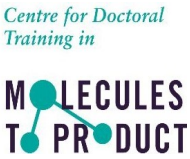
Crystallisation Optimisation and Control using Surrogate Modelling and Adaptive Model Predictive Control

Transfer Report

Version: v 0.1 — First Draft

Joseph McHale

Submitted in accordance with the requirement for the degree of Doctor of Philosophy



The University of Leeds
School of Chemical and Process Engineering

June 6, 2023

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Joseph McHale

Acknowledgements

I would like to thank my supervisors - Elaine Martin, Tariq Mahmud and Keeran Ward, for their support and input into the project, and their help refining and designing the ideas within.

I would also like to acknowledge the Engineering and Physical Sciences Research Council (EPSRC) and thank them for funding this work.

I am also grateful to the Centre for Doctoral training in Molecules to Product and everyone within for their continued support and stimulating discussions on the work, and their willingness to share ideas and help each other develop as researchers.

Abstract

The optimisation of processes with multiple objectives is an intense sector of research currently, with the aim to balance economic, environmental and efficacy goals at the forefront of this Need reference for this. Crystallisation processes are currently still difficult to robustly control and optimise, due to a lack of understanding of the fundamental mechanisms that are required to model said crystallisation to a sufficient level to deal with emergent conditions, or long-term shifts in process parameters Need reference here as well. Direct design control and model based control are some of the known methods to achieve robust control, with recent advances being a combination of the two Need reference here as well. However, these often require high computation time, significant experimental work to determine kinetics, and limited application as parameters such as vessel size change.

The objective of this research is to create a digital twin of a crystallisation system, that can respond optimally to these emergent conditions and long-term shifts, through the use of predictive modelling and control. This predictive modelling is intended to be done via including the usage of historical data from the crystallisation system under different conditions, which should contain inferences on what the current optimal response should be, from what has been observed in the past. This is intended to encompass recent system data as well, to enable a continuously adaptive control system, that in theory improves its responses over time. The methodology proposed in this work could potentially allow significantly faster responses for real-time control due to low computation time, less experimental work, and the ability to more accurately control a wider range of possible parameters. The novelty of this work is seen to initially be in the combination of first-principles and data-driven approaches, the inclusion of historical data from industrial processes and the testing of limits of adjacency (when predictions are beyond an accepted tolerance) for these systems, and potentially the introduction of new parameters other than cooling rate in the control of the desired objective.

Contents

List of Figures

List of Tables

Abbreviations

Table 1: Table of Abbreviations - in alphabetical order

Abbreviation	Meaning
ANN	Artificial Neural Network
AR	Aspect Ratio
CLD	Chord Length Distribution
CSD	Crystal Size Distribution
DRO	Deep Reinforcement Optimisation
EIS	Enterprise Information System
EKF	Extended Kalman Filter
ERP	Enterprise Resource Planning
GA	Genetic Algorithm
KPI	Key Performance Indicator
MLS	Moving Least Squares
MPC	Model Predictive Control
PAT	Process Analytical Technique
PD	Predicted Desireable
PINN	Physics-Informed Neural Network
PLM	Product Lifecycle Management
PU	Predicted Undesireable
RBF	Radial Basis Function
RL	Reinforcement Learning
RNN	Recurrent Neural Network
UD	Unpredicted Desirable
UU	Unpredicted Undesirable

1 Introduction

1.1 Background and Vision

A brief description of the background of the area, and why this research is needed. How this has built into the vision for the project (a pyramid of the vision would be good here!)

1.2 Research Question

During the development of a new crystallisation process in industry, a wealth of data is generated. This data contains information on how certain factors such as impurities, spontaneous secondary nucleations, or other factors affect the crystallisation, but this information is currently unused in most cases due to difficulties of scale-up and complexity, or other confounding factors. When these conditions are experienced further in the product lifecycle, inferences should allow a response based upon these past interactions. The process needs to learn from its current operation, combining this with past inferences from optimum responses to past stimuli, and respond appropriately to optimise the current conditions that the process finds itself in.

> How can a process respond to emergent conditions optimally, given that past data with inferences on the optimum response exists - where the optimum response is complex. Is this best done through self-optimisation such as Reinforcement Learning, past-inference from the Primary and Adjacent systems, or a combination of methods.

1.3 Research Objectives

In order to achieve this envisioned workflow to retain, understand, and apply the data that is produced during process/product development, and then on into its lifecycle, there needs to be a sufficiently organised repository of said data. The parameters that are critical to the process need to be recorded through the process development lifecycle, in a structured manner that reflects the state of the process and its parameters at each step. The current proposed method at the forefront of this is the Digital Twin. This represents a Digital model that mirrors the process as it goes through the various development stages, and records any related information in a clearly understood form. Furthermore, this data is then used to control and optimise the process, and the digital representation. Allowing the conceptualisation of any future or past changes, collaboration within that data for other similar processes at each stage of development, and comparison between the digital representation and the physical system it is twinned to.

This project seeks to focus on the control aspect of the Digital Twin, examining the current limits of Machine Learning combined with Simulation, in the optimal control of a batch cooling crystallisation. The theorised workflow of the creation of a simulation, the training of a surrogate model, optimisation of the system and then further training of a Model Predictive Controller (MPC) system, controller, would allow for potential improvements in the control of crystallisation processes when subject to unexpected conditions that cause a shift from optimal conditions.

Such conditions could include for example - the deterioration of equipment such as heat jacket fouling, or the changing of raw material supplier leading to unexpected impurities. Without this kind of self-optimising control, the process would need retuning, with the large potential costs that come from such, or a decrease in some or most process Key Performace Indicators (KPIs). The optimal response to a shift like this is

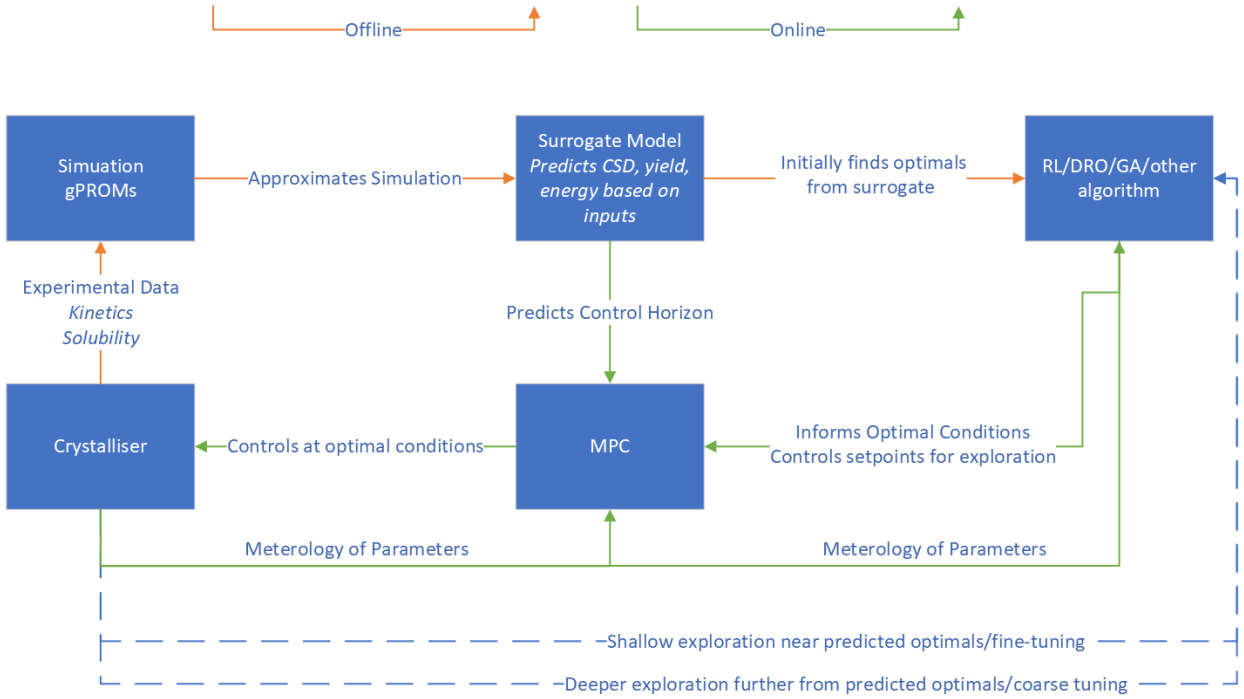


Figure 1: High-level overview of the control system proposed to be developed - to be developed as a general workflow that can be applied to numerous different systems.

complex due to interactions between manipulated variables and the nonlinear relationships within. As such, any minimisation in the required experiments to keep at optimal conditions, or the ability to adaptively optimise, is of keen interest to industry and stakeholders.

The main objectives of the research are:

- Building of a simulation through Population Balance Modelling.
- Build a surrogate model that approximates the Population Balance Model of a crystallisation.
- Train an MPC controller using this surrogate model and compare with other MPC methods, such as more conventional RLS.
- Optimisation of the system - Potentially using Reinforcement Learning (RL), Deep Reinforcement Optimisation (DRO), or other optimisation algorithms like Genetic Algorithms (GA).
- Inclusion of historical and similar systems and the inferences within.
- The transferability of this framework to new crystallisation systems.

The first step has been completed by a previous PhD student, provided validation of the simulation is successful. Either way building a surrogate model and selecting the most efficient development method for this will be the initial starting point - with a hexamine in ethanol batch cooling crystallisation. A summary of this is shown in Figure 1.

1.4 Responsible Research and Innovation

Responsible research and innovation (RRI) is a framework used to scrutinise the potential effects that research is going to have on society. It is intended as a transparent and interactive process, which investigates how research affects various factors such as the economy, the environment and the general public [15]. This is seen as a method with which to reduce the frequency in which research is conducted irresponsibly, which may cause detrimental effects in one or more aspects of society. Examples of this include the development of leaded petrol and chlorofluorocarbons (CFCs), which were later discovered to be harmful to either the environment or general public [16][17].

Two other key areas will also be considered throughout the research project, and reviewed as often is required. These are:

Research Integrity Research Integrity focuses on the four core elements of honesty, rigour, transparency and open communication, and care and respect for fellow researchers [18]. These principles promote a healthy research environment for both the public and researchers themselves. High-integrity research means a higher degree of trust between researchers, but also a higher degree of trust between research institutions and the public [19]. This could be viewed as contrasting with some of the main incentives of research, which means active effort must be applied to ensure research integrity is upheld [20]. Hence, throughout this project and any collaborations or research outputs, research integrity as a principle will be reviewed and maintained at all times, through ensuring the four core elements are upheld.

Research Ethics We have a responsibility to consider the ethical implications of any research we conduct. The areas that need to be considered include experimental methodology and conduct, which need to be reviewed in the experimental design phase to ensure ethical considerations are taken into account. This is true not only in the social sciences, but also in the areas of engineering and chemistry [21]. Data management is also a critical area, namely the initial storage, archiving and dissemination of such.

This can be achieved by reviewing the six key principles of research ethics, as suggested by the UKRI [22].

- Maximise societal benefit whilst minimising harm and risk.
- Respect the rights of groups and individuals.
- Participation should be voluntary and informed.
- Integrity and transparency should be observed.
- Responsibility and accountability should be defined.
- Conflicts of interest should be avoided, or made known if this is not possible.

An ethical review is not required for this project at this time, but the principles can be considered nonetheless, as doing so may improve research quality. The main points that will be reviewed include the storage and dissemination of data as in Section 1.5, and the minimisation of the environmental impact and risk of any experiments, when these are planned and performed.

The three of these frameworks combine to provide a robust methodology with which to evaluate the research project and its impact, and inform how to proceed. They share similarities and differences as Figure 2 demonstrates visually, but with this method we can say all bases are likely covered.

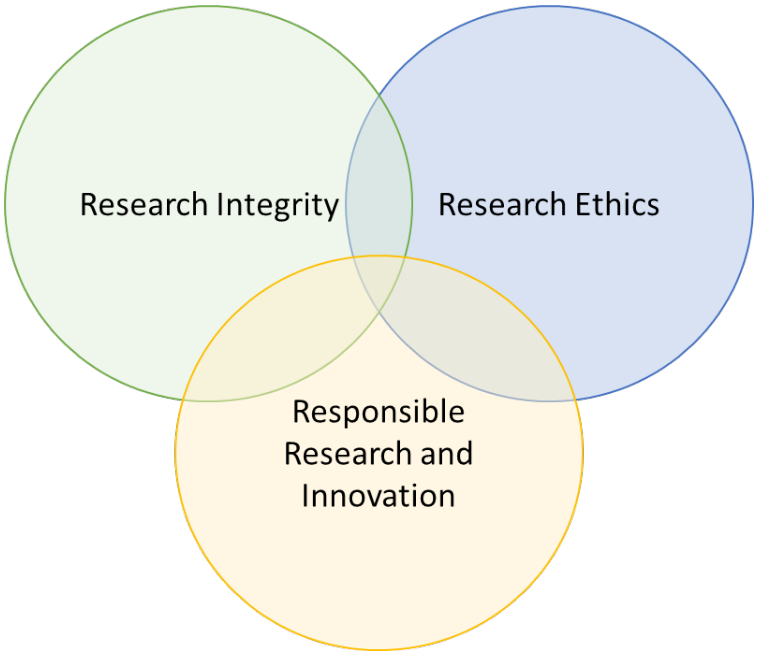


Figure 2: This combination of frameworks ensures the most important areas of RRI, research ethics and integrity are covered and the methodology is transparent to all stakeholders. The concepts have some similarities, but also key differences that justify their usage.

