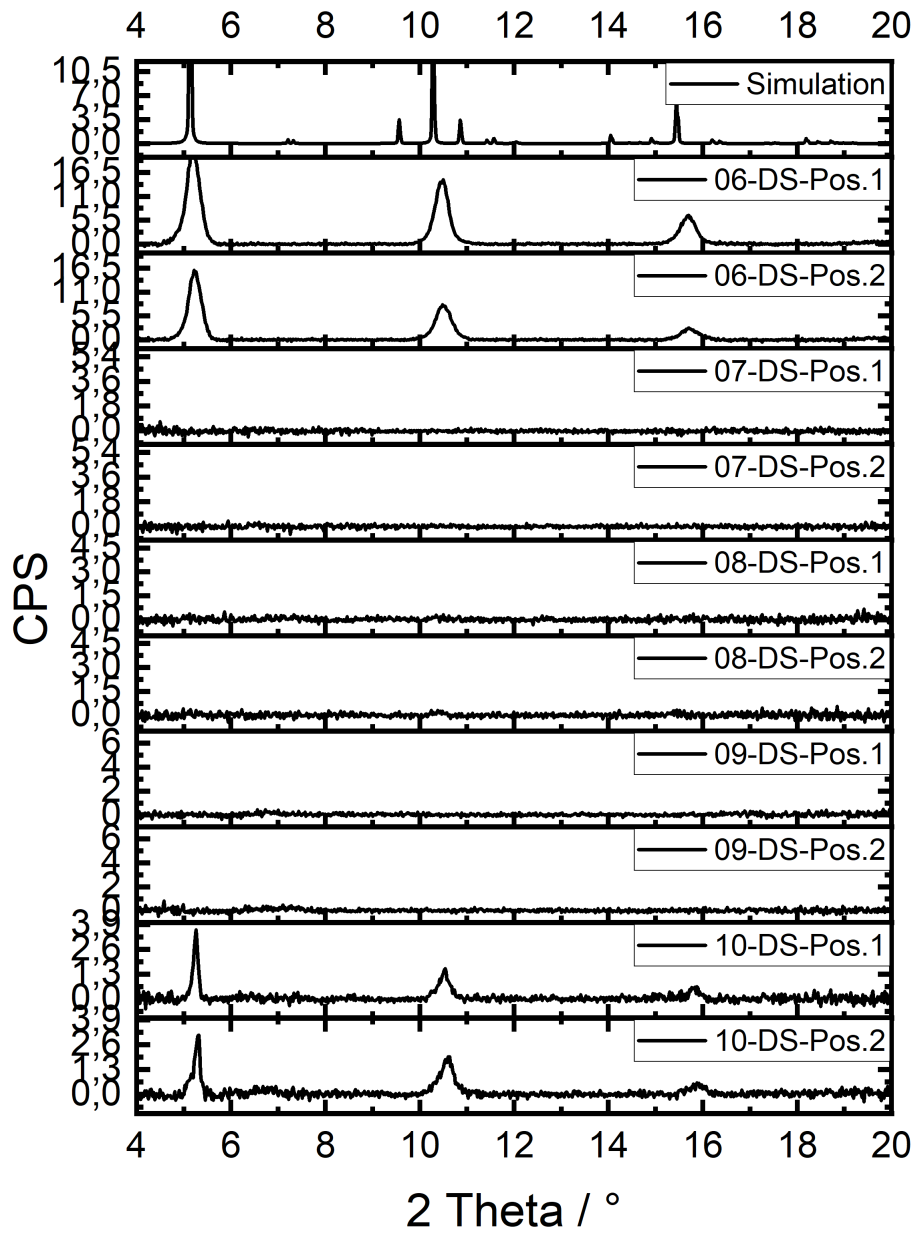


Fig. A.14: X-Ray Diffractograms of the Diverse Set for Parametercombinations 6 to 10.[Reproduced from: [30]]

Genetic Algorithm 1

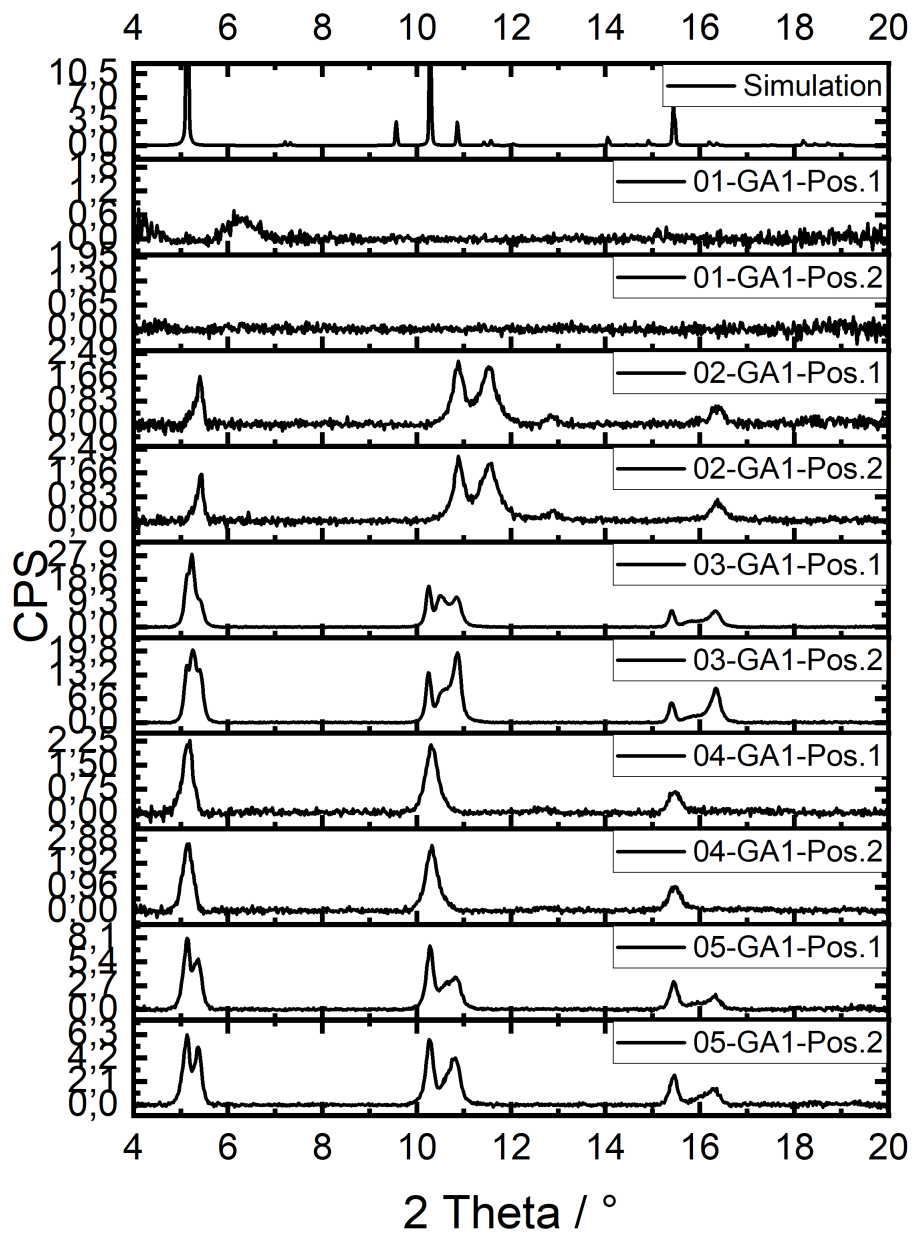


Fig. A.15: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 1 to 5.[Reproduced from: [30]]