AREA Framework The AREA framework was developed in 2013 as a way to clearly lay out how RRI and its main principals can be understood and systematically reviewed. The result was four dimensions of responsible innovation: anticipation, reflexivity, inclusion and responsiveness [23]. This was further develop by the ESPRC into the AREA framework: anticipate, reflect, engage and act, which represents the same four dimensions [24]. This is used as the basis for ensuring RRI is covered in this project, as it splits up the major concepts into understandable and reviewable pieces.

1.4.1 The ORBiT Methodology

Building on the AREA framework as described in Section 1.4, the ORBiT methodology provides a a way to numerically evaluate the standards a project needs to meet, based upon the impact said project is predicted to have on society. This is based on the idea that the more impact an innovation has, the more thoughroughly it should be investigated and tested in terms of RRI, before being allowed to proceed. As we'll see in the following sections, the required RRI scrutiny increases as the technology matures and reaches potential for use in general society.

UN Sustainable Development Goals What are the UN Sustainable Development Goals and how are they applied here.

Technological Readiness Level (TRL) What is TRL, apply it to this project.

RRI Intensity Level (RIL) What is RIL, apply it to this project.

1.5 Data Management Plan

A data management plan on how data from the project will be maintained, backup up, and stored so that confidentiality will be maintained if required, especially from industry or any collaborations.

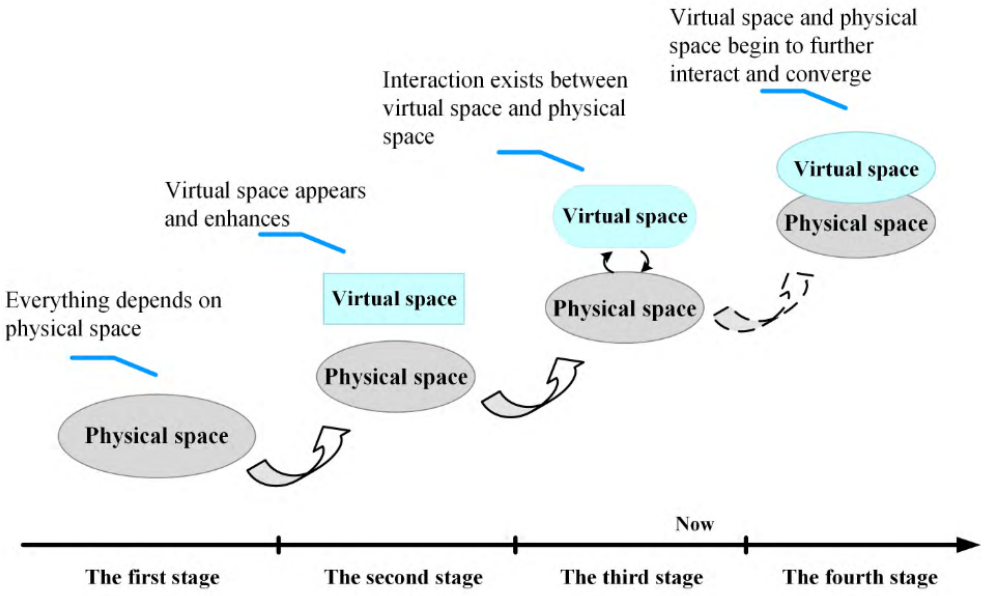


Figure 3: Evolution of the shop-floor and the convergence of the digital and physical in the representation of such [1].

2 Literature Review

2.1 Digital Twins

Digital Twins are typically described as consisting of a physical entity and a virtual model or "twin", which mirrors the physical part through the use of data connections and modelling. They are currently being explored as a way of improving the performance of the physical aspect through computational techniques applied to the virtual twin [2]. For example through the use of a virtual model to investigate the causes of process fluctuation in real-time, and hence enable the root cause to be found, and improved [**?**].

Currently the virtual entity and physical entity of a theoretical production process interact, in a limited fashion. But as Digital Twins and other Industry 4.0 techniques continue to be developed the intention is that they converge. Interacting to a larger degree until the virtual is practically as important is the physical [1]. Figure 3 shows this evolution. In this vision the Digital Twin not only stores a history of the performance of the physical entity, but also carries out optimisation and prediction for it, constantly calibrating itself using data from the physical entity. Properly established communication between the physical and virtual using a Digital Twin should lead to dual-optimisation, the virtual optimising the operation of the physical, and the physical optimising the simulation of the virtual.

Unlike process control, which is usually done with a pre-defined notion of the process, a Digital Twin should be a critical part of the development process from inception. You would develop the process control from the digital twin, and the digital twin would then after the product is developed, optimise the process control.

Concept	Description
Digital Twin	A complete virtual description of a physical product that is accurate to both micro and macro level.
Digital Twin Prototype	The virtual description of a prototype product, containing all the information required to create the physical twin.
Digital Twin Instance	A specific instance of a physical product that remains linked to an individual product throughout that products life.
Digital Twin Aggregate	The combination of all the Digital Twin Instance.
Digital Twin Environment	A multiple domain physics application space for operating on Digital twins. These operations include performance prediction, and information interrogation.

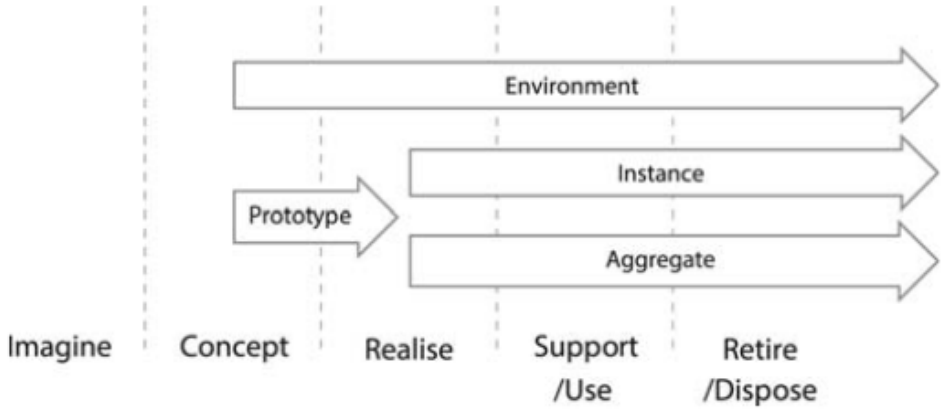


Figure 4: Digital Twin lifecycle example [2], demonstrating the uses of proposed Digital Twin types throughout.

Research in these areas progress so quickly in modern times that it is difficult to maintain a common framework, leading to multiple potentially relatable systems being developed in tandem, duplicating work, and muddying the waters for potential investors or future contributors. Digital twins are a good example of this, where scepticism and confusion are abound and one of the main challenges during initial discussions.

If the original meaning and intention is not referred to, the larger scope of possibilities with Digital Twins may be missed for years to come, even if likely inevitable. The original definition and purpose of the "digital twin" relied around its application as the foundations of product life-cycle management, being developed in tandem with the process itself [**?**][25].

An example of how the Digital Twin lifecycle can be applied in tandem with the product development lifecycle can be seen in Figure 4. In theory this application methodology allows the monitoring and improvement of a product through its entire lifecycle. In which development of the physical element is done with the usage of a digital element in mind, instead of the digital element being applied in hindsight - which results in many of the problems we see in data-driven improvement methodologies being applied as an afterthought.

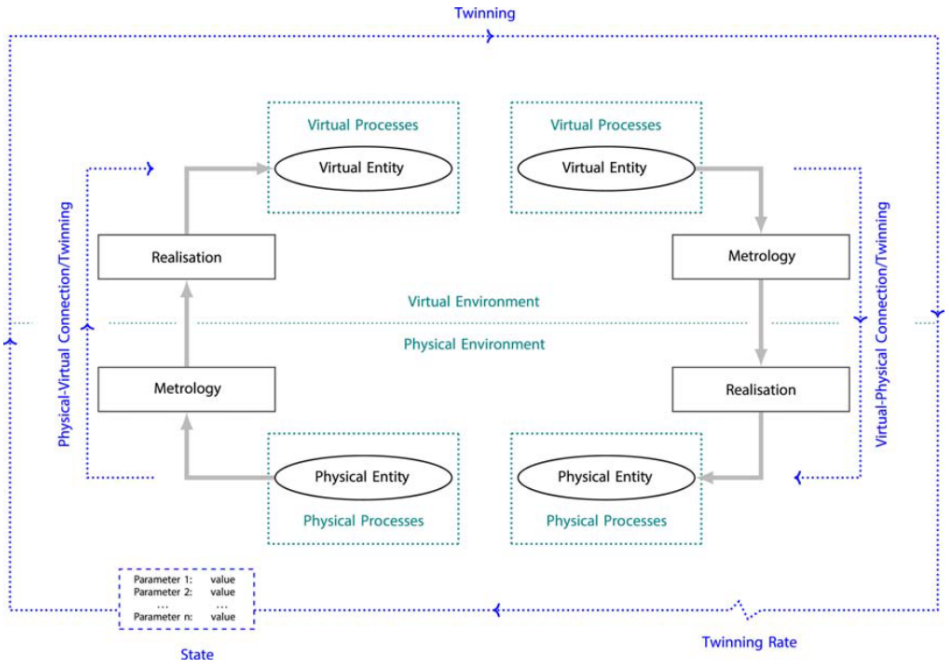


Figure 5: Digital Twin twinning system overview [2].

The basic function of a Digital Twin can be seen in Figure 5, where the physical element is measured, and then transmitted to the Digital Twin where the delta is realised in the virtual element. Similarly the Digital Twin may respond, or be affected by another change. The delta between the virtual element and the physical would then again be measured, and realised by a change in the physical element.

An example of how the Virtual Entity could be set up is shown in Figure 6

A Digital Twin needs a Physical-to-virtual connection, or it is more akin to a Digital Model. A Virtual-to-physical connection is described in the original definition by Grieves, but not present in many Digital Twin publications. Conceptually this is said to be possible, but the benefits of the Digital Twin may be hard to realise without the Virtual-to-physical connection [2]. Although the realisation step of affecting the physical element may instead be achieved with a manual human connection, so is possible in certain situations. Especially as Digital Twins are further applied to existing processes, not developed with a Digital Twin in mind.

Digital Models and Shadows - How the Twin is different There is understandably a fair amount of confusion on how a Digital twin differs to a Digital model. The most basic definition that should allow the average person to understand is shown in Figure 7. In a model data is manual both ways, no direct interactions exist. In a shadow data is collected in real-time, but no connection exists from the shadow to the physical entity [3].

As the twinning diagram in Figure 5 demonstrates, the Twinning Rate is the frequency at which the virtual element aligns with the physical element. This could be represented for example through Hz. Each individual sensor could have its own Twinning Rate, represented by individual parameters. The Twinning

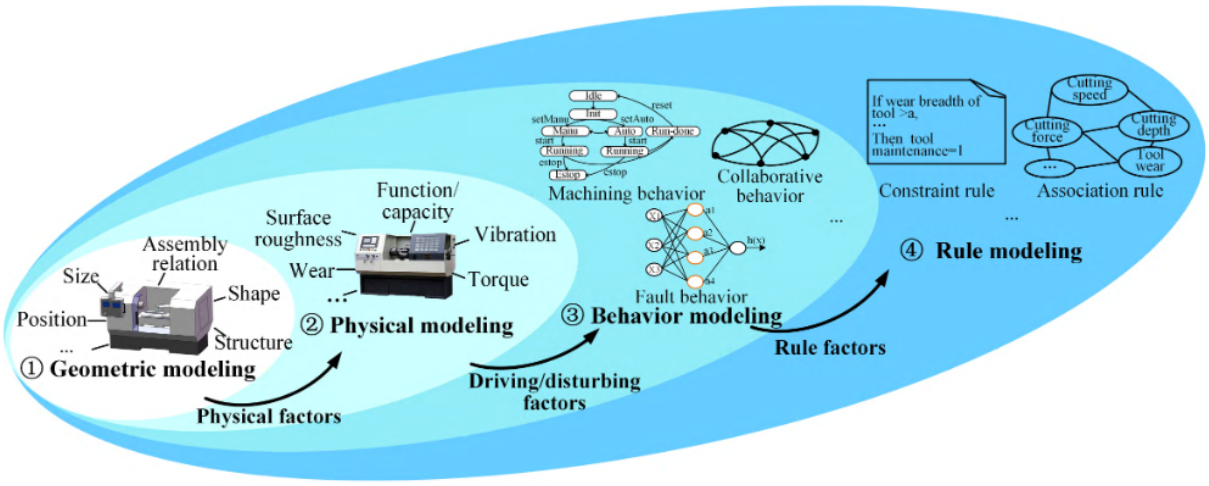
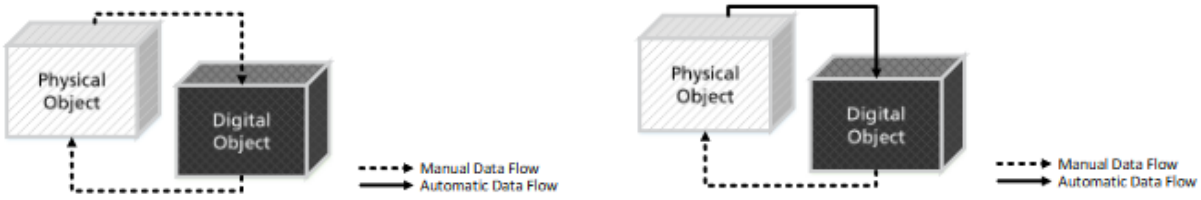


Figure 6: Virtual entity components suggested framework [1].



(a) A Digital Model

(b) A Digital Shadow

Figure 7: Digital Model and Digital Shadow - Data Flow Example [3]. Represents how a Model and Shadow differ from a Twin.

Theme	Description
1. Physical Entity	A 'real-world' artefact, e.g. a vehicle, component, product, system, model.
2. Virtual Entity	A computer generated representation of the physical artefact, e.g. a vehicle, component, product, system, model.
3. Physical Environment	The measurable 'real-world' environment within which the physical entity exists.
4. Virtual Environment	Any number of virtual 'worlds' or simulations that replicate the state of the physical environment and designed for specific use-case (s), e.g. heath monitoring, production schedule optimisation.
5. Fidelity	The number of parameters transferred between the physical and virtual entities, their accuracy, and their level of abstraction. Examples found in literature include: fully comprehensive, ultra-realistic, high-fidelity, data from multiple sources, micro-atomic level to the macro-geometrical level.
6. State	The current value of all parameters of either the physical or virtual entity/environment.
7. Parameters	The types of data, information, and processes transferred between entities, e.g. temperature, production scores, processes.
8. Physical-to-Virtual Connection	The connection from the physical to the virtual environment. Comprises of physical metrology and virtual realisation stages.
9. Virtual-to-Physical Connection	The connection from the virtual to the physical environment. Comprises of virtual metrology and physical realisation stages.
10. Twinning and Twinning Rate	The act of synchronisation between the two entities and the rate with which synchronisation occurs.
11. Physical Processes	The physical purposes and process within which the physical entity engages, e.g. a manufacturing production line.
12. Virtual Processes	The computational techniques employed within the virtual-world, e.g. optimisation, prediction, simulation, analysis, integrated multi-physics, multi-scale, probabilistic simulation.
13. Perceived Benefits	The envisaged advantages achieved in realising the Digital Twin, e.g. improved design, behaviour, structure, manufacturability, conformance, etc..
14. Digital Twin across the Product Life-Cycle	The life-Cycle of the Digital Twin – (whole life cycle, evolving digital profile, historical data)
15. Use-Cases	The applications of the Digital Twin, e.g. reducing cost, improving service, supporting decision making.
16. Technical Implementations	The technology used in realising the Digital Twin, e.g. Internet-of-Things.
17. Levels of Fidelity	The number of parameters, their accuracy, and level of abstraction that are transferred between the virtual and physical twin/ environment.
18. Data Ownership	The legal ownership of the data stored within the Digital Twin.
19. Integration between Virtual Entities	The methods required to enable communication between different virtual entities.

Table 2: Digital Twin terminology table [2].

Rate of the whole Digital Twin could therefore be calculated, either from the slowest Twinning Rate of a constituent component, or some other more weighted measure.

There are a wide numbers of terms used to describe Digital Twins, further complicating developments in the area and compounding any already present scepticism or confusion. For example, when describing the Physical element, a review found that 'vehicle', 'component', 'product', 'system', 'models', and 'artefact' ahd all been used to describe the same aspect. What this review paper then defines as the "physical entity" [2].

In the interest of keep terms consistent the terminology proposed in their review will be used henceforth. The adoption of a standardised system should significantly help understanding in the area if we are to encourage widespread adoption. The terminology can be found in Table 2.

Physical entity describes the, real-world existing part of the digital twin. Physical twin describes this when it is twinned to a virtual entity, which likewise becomes a virtual twin - when twinned.

Physical environment on the other hand describes the "environment" in which the physical entity is situated. In theory meaning any parameters that affect said physical entity, or the output it produces. This could be the temperature, pressure, or whether or not the local football team won the latest match. Anything that could be theorised to have an effect, if attempting to apply a full Digital Twin. The virtual environment is where and how the virtual entity is stored or hosted. Where possible this will be kept unspecific, so to not cause writings to become outdated due to use of specific terminology like "data-warehouse".

Fidelity is an important but general term that may need to be broken down further. Currently it describes anything from the number of different parameters, to the level of abstraction between the virtual and physical, to the rate at which measurements are taken, and so on. A high fidelity Digital Twin is broadly described as one that matches the physical and virtual twins with a high accuracy.

A physical or virtual process represents a change in the physical or virtual state. For example, a manufacturing process. A virtual process is said to take place entirely within the virtual environment. This could

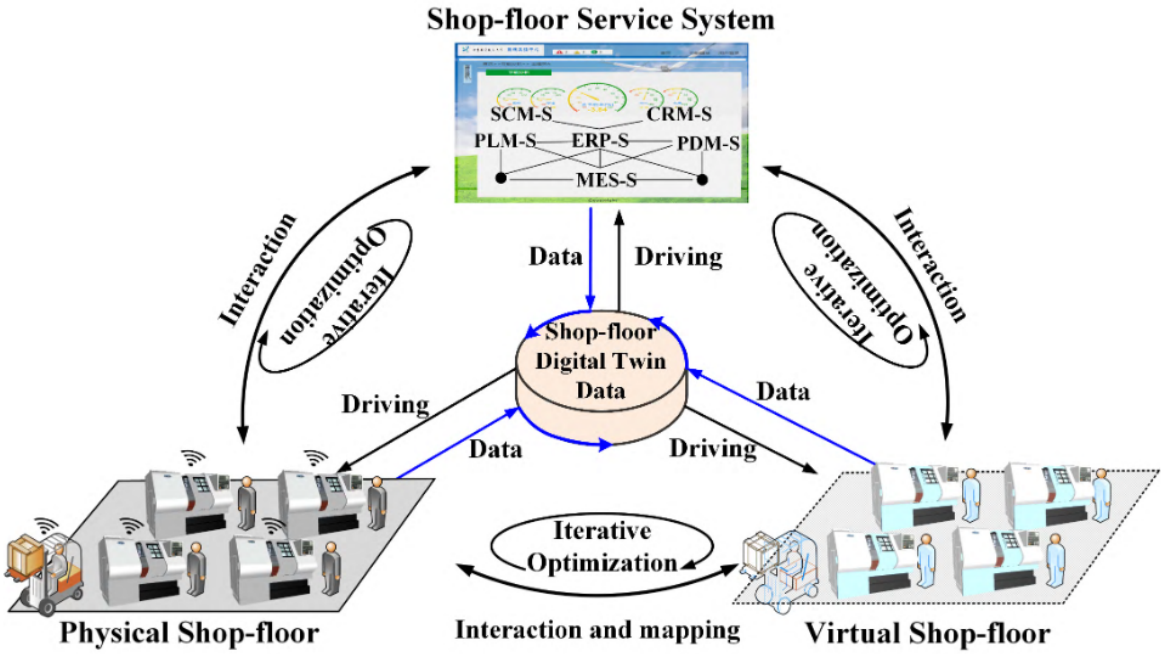


Figure 8: Shop-Floor framework example for a Digital Twin [1], demonstrating one way data can be efficiently exchanged between functions.

include modelling, simulation, optimisation. The results of this could then be fed back to the physical entity.

It is suggested by Grieves that the collection and reuse of historical data is a critical aspect of Digital Twins, enabling optimisations not previously possible. While a worthy objective, the already high costs may be substantially increased by the requirement to store all possible data. It may be worth optimising what needs to be stored and how.

One of the most influential shop-floor digital twin frameworks is shown in Figure 8. This demonstrates how a digital twin could be implemented. A physical entity, sending data to a digital twin database, which also recieves data from the virtual entity, and a "service system", which encapsulates the functions of any Enterprise Information Systems (EIS), for example an Enterprise Resource Planning (ERP) like SAP. The physical and the virtual make demands of the service system, which performs sub-services required for the specific demand. For example, modelling for the virtual entity, visualisation of the current physical entity status, fusing of the data in both physical and virtual, and so on.

2.1.1 Complexity and Emergent Behaviour

A key aim in the use of Digital Twins is the ability to predict behaviours without having to exhibit a physical entity to the conditions that want to be tested, saving time, money, and many other resources. For example, testing what happens when a valve fails. Or coming up with solutions to process fluctuations by virtually changing different parameters or equipment. Therefore, it is critically important to understand what they

can and can't predict, and why. This is where a need to understand complexity and emergent behaviour becomes important [26].

First, it should be understood that a complicated system is one that is completely predictable, and follows well defined patterns. A complex system will for the rest of this discussion be a system that cannot be fully predicted, and will always demonstrate some Emergent behaviour. This is generally due to a series of unfamiliar sequences, unplanned and unexpected, that are not visible or immediately comprehensible [27].

Grieves proposes four main categories for system behaviour, Predicted Desirable (PD), Predicted Undesirable (PU), Unpredicted Desirable (UD) and Unpredicted Undesirable (UU). The meaning of these systems is fairly self-explanatory. PU is the result of ignoring an issue we know exists, PD is "normal running", UD would be where something goes better than it has been predicted, and UU is where the potentially deadly unknown behaviours occur. The suggested purpose of the Digital Twin is to remove as much of the UU behaviours as possible. By predicting the behaviours that we would have missed without their usage. How? Grieves suggests the best way to do this is by developing Digital Twins in tandem with the product itself, or system. Stating that the behaviour of a system "does not simply pop into existence fully formed", and exhibit the behaviours they do due to the way they have been designed.

2.1.2 Digital Twins and Product Lifecycle Management

It has been previously stated that Digital Twins should be developed as part of the Product Lifecycle Management (PLM) process, from conception and design through to the final retirement of the system.

During the "create" stage, a Digital Twin can be used to simulate how a physical entity would perform. It can be used to test scenarios that would have previously required some sort of physical entity, often in a destructive manner. By using a Digital Twin that can be thoughroughly investigated, scenarios can be tested to eliminate the Unpredictable Undesirable behaviour of a system to the furthest extent possible. Instead of designing a "desired" system and carrying out human-centric risk analysis which often misses possible hazards, a systematic virtual approach could be taken in which each parameter is varied to the min and max that they can handle. Of course, not all permutations can be examined, due to the computational power and time this would take in complex systems. Techniques such as machine learning may help with this where possible, such as through identifying patterns we ourselves find difficult to identify. It is currently thought that during this "Digital Twin Prototypes" should be developed, similar to a "Class" in an object-oriented programming language, which can then be developed into a "Digital Twin Instance" during the next phase. A Digital Twin Instance being like an "object" in an OOP, where the parameters of the prototype are detailed to develop a Digital Twin unique to the physical entity it describes.

During the next phase, the "production" phase, the physical entity is built based upon the Digital Twin, and data is sent back to the virtual entity, where they can then optimise each other as discussed in other sections. The data can also then be fed back for future virtual entities, and to optimise the "Digital Twin Prototype" so it may be used more effectively during the next design.

The last phase is the "retirement" phase, where it is critical all the data is stored for iterative use to be fed into the next Digital Twin. It is also important to maintain any aspects needed for the effective and safe disposal of the system, including all substantially and other aspects, and information on how disposal goes.