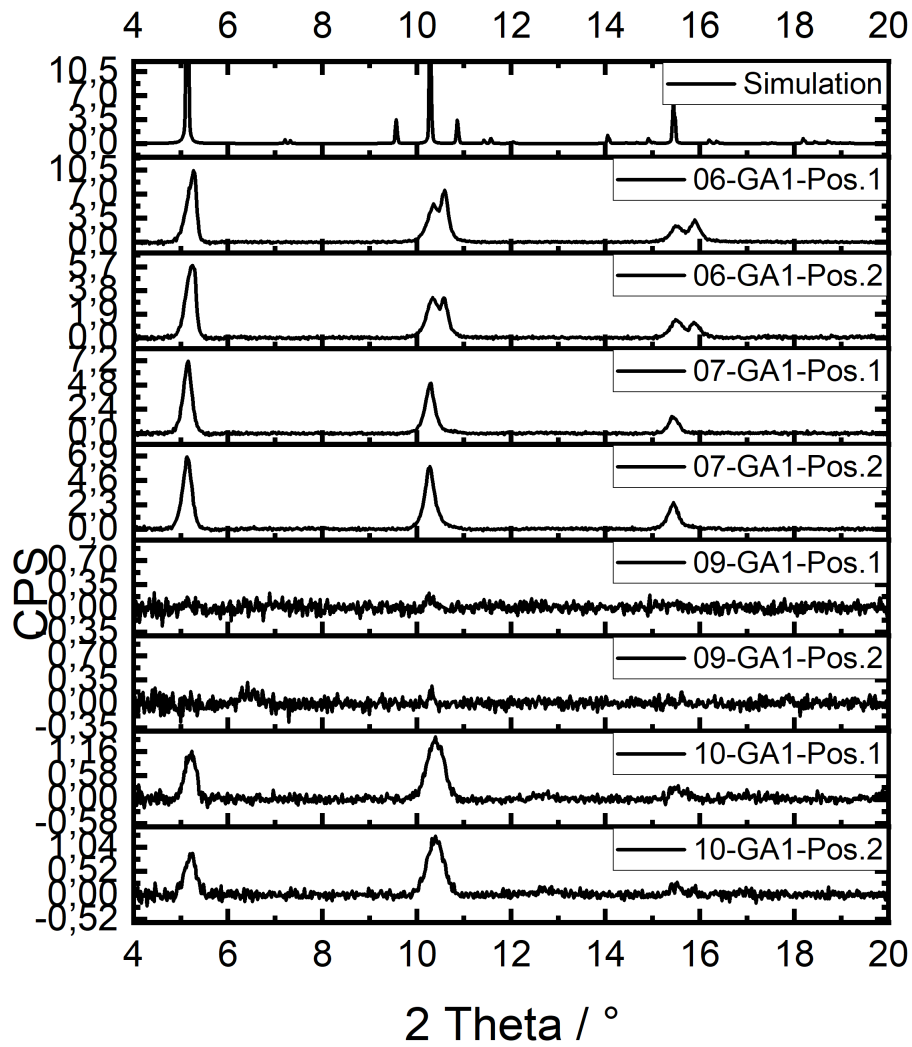


Fig. A.16: X-Ray Diffractograms of the Genetic Algorithm 1 for Parameter combinations 6 to 10. It is to be noted, that experiment number 8 is missing, since the amount of metal-salt was suggested to zero by the genetic algorithm. [Reproduced from: [30]]

Genetic Algorithm 2

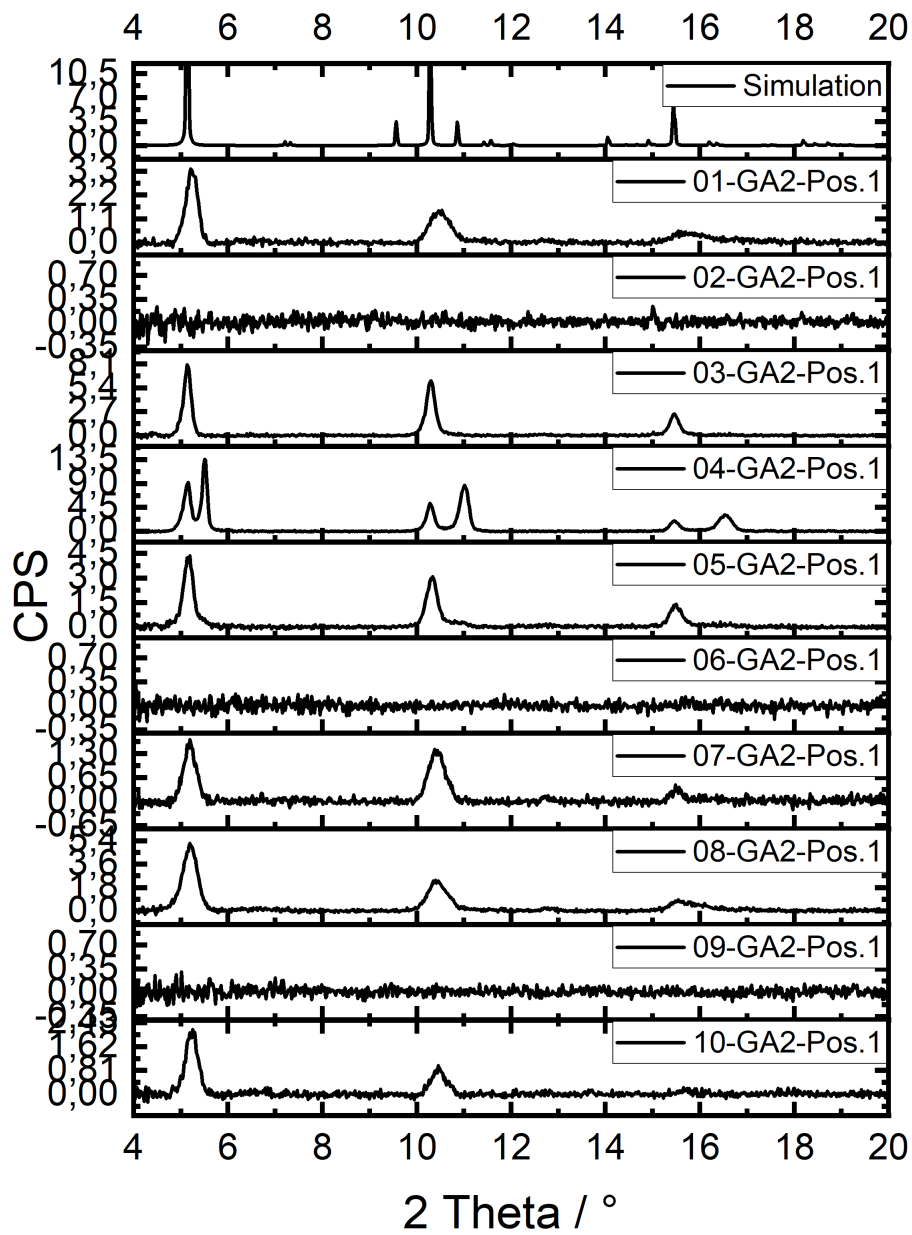


Fig. A.17: X-Ray Diffractograms of the Genetic Algorithm 2 for Parametercombinations 1 to 10.[Reproduced from: [30]]

## A.3 MOF thin film Optimization towards Roughness Reduction

### A.3.1 Synthesis Parameter

#### Diverse Set

Tab. A.13: All parameter combinations and their corresponding fitness value of the Diverse Set (DS).[Adapted and reproduced from [52] - Manuscript accepted for publication]

Ultra-sonication [s]	Modulator (Water) [mL]	n (Linker) [mmol] Cycle 4-40	n (Metal) [mmol] Cycle 4-40	Spray-cleaning [s]	n2 (Linker) [mmol] Cycle 1-3	n2 (Metal) [mmol] Cycle 1-3	Fitness
0	0.0	0.02	6.00	5	10.00	0.02	0.65
100	0.0	2.67	6.00	5	0.02	6.00	0.00
0	0.0	10.00	0.02	5	7.34	6.00	0.00
100	39.6	0.09	0.05	5	7.34	1.34	0.00
27	40.0	10.00	6.00	5	0.02	0.02	0.00
0	0.0	0.02	0.02	0	0.02	1.62	0.76
0	40.0	10.00	0.02	0	10.00	0.02	0.00
100	0.0	10.00	6.00	0	10.00	1.61	0.74
100	29.3	10.00	0.03	0	0.02	6.00	0.00
27	39.9	0.04	5.98	0	7.96	5.99	0.00

#### Genetic Algorithm 1

Tab. A.14: All parameter combinations and their corresponding fitness value of the Genetic Algorithm (GA1).[Adapted and reproduced from [52] - Manuscript accepted for publication]

Ultra-sonication [s]	Modulator (Water) [mL]	n (Linker) [mmol] Cycle 4-40	n (Metal) [mmol] Cycle 4-40	Spray-cleaning [s]	n2 (Linker) [mmol] Cycle 1-3	n2 (Metal) [mmol] Cycle 1-3	Fitness
34	26.7	10.00	0.03	1	2.85	6.00	0.00
0	0.0	10.00	0.02	5	7.34	6.00	0.72
81	0.0	2.52	6.00	5	7.93	2.97	0.62
84	5.5	10.48	1.76	0	10.60	2.32	0.00
12	2.9	10.00	0.03	5	6.65	6.00	0.67
4	3.9	6.27	6.00	5	9.04	0.02	0.00
0	18.2	6.98	3.76	3	9.56	6.38	0.75
63	22.6	0.05	0.05	2	0.33	1.59	0.54
5	15.9	0.02	1.15	5	9.96	0.97	0.00
0	28.8	8.40	0.02	0	4.55	0.41	0.00

**Genetic Algorithm 2**

Tab. A.15: All parameter combinations and their corresponding fitness value of the Genetic Algorithm (GA2).[Adapted and reproduced from [52] - Manuscript accepted for publication]

Ultra-sonication [s]	Modulator (Water) [mL]	n (Linker) [mmol] Cycle 4-40	n (Metal) [mmol] Cycle 4-40	Spray-cleaning [s]	n2 (Linker) [mmol] Cycle 1-3	n2 (Metal) [mmol] Cycle 1-3	Fitness
45	7.8	7.74	1.41	0	11.64	5.46	0.76
10	7.2	2.72	1.07	5	8.41	1.95	0.58
10	3.3	6.62	2.20	5	8.31	1.93	0.00
10	11.2	6.81	6.00	4	9.99	3.64	0.00
84	5.5	10.48	1.76	0	10.60	2.32	0.00
0	24.3	7.18	0.03	4	7.70	1.15	0.72
5	5.3	9.91	2.57	4	9.01	6.34	0.41
54	11.0	7.86	1.48	4	9.96	1.27	0.00
36	7.8	8.29	3.75	3	10.22	4.28	0.81
58	2.8	4.77	6.00	5	8.22	1.04	0.58

**A.3.2 Fitness****Diverse Set**

Tab. A.16: Fitness values of single fitness criteria and overall fitness for the Diverse Set.[Adapted and reproduced from [52] - Manuscript accepted for publication]

Number	$f(\text{phase identity})$	Crystallinity [%]	$f(\text{crystallinity})$	Roughness [nm]	MSE	Thickness [nm]	R/T	$f(\text{roughness})$	Fitness
1	1	74.8	0.75	5.97	7.96	45.50	13.11	0.87	0.65
2	1	80.0	0.80	-	-	-	-	-	0.00
3	1	87.6	0.88	-	-	-	-	-	0.00
4	1	94.0	0.94	-	-	-	-	-	0.00
5	1	87.0	0.87	-	-	-	-	-	0.00
6	1	83.2	0.83	4.54	10.13	55.58	8.17	0.92	0.76
7	1	94.7	0.95	-	-	-	-	-	0.00
8	1	87.8	0.88	7.22	8.08	45.49	15.86	0.84	0.74
9	1	92.3	0.92	-	-	-	-	-	0.00
10	1	95.3	0.95	-	-	-	-	-	0.00

**Genetic Algorithm 1**

Tab. A.17: Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 1.[Adapted and reproduced from [52] - Manuscript accepted for publication]

Number	$f(\text{phase identity})$	Crystallinity [%]	$f(\text{crystallinity})$	Roughness [nm]	MSE	Thickness [nm]	R/T	$f(\text{roughness})$	Fitness
11	1	93.6	0.94	-	-	-	-	-	0.00
12	1	84.1	0.84	5.99	4.99	42.28	14.17	0.86	0.72
13	1	83.7	0.84	14.03	11.70	55.36	25.34	0.75	0.62
14	1	92.9	0.93	-	-	-	-	-	0.00
15	1	93.0	0.93	13.21	6.73	47.31	27.91	0.72	0.67
16	1	93.2	0.93	-	-	-	-	-	0.00
17	1	94.2	0.94	12.68	19.23	63.24	20.05	0.80	0.75
18	1	87.5	0.88	23.94	6.39	61.63	38.84	0.61	0.54
19	1	95.2	0.95	-	-	-	-	-	0.00
20	1	91.5	0.92	-	-	-	-	-	0.00

**Genetic Algorithm 2**

Tab. A.18: Fitness values of single fitness criteria and overall fitness for the Genetic Algorithm 2.[Adapted and reproduced from [52] - Manuscript accepted for publication]

Number	$f(\text{phase identity})$	Crystallinity [%]	$f(\text{crystallinity})$	Roughness [nm]	MSE	Thickness [nm]	R/T	$f(\text{roughness})$	Fitness
21	1	92.6	0.93	9.00	8.82	51.19	17.58	0.82	0.76
22	1	86.6	0.87	21.47	6.82	65.13	32.96	0.67	0.58
23	1	85.0	0.85	-	-	-	-	-	0.00
24	1	90.1	0.90	-	-	-	-	-	0.00
25	1	92.7	0.93	-	-	-	-	-	0.00
26	1	94.5	0.95	13.33	9.29	55.41	24.06	0.76	0.72
27	1	91.90	0.92	23.05	11.24	41.58	55.42	0.45	0.41
28	1	96.6	0.97	-	-	-	-	-	0.00
29	1	91.6	0.92	6.52	15.63	57.17	11.40	0.89	0.81
30	1	84.2	0.84	16.57	8.32	54.25	30.54	0.69	0.58

### A.3.3 X-Ray Diffraction

#### Diverse Set

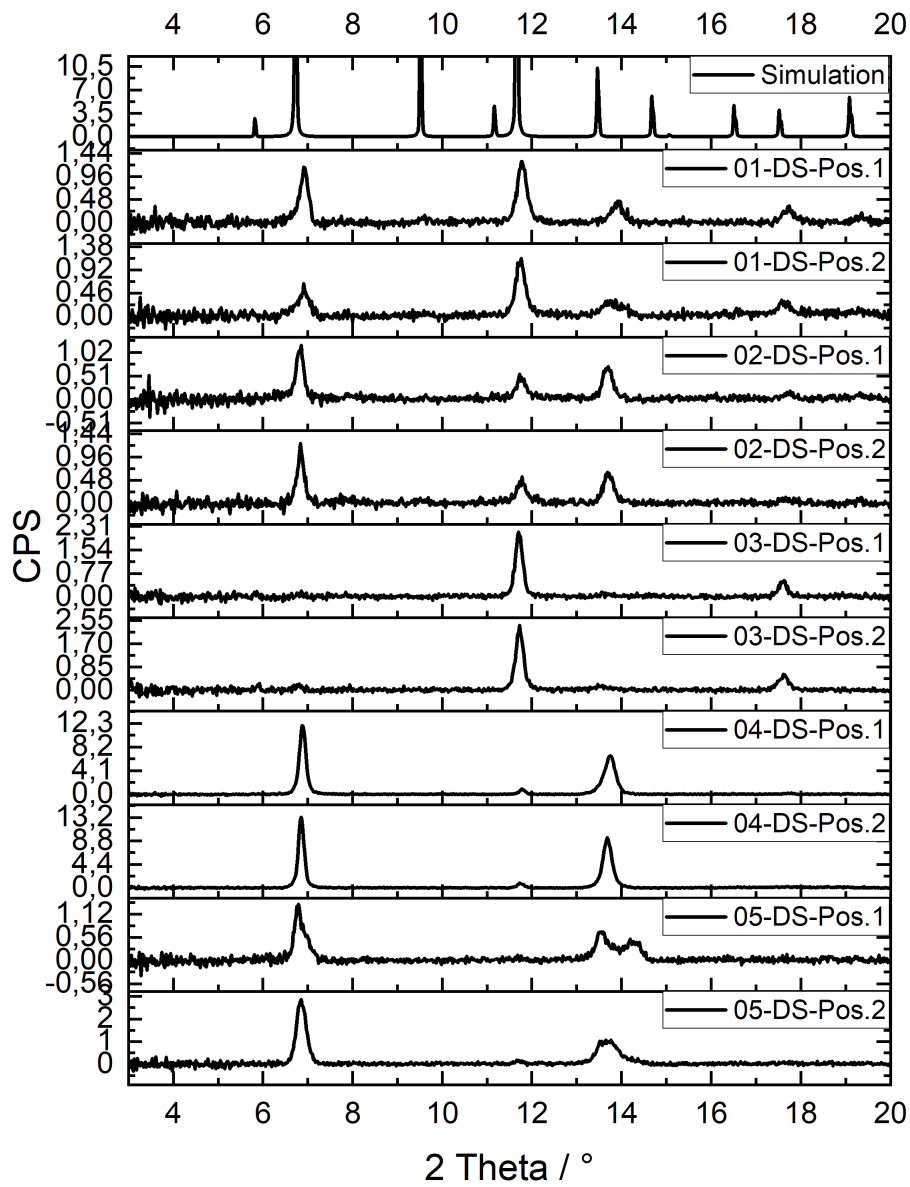


Fig. A.18: X-Ray Diffractograms of the Diverse Set for Parametercombinations 1 to 5.[Reproduced from [52] - Manuscript accepted for publication]