There are many varied parameters that are required when designing a new product/process, and these can be hard to understand and balance. A Digital Twin can help with this through understanding what can be optimised. The data of previous usage could also be combined with this - utilising the history of previous

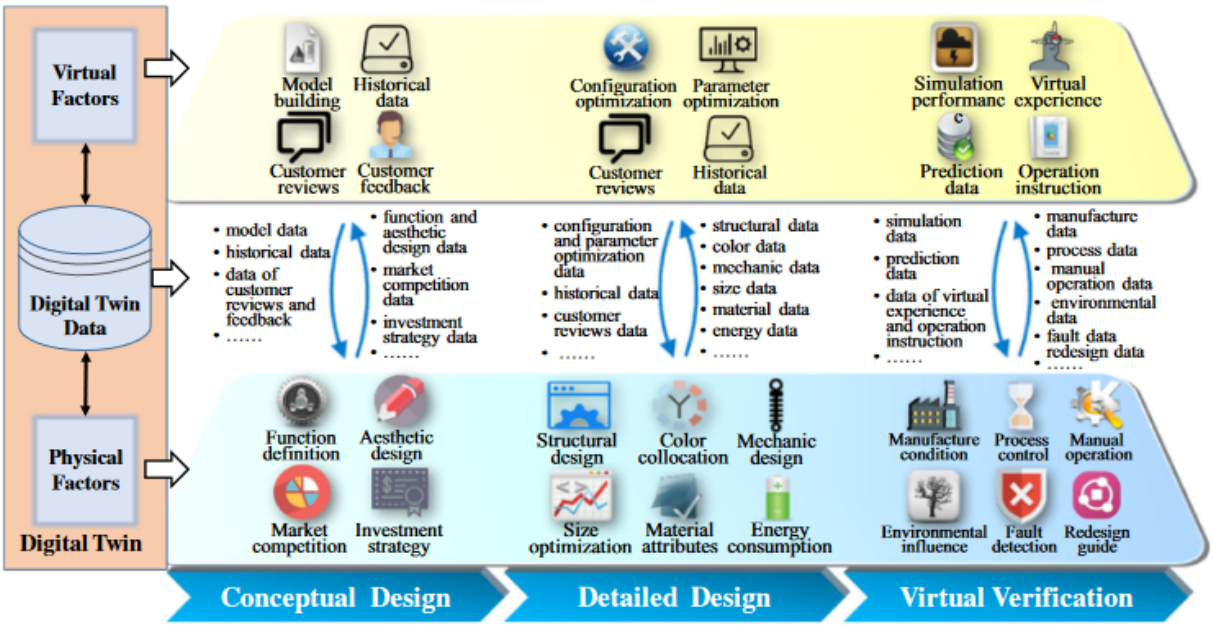


Figure 9: Example of Digital Twin Product Lifecycle Management (PLM) flowchart [4], highlighting how a Digital Twin would be most useful if developing with the process or product, instead of being "siloed"

Digital Twin iterations to improve the new design [4].

It is also suggested that Digital twins could be utilised after a product has been produced in a manufacturing process. For example, if making a car a Digital Twin could be used in the factory itself, but also in how the car is operating after its actual manufacture. Data such as energy usage, fuel economy, wear points, and so on can be gathered. If a sufficient Twin of the car is created and recorded, these factors could be related to parameters that change in the car. For example, the material of construction. This information can then be fed into future Digital twins to optimise for greener materials.

2.2 Crystallisation

Crystallisation is one of many separation and purification methods, defined as a phase change where a crystalline product is obtained from solution [6][28]. With crystallisation the dispersed phase which seperates out as solid particles, also forms the final product, and as such has to meet any required product specifications. This means meeting an ever tightening requirement of size distribution, morphology, agglomeration, and so on - whilst also balancing yield, costs, and sustainability factors [29].

This is a step in a wide variety of industrial processes, from pharmaceuticals to agriculture. Each sector having their own challenges and requirements. Pharmaceuticals for example have stringent requirements to meet for their customers, and require smaller crystals produced by a process such as milling. Agriculture instead may require larger crystals for a controlled release, and may even coat them to delay release [30].

The crystallisation process, to meet these ever more stringent requirements, has to be controlled in a robust and reproducable manner. With sustainability as a key parameter [5]. A number of factors that have to be optimised include:

Polymorphism control, which can affect performance of the product or the manufacturing process itself. This means ensuring you are forming the target polymorph in the case a crystallisation can result in more than one different type.

Morphology is another critical control point, which again can affect performance and manufacturing. Needle-like crystals for example, can cause difficulties in a filtration step as they plug the filter and block liquid from passing [31].

Crystal size distribution is also important, demonstrating a weighting towards the mean crystal size, which is a key KPI, but also potentially causing other problems. Fines for example if present in too large a quantity may lead to caking and then other processing problems, such as impurities. It is important to note that with CSD, you can get varied results for what an instruments calculates, due to the way they each use a different characterstic measurement. For example, CSD from number will often be very different to CSD from volume [32].

Impurities are a critical control point of their own, and can be caused by side-reactions occurring, from previous steps in the process, inadaquate washing, or simply due to the "mother liquor" being trapped within crystals as inclusions.

Agglomeration can happen to varying extents and can encourage the inclusion of mother liquor, or the generation of fines during transport, to name a couple of issues. Factors such as Solvent choice, seeding, and temperature cycling can influence this [33][34].

Growth rate is a major factor, the longer it takes for a crystal to grow and the separation step to complete, the more costly the process to achieve the same product.

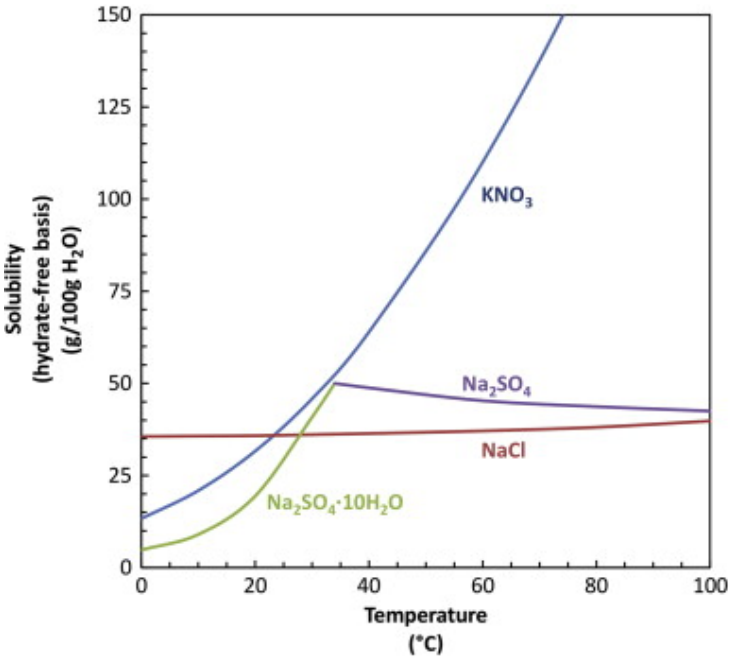


Figure 10: Examples of solubility curves of various compounds in water [5]. This a key factor in controlling crystallisations as once above the solubility curve, supersaturation is present.

Sustainability, not simply the economic impacts are considerations that must be considered at each stage as well. A slower growth rate and longer batch times may potentially lead to higher energy consumption, for example. This could also be the case with requiring additional purification steps due to excess fines, or incorrect morphological control.

There are a number of ways to control a crystallisation process to optimise these KPIs.

The main driving force of a crystallisation is the solubility of the crystalline compound in the solvent it is dissolved in, otherwise called the difference in chemical potential [35]. To crystallise the dispersed phase out of the solvent, we have to decrease the solubility until the dispersed phase crystallises out of the solvent. The rate this happens at is dependent on several factors, and is known as the crystallisation kinetics.

Solubility is a function of solvent composition and temperature, and as such we can decrease the solubility to the point we achieve supersaturation, by varying these factors. We can change the temperature, evaporate some of the solvent, or add an antisolvent. This supersaturation can be represented by the supersaturation ratio as in equation 1, which is a function of the concentration c and solubility $c*$, at a specific temperature and pressure.

$$S = \frac{c}{c*} \tag{1}$$

The solubility and how its relationships with solvent and temperature are different for each compound, as can be seen in Figure 10.

This highlights the need for different methods of crystallisation to extract these compounds efficiently. A summary of these methods are shown in Figure 11.

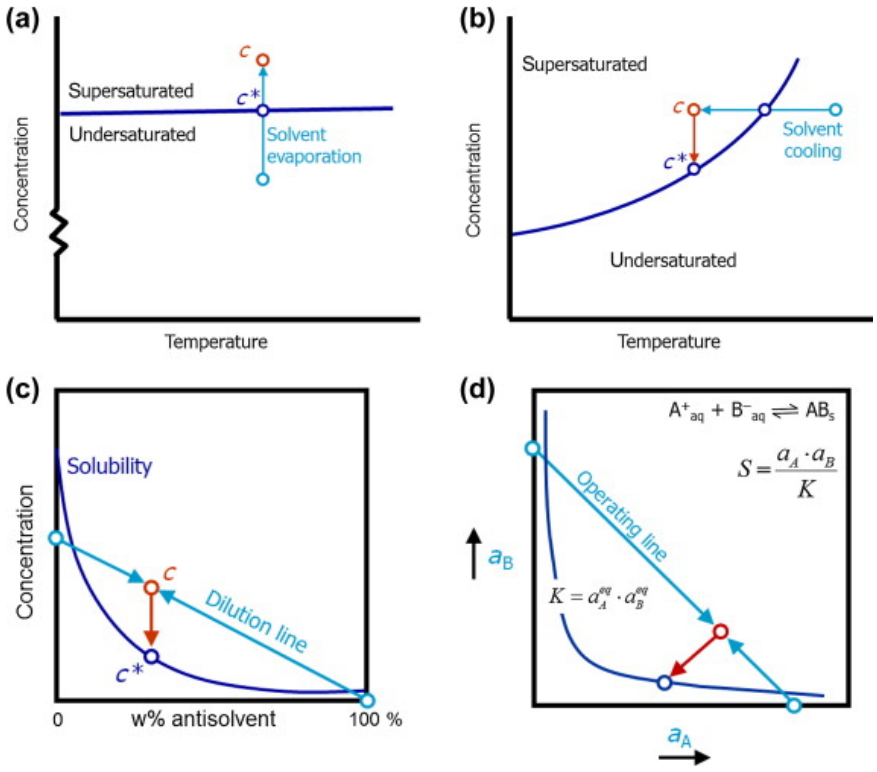
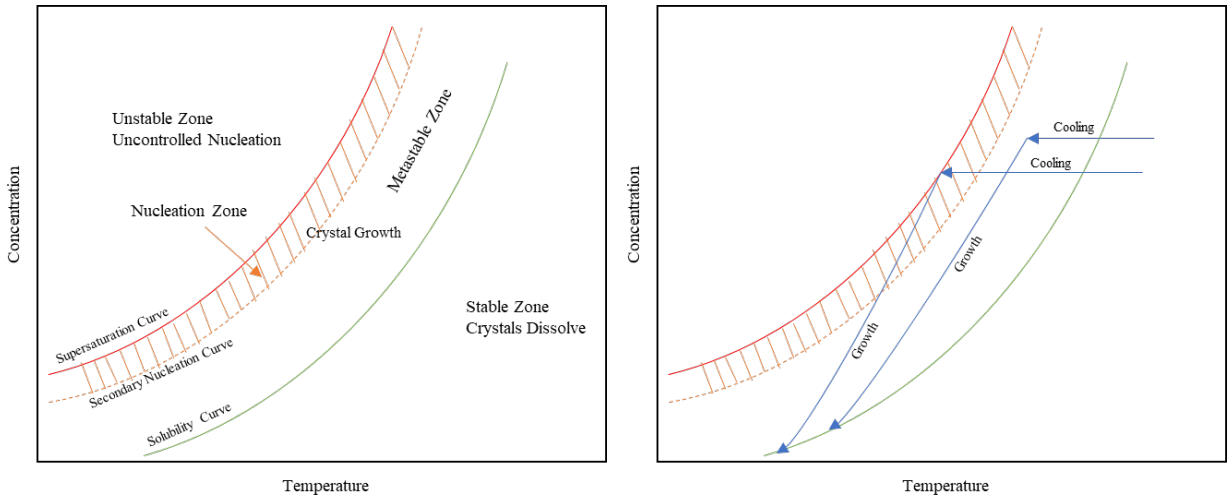


Figure 11: Phase diagrams for four crystallisation methods - (a) Evaporative; (b) cooling; (c) antisolvent; (d) precipitation [5]. Each of these provide a driving force which causes crystals to form at a rate proportional to said driving force.



(a) The Different Areas Crystallisation Zones in a graph of Concentration vs Temperature

(b) Example Cooling Profiles, showing just two different paths a cooling crystallisation could follow. These result in a different driving force, and one pushes further into the metastable zone to the boundary of primary nucleation, which as such would see a larger number of smaller crystals form. This is a common method for unseeded crystallisation, where the initial crystals are formed via this primary nucleation

Figure 12: Crystallisation Zones and Thresholds [6] [7].

When controlling a crystallisation, a main point of consideration is where the setpoint is on the supersaturation curve, and how well this is controlled. You want the crystallisation to follow a particular path to achieve the properties that you require. This is demonstrated visually in Figure 12. By changing the path the crystallisation follows, you alter the number and size of crystals produced.