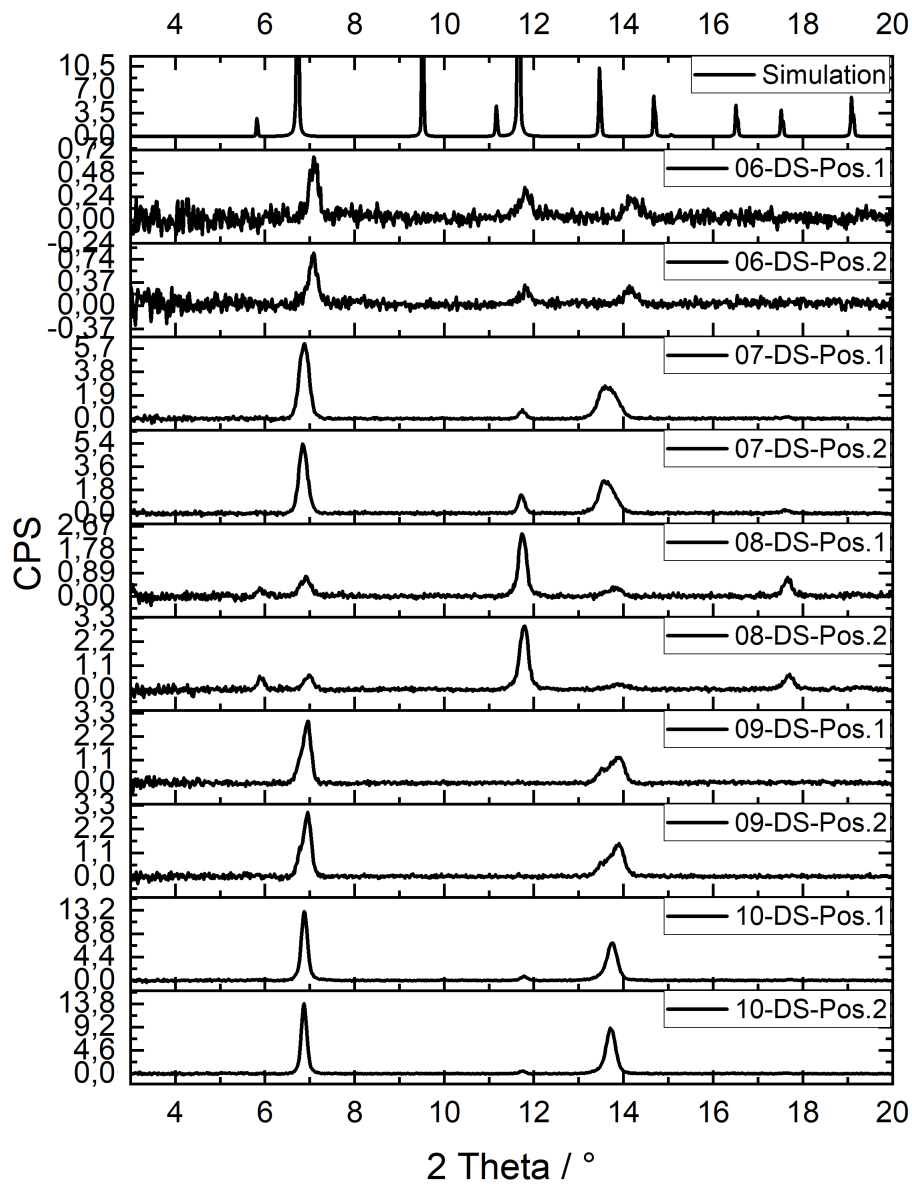


Fig. A.19: X-Ray Diffractograms of the Diverse Set for Parameter combinations 6 to 10.[Reproduced from [52] - Manuscript accepted for publication]

Genetic Algorithm 1

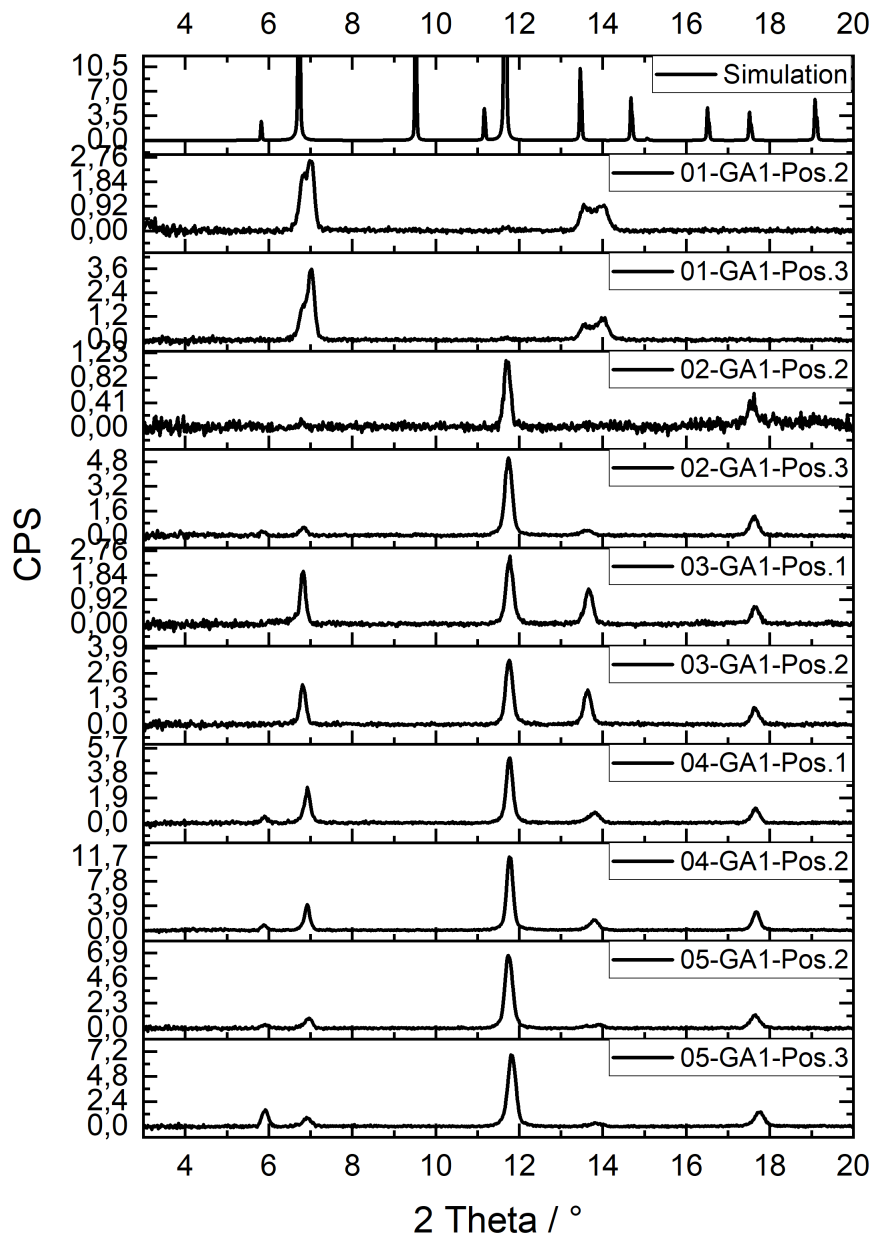


Fig. A.20: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 1 to 5.[Reproduced from [52] - Manuscript accepted for publication]