There are numerous ways for new crystals to form, primary and secondary nucleation groups these. Primary nucleation involves the formation of new crystals, be it homogeneous, where the new crystals form spontaneously in solution, or heterogeneous, where new crystals form on foreign objects such as dust, walls, or the air-solution interface. Secondary nucleation is when new crystals form from the presence of crystals that are already there. This can be from shear from an agitator, collisions between particles, and many other sources.

This competes with crystal growth for each particle of solute, and the final composition of a batch crystallisation is related to how much nucleation occurs in relation to the growth rate. Little nucleation but a high growth rate means large crystals and few fines. High nucleation with little growth means many fines, and smaller crystals.

Crystallisations are complex and difficult to model with high accuracy. They are affected by numerous factors, to the point that a model or simulation which is accurate for one experimental can fail once scaled up, and it is still a challenge being studied [36].

2.2.1 The Challenge of Optimisation

As previously mentioned, simulating or modelling crystallisation processes is not a simple process. It can be done [37], but with narrow design spaces that often fail to work as well in industry. Unintended fluctuations

can occur, such as secondary nucleation outbursts. These throw off predictions and prevent optimisation.

This is partly because crystallisation is fundamentally a molecular process, with changes at the molecular level affecting the performance of the bulk due to changes in how it crystallises. Although research is pushing for improved prediction of how these behave at a molecular level [38] [39], in industry there is always going to be some level of unknown, which can't be predicted.

As such, optimisation specific to a system has to be carried out. This can be at lab-scale, but will then have limited applicability. To truly optimise a crystallisation system we have to be able to optimise it in the environment the actual production takes place. To enable this, in-line analytical techniques are helpful and likely required, such as FBRM. These are also being developed [40] [41] [42], and used to control crystallisation processes to keep them in optimal conditions. By combining the ability to know what is going on in real-time, with knowledge on how to respond through predictive [43] [44] [35] control, we can in optimise the process dynamically. This allow us to solve issues such as the secondary nucleation outbursts, as we can say respond by dissolving some of the excess fines. This requires a balance between the KPIs we discussed in Section 2.2.

Recent work has shown the benefits of restimating kinetic parameters during running, in order to deal with changes in parameters, or uncertainty in initial estimates [45]. The limits of this real-time optimisation do not seem to be comprehensively explored however. This demonstrates an example of how real-time optimisation can be approached to deal with fluctuations and uncertainties in process conditions, however. It also demonstrates a different method of obtaining crystal size distribution (CSD), working backwards to infer the CSD from the Chord Length Distribution (CLD) by mapping out the expected CLD from a certain CSD, instead of mapping out the expected CSD from a given CLD.

Industry face the issue that a process changes over time. Raw materials are not always consistent, suppliers are changed, sustainable alternatives are sourced. Optimisation and control therefore needs to be able to be dynamic and shift for these changes, or be able to be carried out in a cheap and quick manner. The comparison of situations is also important, and being able to simulate or efficiently evaluate the extent to which a change will affect a crystallisation is therefore critical. This is one reason Digital Twins may be an important push forward if achieved.

2.2.2 Digital Twins in Crystallisation

Digital Twins have been used in crystallisation systems already, though it is often used as a Digital Simulation of the system, and offline optimisations performed. Not as a real-time representation of the process [46] [47], as previously stated as the intention of Digital Twins. Though proper twinning has also been seen and is certainly being researched [48].

2.2.3 Simulation Development

Due to the cost of experimental work, and subsequent cost of optimisation, simulation is critical in crystallisation process development. Even used offline they can assist with design and optimisation significantly, and with the proposition of linking to the PLM lifecycle with Digital Twins, are a required step to achieve good Digital Memory of a process.

For our specific aims, we need a sufficient amount of data to carry out surrogate modelling and any further machine learning techniques. Especially for methods such as reinforcement learning, this means significantly more data than would currently be feasible to obtain by experimentation alone. Though we may look at reducing the simulation step in later work, pretraining using simulation is still recommended [49].

Population Balance Modelling Population Balance Models are used to calculation how a crystallisation is going to progresses. This is a mathematical description of how a number of particles change during a process [6]. For these to be developed, kinetics are required to calculate how the main mechanisms are affecting crystal growth and nucleation [50] [51]. It is in determination of these kinetics that uncertainties lie, and limit the application of these population balances to industrial scale crystallisation. Early on this became apparent as experimental determination of growth kinetics varied significantly between experiments [52].

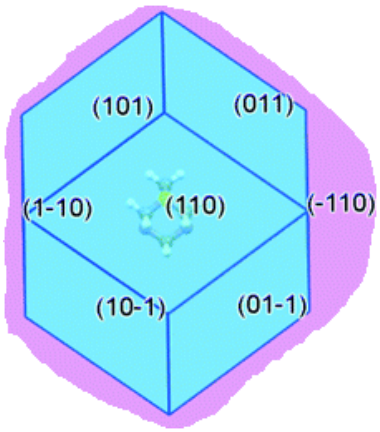
These population balance equations can be solved by software such as gPROMs, or more fundamental methods such as the quadrature method of moments or method of characteristics [14]. The population balance can be represented in either a 1D form, where it is assumed that the crystal is cubic and the kinetics are the same for each face, or 2D, where the crystals can have faces that change at different rates. The difference in error between these methods is assumed to be less when the actual crystals are cubic, and higher when the aspect ratio increases, for example with needle-like crystals. Therefore, a very high aspect ratio crystal such as very long needles - is likely going to be modelled significantly more effectively using a 2D population balance. In order to use PATs correctly, for example, a 2D population balance may be required to accurately control the process [36].

The Estimation of Kinetics Accurate determination of kinetics, as mentioned, is one of the main barriers in using Population Balance Modelling. For example take secondary nucleation - in small units, collisions between crystals and equipment dominate. In larger crystallisations, it is more common for crystal-crystal collisions to dominate. Not only this, but the hydrodynamics of the system also have an effect, secondary nucleation is more likely to occur at a higher rate in areas of high-intensity mixing.

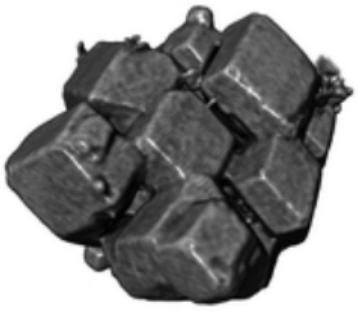
However, the advancement in computational capabilities are slowly allowing more and more complex models to be solved, which allows increasingly complex kinetics to be developed. Using nonlinear optimisation techniques and introducing additional parameters such as breakage and agglomeration [53], have assisted with making these models more accurate. Especially when combined with the in-line techniques discussed in Section 2.2.1. These in-line techniques have also been suggested to show that even different cooling rates cause different regimes, and as such different kinetics [54].

2.2.4 System of Choice

As a model system to begin the project, the initial crystallisation of choice will likely be a batch cooling crystallisation of Hexamine. This is a cubic unit cell, with a single family of growth faces which results in a rhombic dodecahedral shape as in Figure 13a. A simulation has been developed by a previous PhD student, Ryan Leeming, in gPROMS. Hexamine is known to experience agglomeration, as in Figure 13b and as such the kinetic parameters of such will likely need to be considered [8]. But are not currently present in the simulation.



(a) The Spatial Structure of Hexamine, a cubic unit cell which results in a rhombic dodecahedral shape.



(b) An example of Hexamine Agglomeration, the rhombic dodecahedrons agglomerate over time to form these clumps.

Figure 13: Hexamine - Diagrams and Examples [8]

2.2.5 Process Analytical Techniques (PATs)

Process Analytical Techniques (PATs) are critical for the monitoring and control of crystallisations, especially if this is not an open-loop design, and feedback or feedforward control is desired. A recent review of these techniques [9] demonstrates the range of possibilities here. This review highlights that there is a lack of multi-objective optimisation studies, which would involve the balancing of KPIs which a specific weighting or importance in mind. A summary of PAT techniques can be seen in Figure 14. Though, control of the main factors such as polymorphism or fines using temperature has seen significant exploration. Some combinations of control such as temperature combined with antisolvent addition have also been considered [55].

PAT Tools	Working Principle	Application
FBRM	Receive the reflection signal generated by the laser beam through the particles, and calculate the time the laser beam passes through the particles to obtain the chord length distribution	Real-time monitoring of particle size and particle number changes as well as nucleation and growth processes [13–15]
PVM	The video microscope on the probe can intuitively display the particle status, continuously capture high-resolution images under various process conditions, without the need for sampling or offline analysis	Monitor nucleation, growth, coalescence, fragmentation, shape evolution, and polymorphic behavior of the crystallization process, endoscope probes can detect color-related changes [16–18]
ATR-FTIR	The crystal sample absorbs infrared light at the incident frequency, the reflected light intensity is reduced, and a spectrum similar to the transmission absorption is generated, thereby obtaining structural information of the chemical composition of the sample surface	Infrared measurement of solute concentration during crystallization [8]
ATR-UV/Vis	The ultraviolet-visible absorption spectrum is an electronic spectrum, which is produced by the transition of valence electrons. The composition, content, and structure of a substance can be analyzed, determined, and infer determined by using ultraviolet-visible absorption spectrum and the degree of absorption of ultraviolet-visible light by substance molecules or ions	In situ monitoring of nucleation [9], crystal form transformation and supersaturation changes [10]
Raman	Raman spectroscopy is a kind of scattering spectrum, which is analyzed with different incident light frequencies to obtain information on molecular vibration and rotation. Raman frequency offset is represented by the abscissa, and Raman intensity is represented by the ordinate, which is complementary to the infrared spectrum and available for analyzing information related to intermolecular bond energy	Identify differences in polymorphic forms and enable qualitative and quantitative studies on solvent-mediated transformation, solution concentration, and the ratio of different crystalline forms in solid mixtures [20]

Figure 14: The main PAT areas for controlling crystallisation [9]

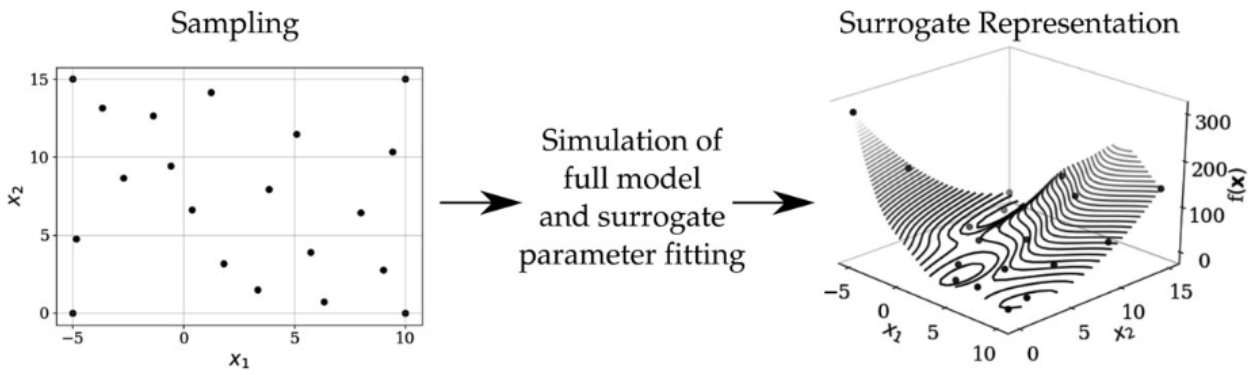


Figure 15: A simplified example of a surrogate model of two inputs and one output variable [10]. The variables are approximated across a surface, which is effectively a map of the output that is estimated from the input variables at each position on the surface.

2.3 Surrogate Modelling

A surrogate model is a model which approximates a usually higher fidelity system. This is often something akin to a simulation with a higher computational cost, which would be too intensive a calculation to run across enough permutations to train most machine-learning implementations. The surrogate model in this case reduces this computational cost and the time required for each permutation drastically, allowing data to be generated at a speed that is sufficient to train even more data-intensive machine learning models, such as reinforcement learning. There are examples of machine-learning based surrogate models being utilised to provide training sets for techniques such as multi-variable and multi-objective optimisation [13].

Surrogate modelling is used in many Chemical Engineering applications [10] to simply complex reactions and interactions, allowing them to be optimised, especially in cases where the full complexity of the process is not understood. This is often done by mapping a complex simulation onto a surface, allowing the outputs to a given input to be estimated, with the assumption that the surface accurately represents the system. An example with two inputs and one output is shown in Figure 15.

The determination of how to create a surrogate model, and what type should be used, is not simple. Several factors are involved, complexity, noise, mixed data sets, and so on. Though comparisons of surrogates can be done [56] [57], these are limited to the model they are done on, and usually cannot be extrapolated to other systems easily.

There also exists the problem of surrogate complexity. In order to accurately approximate a system using a surrogate model, you want it to be able to predict a wide range of inputs and outputs. You can do so by expanding the design space as in Figure 16, until you have a sufficient level of coverage for your application. However, the more you do this the more likely it is that the surrogate loses its benefit, and becomes computationally complex itself. As such a balance often has to be struck.