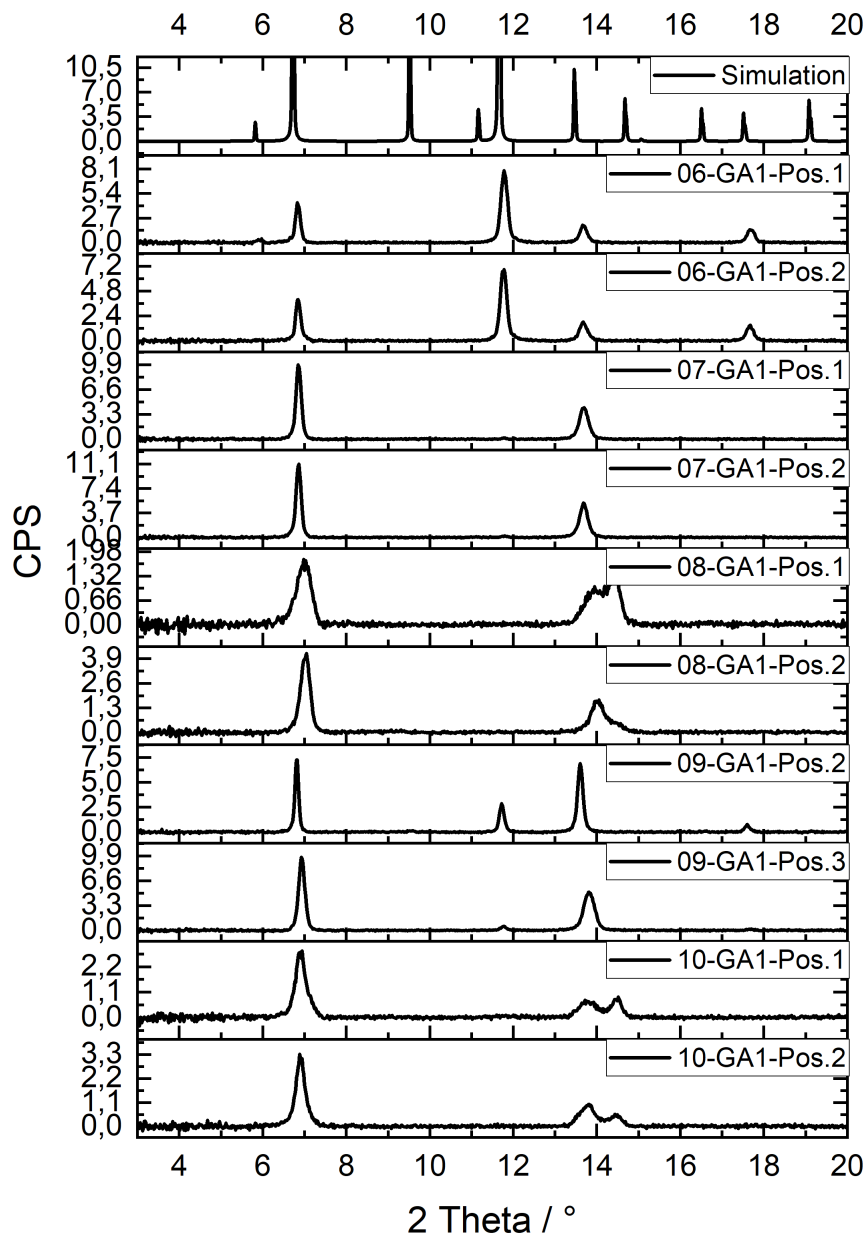


Fig. A.21: X-Ray Diffractograms of the Genetic Algorithm 1 for Parametercombinations 6 to 10.[Reproduced from [52] - Manuscript accepted for publication]

## Genetic Algorithm 2

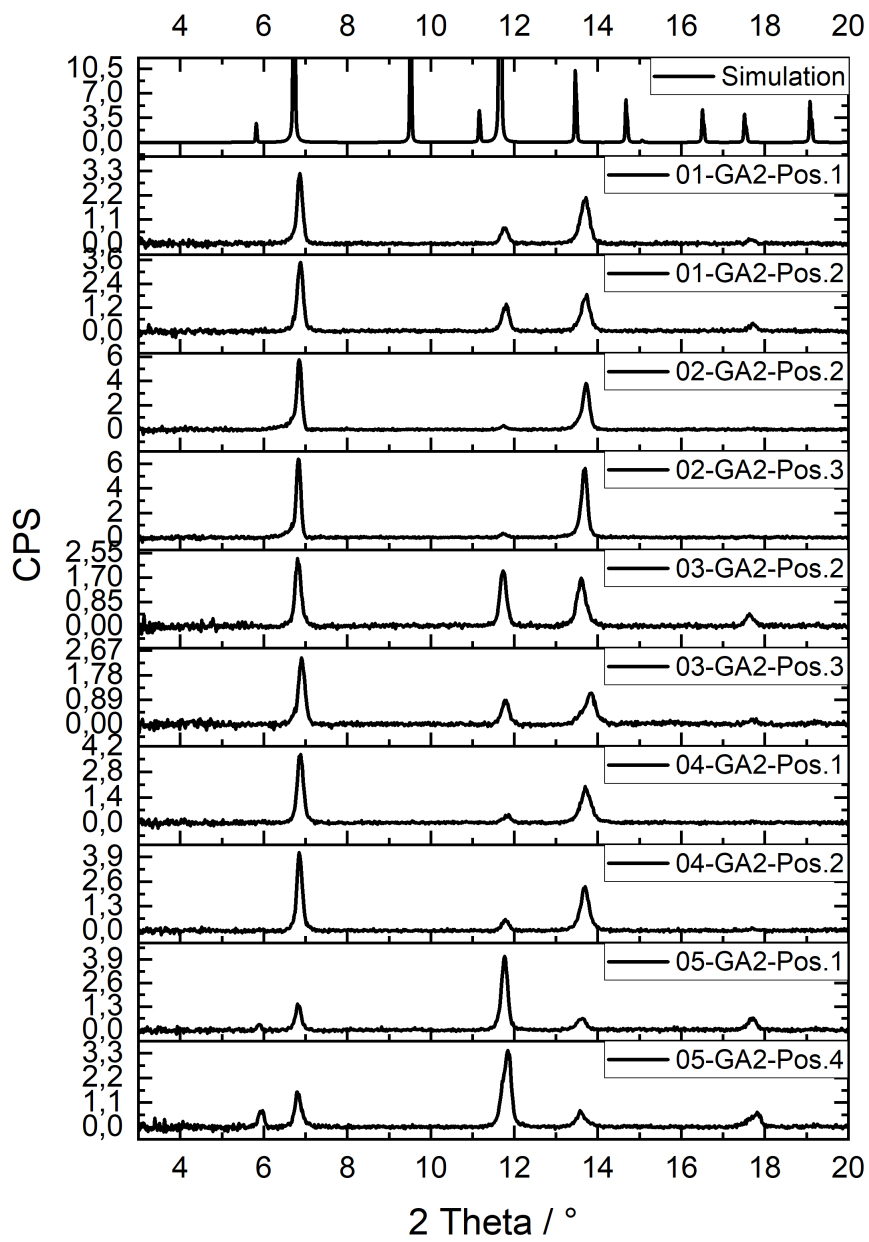


Fig. A.22: X-Ray Diffractograms of the Genetic Algorithm 2 for Parameter combinations 1 to 5. [Reproduced from [52] - Manuscript accepted for publication]