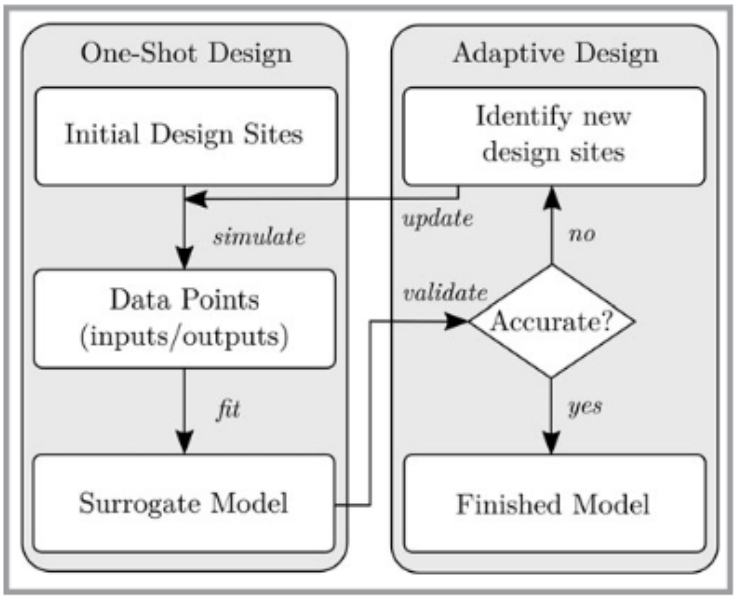


Figure 16: An example of expanding surrogate design space [10]. This would be iteratively run until desired accuracy is reached, without creation of an overly complex surrogate.

The main types of surrogate models include Polynomials, Kriging, Artificial Neural Networks (ANNs), Radial Basis Functions (RBFs), Moving Least Squares (MLS), and many others. The intricacies of these will not be discussed further in this review but McBride et al.[10] provide a good cursory overview. Instead of choosing a single method, it's possible to use a weighted combination of methods, where the error from the accurate prediction is measured and a summation of the separate surrogates is combined in a weighted manner, with the surrogate providing the highest weighting being the one that has the least error. This is often known as an ensemble of surrogates [58] [59] [13].

2.3.1 Neural Networks

As discussed in Section 4.2, there are different types of surrogate models that can be made, to best approximate the input to output relationship. However, due to one of the aims of the project being the use of historical data generated through process development, or data from similar systems, neural networks will likely be the most applicable. This is due to being able to retrain the networks to varying extents, to allow them to learn from new data, as has been demonstrated previously [49]. As such, this is the focus of more in-depth literature review, though this may change further into the project if they are discovered to be less suitable. These have been shown to be able to create a surrogate model of PBMs in previous literature [**?**].

A neural network is intended to replicate the way a brain functions by acting as a complex, nonlinear and parallel model/computer. They can generalise information, allowing the solution of very complex problems, but can rarely solve problems that are not split into a number of relatively simple tasks, which match their actual capabilities. As mentioned, they can deal with nonlinear problems, and are adaptive, meaning the synaptic weights which link each neuron together can be changed in real-time, allowing the development of robust models which cope well with disturbances. At the most basic level, these neurons are developed as shown in Figure 17.

The inputs are weighted by the individual synaptic weights, and summed into an output. This output is

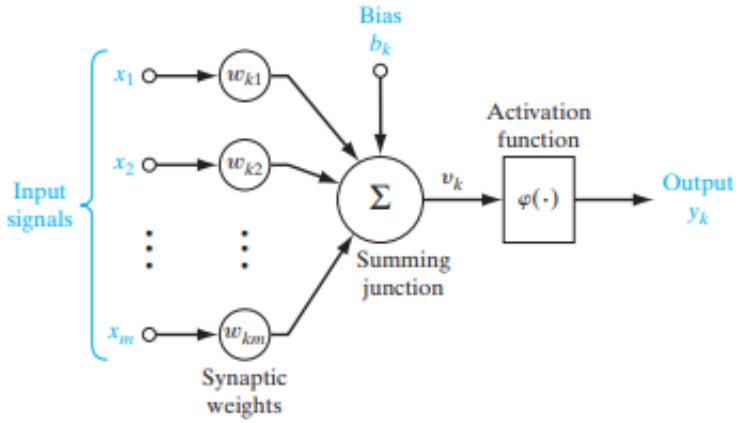


Figure 17: Example of a Neuron [11], containing the basic functions including snaptic weights, a summing junction, and activation or threshold function.

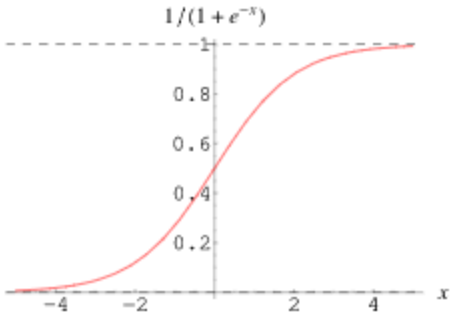
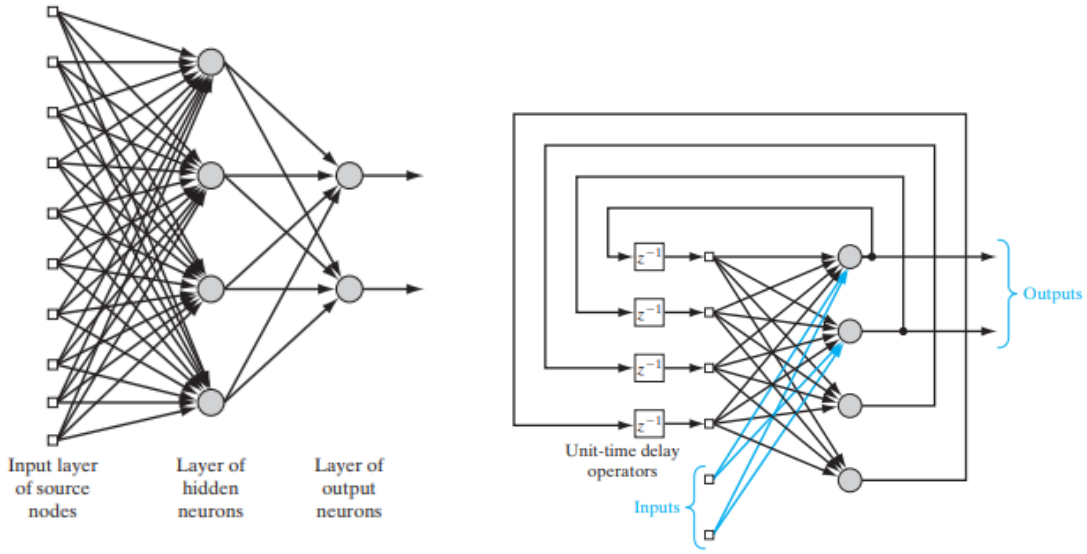


Figure 18: Example of Sigmoid function. A common version of activation function in a Neuron as in Figure 17.

summed with a bias, and passed through the activation or squashing function, which scales the actual neuron output to a given value. Each of these factors can vary depending on the neural network. For example, the activation function could be a threshold value that allows no output until above a certain value. It could also be sigmoid, in that the output varies in a sigmoid pattern as in Figure 18. This is possibly the most common type of activation function.

These neurons are combined into graphs or networks, with how they are combined also being a major factor in their operation. For example, they can be combined in a feedforward system, where information only flows forwards, as in Figure 19a. Or they can be combined with a feedback loop, known as a recurrent neural network, where information also flows backwards, as in Figure 19b.

These neural networks (NNs) are optimised via an error/cost function. Here the error between actual and predicted and measured, and minimised through trying different values. There are many different ways to optimise this cost, some are better at more quickly converging to a solution than others, dependant on what you are trying to predict using the NN. One of the more common methods to achieve rapid convergence is known as gradient descent, where quite intuitively the gradient or change in cost function is used to determine whether you are heading to an optimal solution, or not. This again has many varients to deal with different issues such as getting stuck in local optimums [60].



(a) A Feedforward Neural Network, where information flows in one direction only.

(b) A Recurrent Neural Network, where information may also flow backwards.

Figure 19: Neural Network Example Types [11]

A major breakthrough in NNs came when deep learning was developed and refined to be efficient enough to use in real applications, though the concept did not originate with neural networks themselves [61], along with the required computational power to apply it. Deep learning at the basic level simply means an NN with multiple hidden layers [62]. These multiple hidden layers all have weightings to each other like a conventional NN, but we are not able to see the inputs or outputs. Hence they are known as hidden, similar to a black box algorithm. The idea remained relatively unused until new techniques allowed large deep learning networks to be trained efficiently [63] [64], through training one layer at a time. Allowing more complex concepts to be mapped using these deep learning techniques. This is because instead of a single hidden layer of neurons, many layers could now be used and trained.

These have been used to control crystallisations before, trained on a "golden" reference data set, and then fed the experimental data to complete the neural network and its responses to given parameter shifts [65]. In this they perturb parameters that affect the final CSD obtained, and the control method succeeds in returning to the CSD setpoint using a measured CLD as an error measurement to the reference CLD. This is however only done by altering the cooling profile, not by actively changing the other possible parameters.

Physics-Informed Neural Networks A type of neural network that may yield a higher ability to extrapolate from beyond the domain the data has been obtained from, which may be of use especially once the project begins to look more into using past experimental data, or data collected during operation, is the Physics-Informed Neural Network (PINN) [66]. While most types of Neural network are purely data-based, PINNs use the fundamental equations that inform how a system operates to refine the loss function of a neural network, meaning that these equations have weighting on what the outputs will be from a given set of inputs. As we already have a set of fundamental equations that can inform how a crystallisation progresses, these may be of keen interest in creating an accurate surrogate model. These have been used in the past to

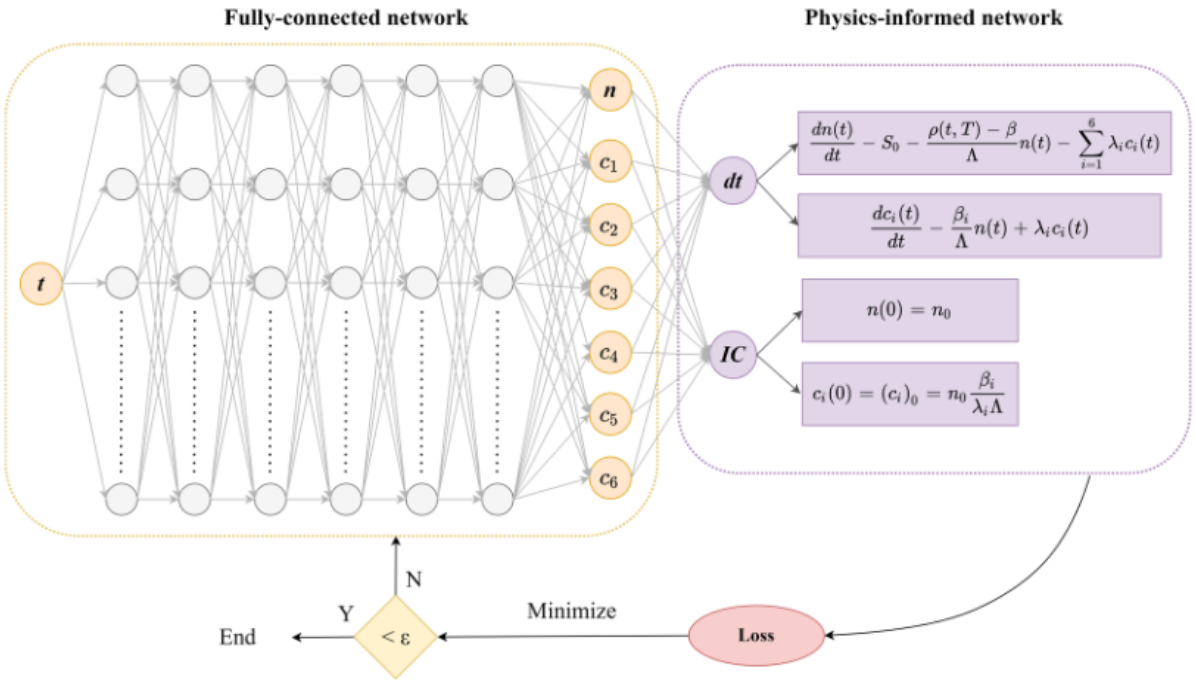


Figure 20: An example of PINN structure [12]. The Partial Kinetic Equations are combined with the Neural Network via a loss function.

this effect in the modelling of nuclear reactors [12] and solving fluid mechanics problems [67]. An example of their structure for the solving of partial kinetic equations can be seen in Figure 20.

The key benefits of this type are that they require less data than the purely data-based NNs, through this encoding of fundamentals. The extent to which this improves extrapolation, and reduces experimental or other data requirements, has not been thoughroughly investigated. Their application to crystallisations is sparse, though they have been linked to MPC in recent work [68].

Transformer/Self-Attention Neural Networks Transformers are a relatively new type of neural network which was built with long-term memory in mind. As such, they work more effectively than conventional neural networks when used in situations where "previous" data affects what the next data shall be. This could be any sequential system, time based being a common example, sentance structure being another. These neural networks have seen most of their use in large language models, where they predict the next word in a sentance based upon the past words in said sentance. This can be used to translate, or create chatbots with memory [69]. They can also be used to map amino acid sequences to properties used for drug discovery [70].

Recently these are being more keenly investigated for their usage in modelling time-dependant physical systems [71], where said systems change dependant on multiple time scales, such as chemical reactions or molecular dynamics.

Hybrid Surrogate Models Hybrid surrogate models combine several different surrogate modelling methodologies, be it a polynomial or a neural network, or several different types of neural network, and weighted them according to the error that they give at each data point. This leads to a theorhetically higher accuracy

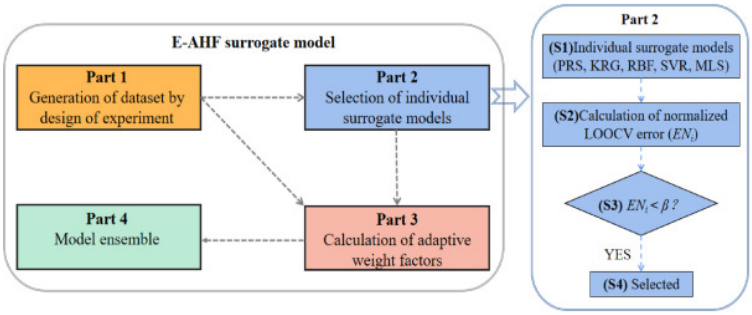


Figure 21: Example of the development flow for a hybrid surrogate model [13], sampling method and individual surrogate models used may be changed. PINNs, transformers, etc... as discussed above may be used, for example

surrogate model, as the best fit for the given data point is found, instead of finding the best overall fit that will often have at least some level of error for at least a portion of the data. This has already been applied in multi-variable and multi-objective optimisation [13], as discussed in Section 2.5.

This is again a balance between the accuracy of the model and computational complexity of the surrogate calculation, a flowchart for the development of these hybrids is shown in Figure 21. In this Jiang et al. suggest using Design of Experiments to generate a dataset, whether or not this is the preferable sampling technique to a near-random method such as Latin hypercube sampling, may need further review depending on method.

A factor that isn't investigated as far as was found in this review, was a comparison to the accuracy of individual surrogate models, and how much computational time increases due to using several individual surrogates combined, if it does.

2.4 Control

Controlling a crystallisation requires purposefully changing some of the parameters that determine the KPIs of the process, such as crystal size distribution. This has in the past been attempted using open-loop temperature versus time to optimise the required characteristics, using first-principles models with growth and nucleation kinetics, which are determined from a series of experiments, and uncertainties/disturbances taken into account [72] [73]. This is the model-based approach, where a detailed mathematical model is used with optimisation techniques to determine the optimal temperature-time profile [55]. Direct Design can also be used [**?**], assuming that the desired region is in the metastable zone of the solubility curve - where nucleation rate is low, but crystal growth rate is high. High supersaturation encourages growth, but also nucleation. Hence there is a balance between growth and nucleation to achieve the desired profile [35].

Model Predictive Control can combine these two approaches, using a predictive model of the crystallisation combined the the direct design feedback response, to achieve a more robust control system [14].

2.4.1 Model Predictive Control

Model Predictive Control (MPC) is a control method, that unlike open-loop control where an optimal profile is followed without any feedback from sensors or other meterology, responds to deviations from what is expected. Specifically, MPC does this by using a model to "predict" how a system will react to a control parameter change. It therefore knows how much to change that parameter by, to get back to a setpoint. MPC predicts a certain time period in front of the current state, and this is known as the horizon that it is predicting.

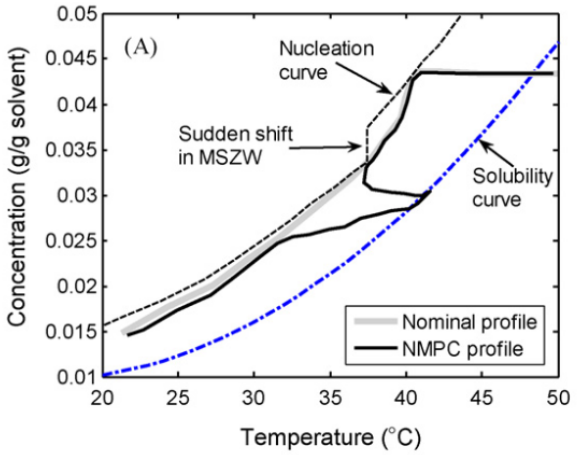
This has wide applications in industry and has been used for decades to control processes.

To cope with uncertainties in model predictions, methods such as Extended Kalman Filters (EKF) are used. These help make the MPC robust, in that it can cope with parametric uncertainties inherently, by determining how uncertain a given model prediction is, and ensuring the target is within the confidence region [74].

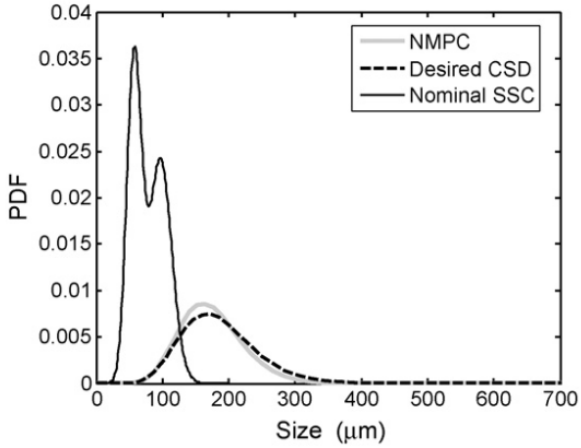
MPC in Crystallisation MPC has been used in crystallisation control in a number of forms [14] It has been applied to respond to changes that occur during the crystallisation, such as sudden spikes in secondary nucleation and as such the introduction of fines. In this case, the NMPC Nagy developed responded by significantly increasing the temperature, instead of following a traditional cooling ramp. This is due to the NMPC predicting the final CSD is going to be incorrect due to the introduction of fines, and as such these fines go through dissolution, the concentration actually increases for a period as crystals dissolve back into solution, and then the cooling rate is returned to normal as the NMPC iteratively predicts a CSD closer to the target.

This is only possible through measuring concentration, temperature and critically CSD - using an FBRM with a conversion algorithm, to convert chord length distribution into the CSD. The system can therefore detect the introduction of fines and respond appropriately. See Figure 22a for a visual of this response, and Figure 22b as the resultant improved final CSD compared to the nominal base case and target.

MPC Optimisation Separate from optimisation of the system, MPC itself could also be optimised. As this uses a model to predict and control the system, how accurate the model is directly affects how well it can control said system. Therefore, improving the accuracy of the model can improve the control efficiency. In the



(a) NMPC responding to introduction of fines. The fines are dissolved after being detected by the FBRM, as MPC responds by increasing temperature.

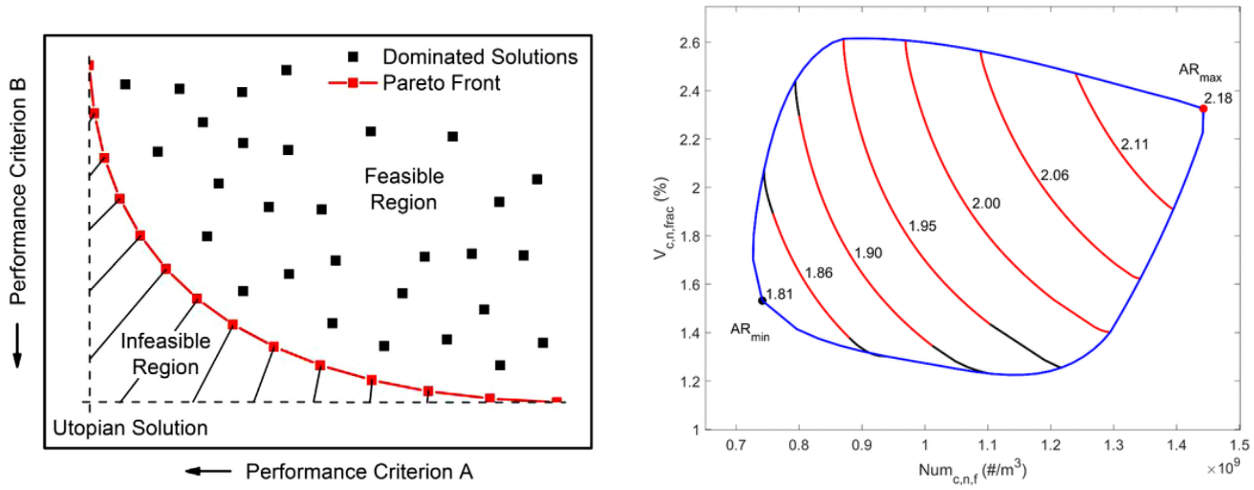


(b) NMPC final CSD compared to target and nominal base case, as a result of this robust MPC control.

Figure 22: NMPC robust control and its response to perturbations introduced purposefully to the crystallisation [14].

case of control using a surrogate model, with say an artificial neural network, this would mean achieving an improved convergence between the surrogate and a higher-fidelity estimation such as a gPROMs simulation.

An example of this is a reinforcement learning based MPC used to control a batch crystallisation, which has been attempted combined with a Kalman filter [75]. Benyahia et al. also make use of Transfer Learning, which means training the reinforcement learning agent iteratively, continuously "learning" and optimising the MPC as the operation runs. This transfer learning can take a wide variety of forms, learning from the current system, or similar systems [76].



(a) A Generalised Pareto Curve [77].

(b) A Pareto balancing AR, Nucleated Number, and Nucleated Volume [78].

Figure 23: Pareto Curves.

2.5 Global Optimisation

Global optimisation in this case means finding the global optima that is achievable given the constraints that are presented. This will be different for each use case in a process, and unique to the industrial application. This is different to the optimisation of a control system as discussed in Section 2.4.1, as instead of reaching a given setpoint in the most efficient manner, it means finding what that setpoint should be. In the case of crystallisation, this could be making the largest crystals, or achieving the shortest batch time. In many cases it could be that the setpoint for CSD is known, and as such, the other main factors are the manipulated variables we are optimising, such as energy use or yield.

There is also the situation where multiple objectives are considered important, and a balance between them is required. This is known as multi-objective optimisation and is an area of interest in many sectors, chemical reactions [77] and processes [13] included. An example can be visualised in Figure 23a, which shows a pareto curve in which two theorised outputs are optimised, and the balance between them highlighted. In this, the optimisation algorithm searches for where the pareto front lies, which is along the boundaries of the feasible and infeasible regions. The numbers of calculations, experiments or steps to reach this pareto front is how the efficiency of a multi-objective optimisation algorithm is measured, in this case representing convergence with the optimal value. A recent example of this involves balancing Aspect Ratio (AR) and Nucleation Rate [78], seen in Figure 23b.

In Jiang et al., they use a hybrid surrogate to first carry out multi-variable optimisation, where they determine a value they want to maximise, and a value they want to minimise, outside the bounds of the training set. They use the surrogate to rapidly converge a genetic algorithm on these minimum and maximum values, and then compare this to what the intensive mechanistic model calculates, with relatively small error. The two objectives are then balanced using a NSGA-II algorithm, to give a pareto front that allows an informed decision on optimal operating conditions.

This could also be done with crystallisation using a similar workflow, as we have population balance

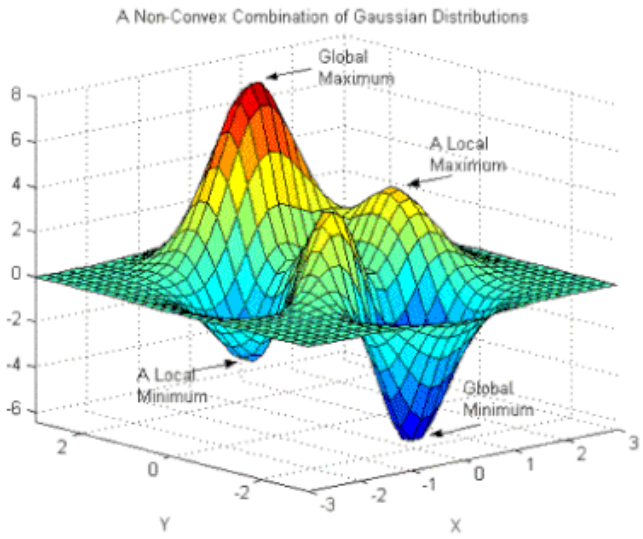


Figure 24: A Non Convex surface with a global optimum and minimum, several local maxima and minima representing possible issues for optimisation.

equations that could be intensive for a 2D simulation with high fidelity, and potentially multiple objectives to balance such as CSD and batch time.