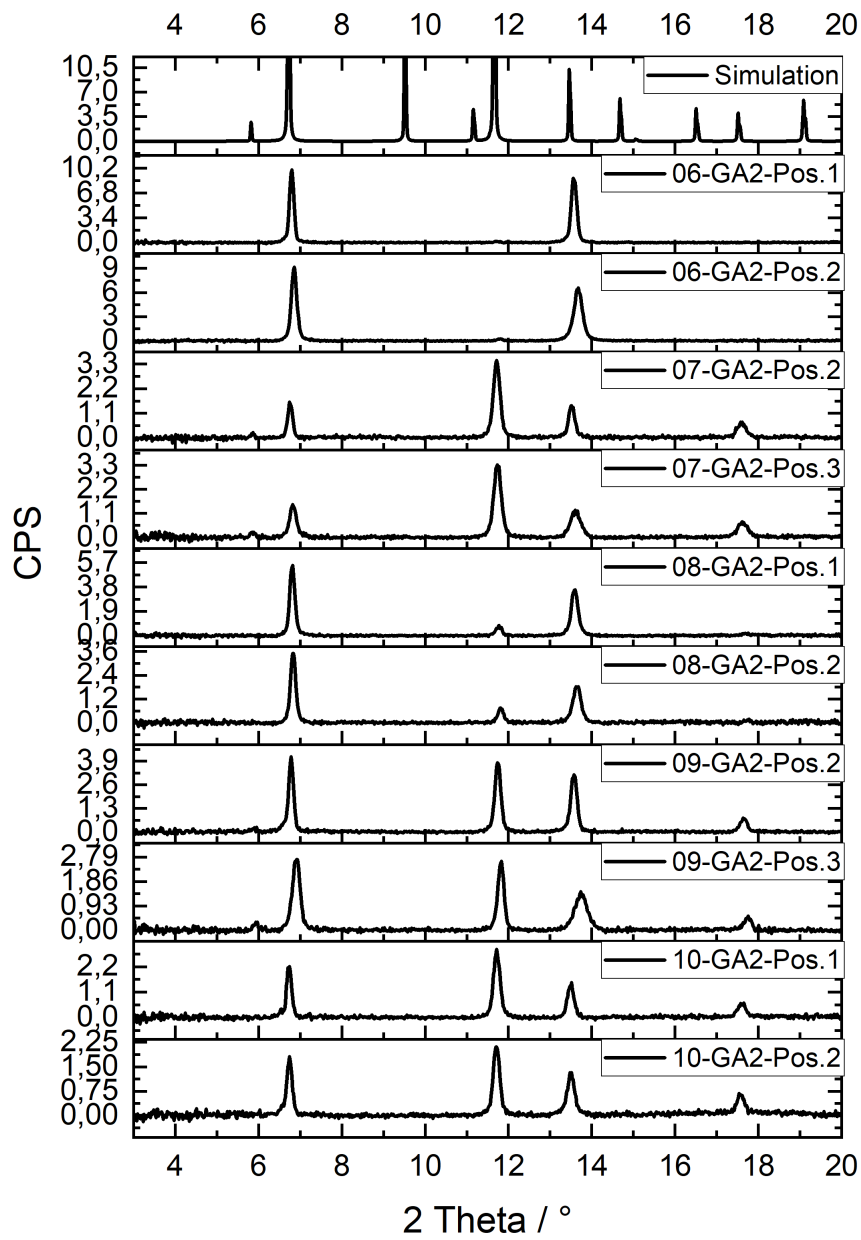


Fig. A.23: X-Ray Diffractograms of the Genetic Algorithm 2 for Parametercombinations 6 to 10.[Reproduced from [52] - Manuscript accepted for publication]

## A.4 Comparison of Genetic Algorithm and Bayesian Optimization

### A.4.1 Provided training dataset from GA Optimization for BO

Tab. A.19: Provided training dataset from GA Optimization for BO.[Adapted and reproduced from [89]]

Number	Temperature [°C]	Time [min]	Metal [mmol]	Reactantsratio [eq L of M]	Modulorratio [eq of M]	Fitness	XRD	Particlesize [nm]	PdI
1	140.00	45.00	2.00	50.00	10.00	0.20852	1	322.75	0.33
2	40.00	10.00	1.00	50.00	10.00	0.00000	0	-	-
3	40.00	45.00	2.00	0.50	5.00	0.00000	0	-	-
4	90.00	45.00	0.01	50.00	0.00	0.00000	0	-	-
5	140.00	10.00	2.00	25.25	0.00	0.08247	1	906.40	0.25
6	140.00	27.50	0.01	0.50	10.00	0.00000	0	-	-
7	40.00	27.50	2.00	50.00	0.00	0.28953	1	336.75	0.03
8	40.00	45.00	0.01	25.25	10.00	0.00000	0	-	-
9	90.00	10.00	2.00	0.50	10.00	0.26010	1	309.30	0.20
10	140.00	10.00	0.01	50.00	5.00	0.00000	0	-	-
11	40.00	14.18	0.29	25.47	10.00	0.00000	1	401.45	1.00
12	144.45	44.38	2.04	53.78	10.13	0.07096	1	729.95	0.48
13	83.88	44.73	0.77	50.00	0.00	0.15851	1	511.65	0.19
14	40.00	17.17	1.10	50.00	0.25	0.00000	0	-	-
15	140.00	10.92	0.01	19.60	8.06	0.00000	0	-	-
16	40.33	36.84	2.17	14.45	0.00	0.15045	1	440.00	0.34
17	140.00	10.00	1.53	48.31	3.75	0.02432	1	3399.00	0.17
18	63.34	29.44	1.86	9.79	9.02	0.06246	1	1368.00	0.15
19	40.00	27.50	2.00	50.00	0.00	0.40141	1	233.80	0.06
20	140.00	27.39	1.82	2.60	8.97	0.01606	1	2357.00	0.62
21	56.33	28.56	1.90	50.00	0.00	0.13505	1	601.25	0.19
22	59.71	30.37	1.95	10.16	8.00	0.08536	1	383.10	0.67
23	40.00	18.73	1.87	50.00	0.07	0.33472	1	276.20	0.08
24	130.14	11.77	1.22	48.66	2.34	0.66127	1	139.05	0.08
25	123.60	16.43	1.11	49.65	0.80	0.64853	1	141.55	0.08
26	40.00	22.19	0.39	42.13	1.80	0.00000	0	-	-
27	136.44	15.87	1.98	44.84	4.31	0.46595	1	189.40	0.12
28	51.73	41.39	0.91	23.82	0.00	0.08515	1	636.55	0.46
29	93.33	15.22	2.46	42.61	9.72	0.00615	1	495.80	0.97
30	40.32	20.97	1.63	21.02	4.34	0.00000	0	-	-

### A.4.2 Suggested Parametersets and Fitness Evaluation Values of BO

Tab. A.20: Suggested Parametersets and Fitness Evaluation Values of BO.[Adapted and reproduced from [89]]

Number	Temperature [°C]	Time [min]	Metal [mmol]	Reactantsratio [eq L of M]	Modulorratio [eq of M]	Fitness	XRD	Particlesize [nm]	PdI
1	138.9	41.00	1.89	1.7	6.3	0.0000	1	773.10	1.00
2	60.6	37.00	0.16	14.6	7	0.0352	1	1807.00	0.36
3	109.5	27.00	1.57	47.7	9	0.0798	1	305.90	0.76
4	137.5	37.00	1.53	37.2	0.7	0.6192	1	151.90	0.06
5	41.6	11.00	1.79	21.1	2.3	0.0227	1	644.30	0.85
6	63.3	34.00	0.43	47.9	2	0.1636	1	306.60	0.50
7	144.2	35.00	0.48	46.7	7.8	0.0402	1	734.10	0.71
8	101.1	38.00	0.83	11.7	9.7	0.0000	1	1033.35	1.00
9	101.8	35.00	0.68	33.5	9	0.0319	1	1052.10	0.66
10	115.4	40.00	2	19.4	6.6	0.0839	1	587.15	0.51

### A.4.3 X-Ray Diffraction

#### Genetic Algorithm 2 - ZIF-8

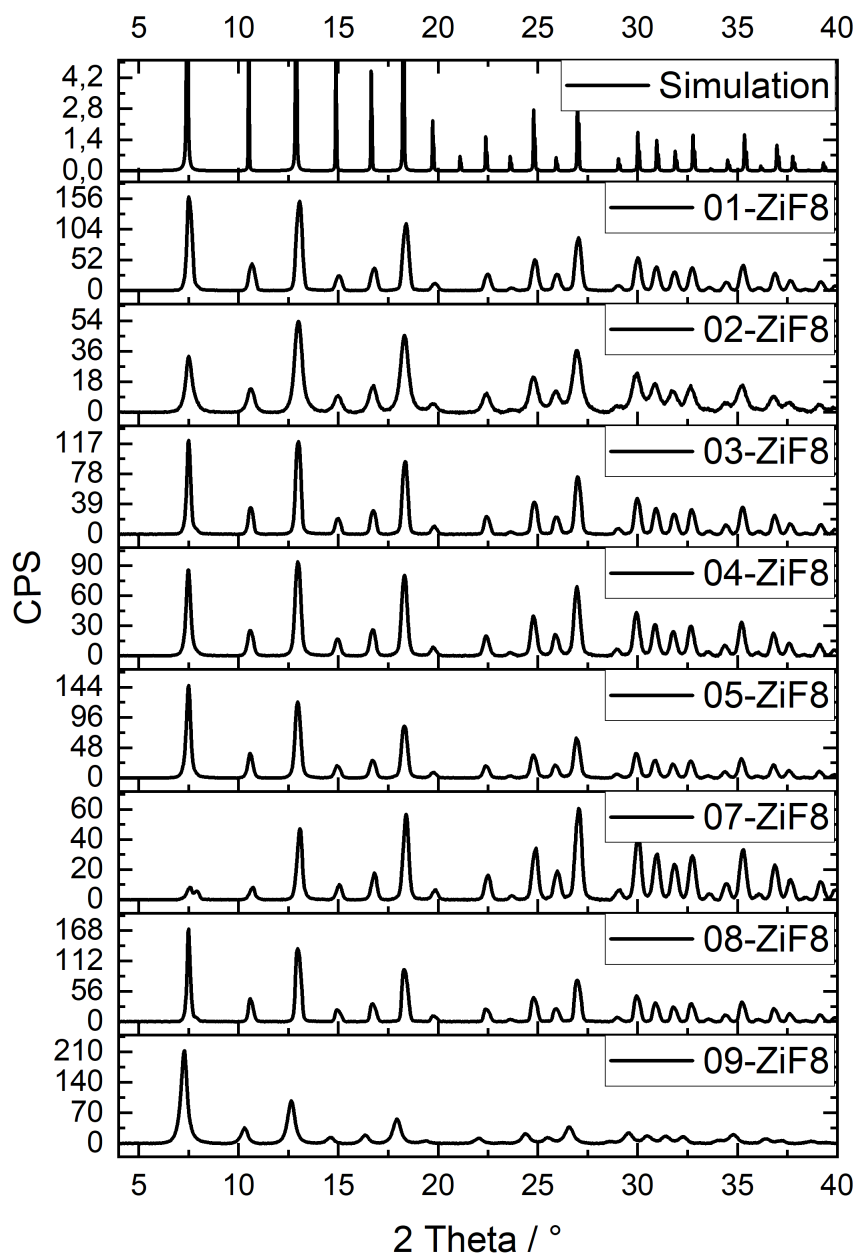


Fig. A.24: X-Ray Diffractograms of the second generation of GA Optimization (GA2) for ZIF-8.

Bayesian Optimization - ZIF-8

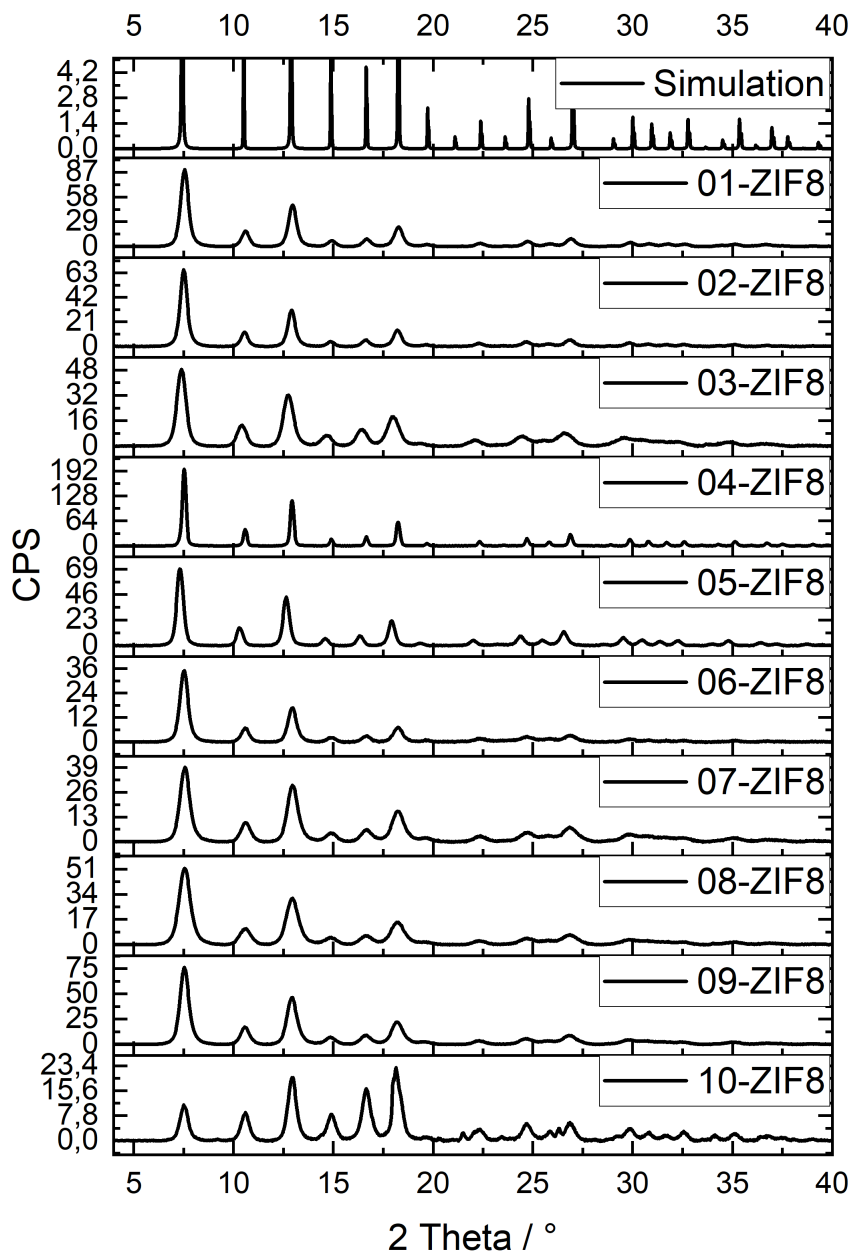


Fig. A.25: X-Ray Diffractograms of the Bayesian Optimization for ZIF-8.

## A.5 Modification of SyCoFinder

### A.5.1 Final Code of the adapted Diverse Set

Listing A.1: Final Code of the adapted Diverse Set. The original code was published under [91]. [Adapted and reproduced from [92]].

```

from __future__ import print_function
import numpy as np
import itertools
import time

def check_sample(a, var_LB, var_UB):
    # this function check if the randomly generated
    # sample point
    # lies on the phase space. In other words, it applies
    # the constraints.
    # For example, the volume of solvent in our study
    # should not be above 6ml.
    a = np.asarray(a)
    a = np.multiply(a, (var_UB - var_LB)) + var_LB
    vol = np.sum(a[4:9])
    cond = vol <= 6 and vol >= 1 # pylint: disable=
    chained-comparison
    return cond

def compute_distance(x, y, w):
    # This function compute the pairwise distance between
    # two vectors.
    # It weighs the distance based on the importance of
    # variables.
    return np.linalg.norm(w * (x - y))

def MaxMin(selected_set, nsamples, NPS1, var_importance):
    # This function returns the most diverse set of
    # parameters weighted with variable importance.
    selected_indices = []
    if len(selected_set) > nsamples:
        print("Already selected set, no need for minMaX!"
              )
        return selected_set, selected_indices
    if len(selected_set) >= len(NPS1):
        print("Already covered the entire space!")
        return selected_set, selected_indices
    if not selected_set:
        selected_indices.append(1)
        selected_set = [NPS1[0]]
    prtime_start = time.time()
    prtime = time.time()
    while (len(selected_set) < nsamples) and (len(
        selected_set) < len(NPS1)):
        distances = np.zeros([NPS1.shape[0], 1])
        for i, psamp in enumerate(NPS1):
            min_dist = 100000000
            for samp in selected_set:

```



```

        d = compute_distance(samp, psamp,
                             var_importance)
        if d < min_dist:
            min_dist = d
        distances[i] = min_dist
    extime = time.time() - prtime
    prtime = time.time()
    print(
        (
            "Index:", np.argmax(distances),
            "Distance:", distances[np.argmax(
                distances)],
            "Execution time for current landmark:",
            extime
        )
    )
    print((np.argmax(distances), distances[np.argmax(
        distances)],
          " execution time for current landmark: ",
          extime))
    selected_indices.append(np.argmax(distances))
    selected_set.append(NPS1[np.argmax(distances),
        :])

    print(("Total execution time of MaxMin: ", time.time
        () - prtime_start))
    return selected_set, selected_indices

# Transferring categories to vectors
def one_hot_encoder(number_of_categories):
    # Initializing a matrix representing One-Hot-encoding
    .
    one_hot_encoded = np.zeros((number_of_categories,
        number_of_categories))

    # Iterating over each category and setting the
    # according entry in matrix to 1.
    for i in range(number_of_categories):
        one_hot_encoded[i, i] = 1

    return one_hot_encoded

def reverse_one_hot_encoder(one_hot_encoded_vectors,
    category_names):
    # Check if the input is one-dimensional
    if one_hot_encoded_vectors.ndim == 1:
        # Find the index of the non-zero entry
        index = np.argmax(one_hot_encoded_vectors)
        return [category_names[index]]

    # For multi-dimensional input, find the index of the
    # non-zero entry in each one-hot encoded vector
    indices = np.argmax(one_hot_encoded_vectors, axis=1)

    # Map the indices to the corresponding category names

```

```

reversed_categories = [category_names[i] for i in
    indices]
return reversed_categories

def combinatorial_product_spaces(A, B):
    combined = []
    for x in A:
        for y in B.tolist():
            combined.append(x+y)
    return combined

def compute(var_importance, var_LB, var_UB,
    number_of_categorical_variables,
    number_of_categories_list, num_samples=10,
    ngrids_per_dim=5):
    """Compute most diverse set of inputs.
    :param var_importance: list of weights
    :param var_LB: list of lower bounds
    :param var_UB: list of upper bounds
    :param num_samples: number of samples to pick
    :param ngrids_per_dim: number of grid points in each
        variable
    :return result: String containing coordinates of
        sample points.
    """
    num_variables = len(var_importance)
    #Checking that var_importance fit's the number of
    variables
    num_continuous_variables = num_variables -
        number_of_categorical_variables
    if not num_variables==
        number_of_categorical_variables + len(var_LB):
        print("Error in number of variables!")
        return None

    # building the countious space
    grids_dim = np.linspace(
        0,
        1,
        num=ngrids_per_dim,
        endpoint=True,
    )
    print(("On each dimension, we sample: ", grids_dim))

    NPS_continuous = itertools.product(grids_dim, repeat=
        num_continuous_variables)
    NPS_continuous = [list(i) for i in NPS_continuous]
    print(("In total, there are ", len(NPS_continuous), "
        samples in the continous space\n"))

    NPS = NPS_continuous

    # array shall contain possibilities for categorical-
    vector-values without using a grid

```

```

for i in range(number_of_categorical_variables):
    one_hot_encoded_vectors = one_hot_encoder(
        number_of_categories_list[i])
    NPS = combinatorial_product_spaces(NPS,
        one_hot_encoded_vectors)
    NPS = np.array(NPS)
print(("In total, there are ",
    number_of_categories_list, "samples in the
    categorical space\n"))
print()
continous_importance = var_importance[: -
    number_of_categorical_variables]
categorical_importance = [var_importance[-
    number_of_categorical_variables]/np.sqrt(2) for i
    in range(number_of_categories_list[0])]

new_variable_importance = continous_importance +
    categorical_importance
print("The total size of search space is ", len(NPS)
    )
print(new_variable_importance)
#Here MaxMin method should be applied on both
    numerical and categorical values
norm_diverse_set, sel_ind = MaxMin([], num_samples,
    NPS, var_importance=new_variable_importance)
print("After MaxMin, length of diverse_set:", len(
    norm_diverse_set))
print(norm_diverse_set)

# Select only the continuous part
print(sel_ind)
diverse_set_continous = np.array(NPS)[sel_ind][:, :
    len(var_LB)]
diverse_set_continous = diverse_set_continous * (np.
    array(var_UB) - np.array(var_LB)) + np.array(
    var_LB)
# Select only categorical part
diverse_set_categorical = np.array(NPS)[sel_ind][:,
    len(var_LB):] # Select only the categorical part
print("Length of continuous set:", len(
    diverse_set_continous))
print("Length of categorical set:", len(
    diverse_set_categorical))

# Reverse the categorical part to original categories
reversed_categories = reverse_one_hot_encoder(
    diverse_set_categorical, categories)
print("Reversed Categories:", reversed_categories)

#Cobination of both continous and categorical sets
diverse_set_pre = np.concatenate((
    diverse_set_continous, diverse_set_categorical),
    axis=1)
print("Diverse Set not reencoded, yet:\n",
    diverse_set_pre)

```

## Appendix

```
# Convert reversed_categories to a column vector
before concatenating
reversed_categories_column = np.array(
    reversed_categories)[: , np.newaxis]
print("Reversed Categories in column:\n",
    reversed_categories_column)

# Combine both continuous and categorical sets
diverse_set = np.concatenate((diverse_set_continuous,
    reversed_categories_column), axis=1)
print("Length of diverse_set:", len(set(map(tuple,
    diverse_set))))
print("Diverse Set:\n", diverse_set)
return diverse_set
```

## A.6 Publications

Lena Pilz, Meike Koenig, Matthias Schwotzer, Hartmut Gliemann, Christof Wöll, Manuel Tsotsalas, *Advanced Functional Materials*, Manuscript accepted for publication.

Lena Pilz, Carsten Natzeck, Jonas Wohlgemuth, Nina Scheuermann, Simon Spiegel, Simon Oßwald, Alexander Knebel, Stefan Bräse, Christof Wöll, Manuel Tsotsalas, Nicholas Prasetya, *Journal of Materials Chemistry A*, 2023, **11**, 24724-24737.

Aya Ali, Mustafa Basuni, Reham Shams-Eldin, Lena Pilz, Tawheed Hashem, Marita Heinle, Alexei Nefedov, Muhamed Amin, Manuel Tsotsalas, Goran Drazic, Abdou Hassanien, Mohamed H. Alkordi, *ACS Applied Nano Materials*, 2023, **6**, 5, 3627-3636.

Lena Pilz, Carsten Natzeck, Jonas Wohlgemuth, Nina Scheuermann, Peter G. Weidler, Ilona Wagner, Christof Wöll, Manuel Tsotsalas, *Advanced Materials Interfaces*, 2023, **10**, 2201771.

## A.7 Conference Contributions

### Poster Presentation

'Machine learning based synthesis-optimization in metal-organic framework thin films'

**5th European Conference on Metal Organic Frameworks and Porous Polymers (EuroMOF)**, Granada, Spain, September 2023

### Poster Presentation

'Machine Learning Optimization in Metal-Organic Framework Thin Films'

**Heraeus Seminar**, Bad Honnef, Germany, May 2023

### Poster Presentation

'Machine Learning Optimization in Metal-organic Framework Thin Films'

**34. Deutsche Zeolith-Tagung**, Vienna, Austria, February 2023

### Oral Presentation

'Machine Learning-Driven Synthesis Optimizations in Thin Films and Nanoparticles'

**MRS Fall Meeting**, Boston, USA, November 2022

### Oral Presentation

'Chemical Intuition for the Nano-Space: SyCoFinder applied to SURMOF and NanoMOF Synthesis'

**33. Deutsche Zeolith-Tagung**, Frankfurt, Germany, March 2022

### Poster Presentation

'Chemical Intuition for the Nanospace: SyCoFinder Applied to SURMOF and NanoMOF Synthesis'

**4th European Conference on Metal Organic Frameworks and Porous Polymers (EuroMOF)**, Krakow, Poland (online event), September 2021

## A.8 Curriculum Vitae

### Personal Details

---

Name:	Lena Pilz
Date of Birth:	February 11, 1993
Place of Birth:	Überlingen

### Education & Career

---

2021 – 2024	PhD, Chemistry, Karlsruhe Institute of Technology
2017 – 2021	Master of Science, Chemistry, Karlsruhe Institute of Technology
2012 – 2017	Bachelor of Science, Chemistry, Karlsruhe Institute of Technology

### Experience

---

#### Vocational

2019 – 2021	Student Assistant, International Scholar and Welcome Office, Karlsruhe Institute of Technology
2019 – 2020	Student Assistant, Institute of Functional Interfaces, Karlsruhe Institute of Technology
2017	Student Assistant, Institute of Functional Interfaces, Karlsruhe Institute of Technology

#### Visits abroad

2022	Visit to the research group of Prof. Seth M. Cohen as part of the <i>KHYS Networking Grant</i> , UC San Diego, USA
2023	Visit and Collaboration with the research group of Prof. Mohamad Moosavi, Toronto, Canada

### Voluntary Activities

---

2012 – present	Voluntary work in the HaDiKo events e.V., Hans-Dickmann Kolleg, Karlsruhe
2017	Volunteering in the extended board of HaDiKo e.V., Hans-Dickmann Kolleg, Karlsruhe