A problem unique to some systems, is the concept of a nonconvex pareto front or optimisation process, and this may need to be considered depending on how the project progresses. Either for applying gradient descent, which is notorious for having trouble with local minima or maxima, or within a pareto itself. An example in Figure 24 shows this visually as a 3D-plot, where local optimums are highlighted, but a single global optimum exists in the non convex surface. This global optimum would be the optimal value the algorithm would want to converge to. The process of avoiding local optimums inherently adds additional experimental work or computational time, so a balance would need to be developed.

2.5.1 Deep Reinforcement Learning

Deep Reinforcement Learning has been utilised to optimise chemical reactions, namely to maximise yield. This was found to be more effective than other optimisation algorithms such as SNOBFIT or CMA-ES [49]. It also demonstrates the capability of being pretrained and then achieving faster convergence when predicting new, but similar data. As previously discussed in Section 2.4.1, reinforcement learning has already been used to control crystallisation processes to an extent, including the use of transfer learning [75]. It has also been used combined with feedback control via image analysis, to control for CSD in an antisolvent crystallisation of NaCl [79]. It should be noted they manipulate both the antisolvent flowrate and the temperature to achieve this, demonstrating its use with multiple manipulated variables, but its application to the global optimisation of systems has not been attempted yet as far as this review could determine.

2.6 Summary

In this literature review we have covered many topics at a shallow level, Digital Twins, Crystallisation, Surrogate Modelling, MPC Control and Global Optimisation. As orginally shown via Figure 1, the combination of these topics would potentially achieve the creation of an adaptable Digital Twin for a Crystallisation, which could be applied to a wide variety of different crystallisations. Surrogate modelling allows the MPC method to be achieved even with very complex simulations or population balance models, while the use of neural networks or a weighted method that involves such, should allow adaptive surrogates which can be retrained without losing their original inferences completely. The global optimisation of these systems and ability of reinforcement learning to adapt to a shift in the global optimum, reoptimising in accordance with the weightings of each multi-objective objective, would allow the solution of the problems raised by industry. Supply chains shift, process changes occur after outsourcing, and rigorous optimisations that were proven as such lose their carefully tuned KPIs. A control and optimisation method capable of adapting to this, like the workflow proposed, would allow KPIs to be reoptimised after changes occur, and improve this issue.

In terms of what has been studied already, each individual aspect has seen some research. Hybrid weighted surrogate models that involve neural networks have been developed, but not applied to crystallisations. Global optimisation has not been carried out significantly but is being further investigated recently using Paretos, however Reinforcement Learning has only been applied to control of crystallisations not multi-objective optimisation. The control of crystallisations using more factors than simply temperature is sparse, though a combination of antisolvent and temperature has seen a number of studies. The application of a workflow to multiple crystallisation systems to test viability is not often done.

3 Materials and Methods

Why are the methods chosen being used? How were they chosen and why. This includes experimental, modelling, and data analysis. Also explain the benefits and pitfalls of each method chosen, and how these have been mitigated as best they can. Also discuss things you may miss, or why parts may "come up short"

The objective of this research, as described in Section **??**, is to create a digital twin of a crystallisation system, that can respond optimally to emergent conditions, through the use of predictive modelling and control. This predictive modelling is intended to be done via including the usage of historical data from the crystallisation system under different conditions, which should contain inferences on what the current optimal response should be, from what has been observed in the past. This is intended to encompass recent system data as well, to enable a continuously adaptive control system, that in theory improves its responses over time.

As a proof of concept, this will initially be done with a single optimisation objective, endpoint crystal size distribution (CSD). The goal being to minimise the deviation of actual CSD achieved from the desired CSD. Unseeded crystallisation in inherently unpredictable, and results in a different initial CSD each time. As such, this was chosen because if this dynamic initial CSD can be measured and predictively controlled to achieve a desired final CSD, proof of concept of CSD control will have been achieved. Hexamine as batch cooling crystallisation has inherent difficulties in CSD control, due to its apparent lack of secondary nucleation Citation needed for this lack of secondary nucleation!. If unseeded crystallisation can be controlled, by primary nucleation and the resultant supersaturation trajectory to reach the desired CSD, CSD control of the Hexamine system may be enabled. Due to the lack of secondary nucleation this may still encounter significant difficulties, and as such dissolution, and any other factors that drive a change in CSD may have to be investigated at a later point.

An initial proposal on how the full digital twin for control can be seen in Figure 25.

The following sections contain how this has been attempted thus far in the project, and the reasoning behind such.

3.1 Hexamine System and Data Sources

A section on the hexamine system, where the data has come from, how that data was collected and validated. Any pictures, the methodology of the experiments that were carried out. Etc...

3.2 Population Balance Modelling/Simulation

A section that describes the simulations used, if gPROMs, or the population balance modelling behind it.

There are many different approaches to solving population balance equations, each with their own strengths. Methods such as standard method of moments (SMM) or quadrature method of moments (QMOM) only provide average characteristics, such as the mean size of crystal and standard deviation. They aren't capable of representing the whole distribution. Other methods such as method of characteristics (MOCH) or method of classes (MOC) can be used to predict the entire size distribution, but are more computationally intensive. The combining of these average characteristic and whole distribution solution methods have also been attempted, in order to achieve computationally efficient solution of a whole distribution [80].

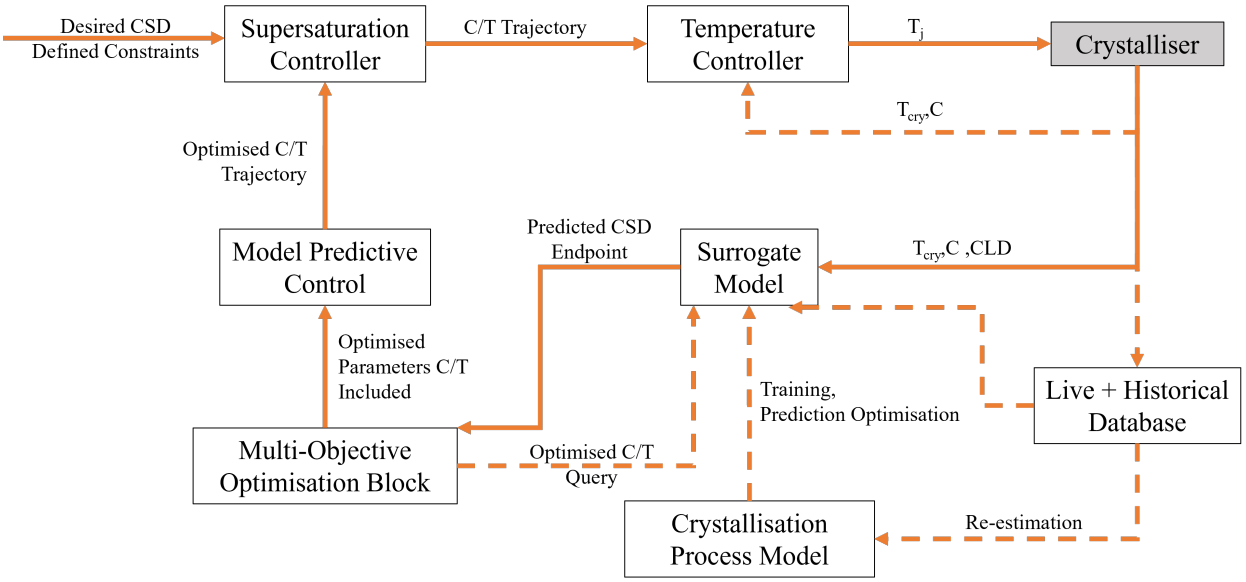


Figure 25: This theorised digital twin demonstrates how the adaptive control could be achieved. The supersaturation trajectory is continuously optimised to achieve the optimal endpoint CSD, for the constraints it is given. This is given to a supersaturation controller which maintains this desired supersaturation trajectory via changing the cooling/heating rate. The real-time status of the crystallisation is input to a surrogate model, which predicts what the endpoint CSD will be with the current supersaturation trajectory. The deviation between this predicted endpoint CSD and the desired CSD is then minimised through optimisation of the superaturation trajectory, again using the surrogate model to predict how a change in trajectory will change the endpoint CSD.

3.2.1 Kinetics

A section that describes the experimental setup used to validate and derive kinetics for the simulation/population balance model.

3.2.2 ASPEN

The parameters and functions that we have determined in sections X, Y, Z... are then input into an ASPEN flowsheet. This uses method X and Y... to model how the crystallisation progresses.

Paragraph/section on the methods that ASPEN uses

To train the supersaturation controller models, sufficient data was required on the how the concentration changes as the temperature of the crystallisation is reduced. This simulated data could then be used to initially train the supersaturation controller, for initial testing to be carried out. There pros and cons to this method instead of experimental, as is discussed in Section section needed on data collection, the use of simulations and their limitations and drawbacks, and the same for experiments. This simulated data will be stress tested on its applicability, and if found insufficient for the use of a supersaturation controller, potential hybrid models could be investigated. As will be investigated later in the project with the more potentially more complex crystal size control.

3.3 Surrogate Modelling

A section that describes the surrogate modelling done so far, and the methods behind these models.

3.3.1 Supersaturation Controller

A section on surrogate modelling of supersaturation, to build a controller, look at the paper Development of a recurrent neural networks-based NMPC for controlling the concentration of a crystallization process for an example of how this can be done - it goes into most of it!

In order to control the crystallisation, a common method is maintaining the supersaturation at a given setpoint, which is optimal for the desired crystal growth or dissolution.

A possible method, useful for simple to control crystallisations, is direct-design control. In direct-design, a predefined supersaturation profile is followed. This as such cannot respond to unexpected changes in the crystallisation, but may still be sufficient for systems with a low degree of variation between batches. To develop the predefined supersaturation profile, a series of experiments would have to carried out to determine what profile is optimal to achieve the desired crystallisation parameters, such as CSD, morphology or batch time.

This can be either absolute supersaturation control, which follows a constant supersaturation set point in the phase diagram [**?**], or relative supersaturation control, where a constant relative supersaturation is maintained instead [81]. This results in a decreasing absolute supersaturation during a batch, which limits secondary nucleation increases towards the end of batch run time. This has been shown to produce fewer fines, and as such enabled less costly downstream processing such as filtration [6].

Draw up a graph that shows the difference, as in Placeholder Example 26

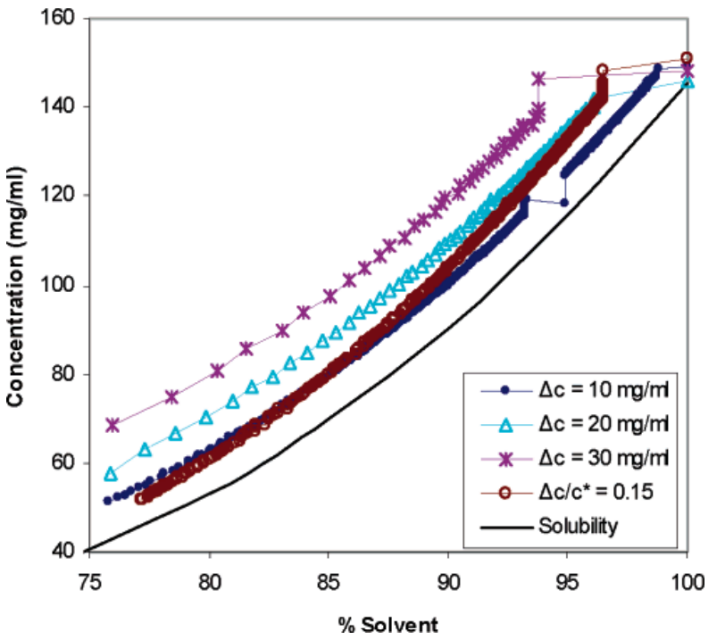


Figure 26: Change this to our own drawing! This shows the difference betyween a set supersaturation and a dynamic relative supersaturation, in which the absolute value gets lower as solubility decreases

For this research project, relative supersaturation will be the control point used, as this will likely allow finer control of the endpoint CSD due to this interaction of increased secondary nucleation towards the end of a batch.

We therefore need to control this relative supersaturation at the theorised optimal profile. The first step in this is achieve the relative superaturation control, and then to use this relative supersaturation controller to follow the supersaturation profiles that are determined to be optimal by the MPC controller of our desired characteristics, starting with CSD.

Artificial Neural Networks The first control method tested was an artificial neural network (ANN), this was to provide a method to control the relative supersaturation at a given set point.

Details on how the ANN was done and developed can go here.

3.3.2 Crystal Size Distribution

A section on surrogate modelling of CSD via recursion.

This surrogate model has to be able to predict the endpoint crystal size distribution (CSD), given a number of input parameters.

Artificial Neural Networks The simplist solution is often the best, and as such, the first method of neural network surrogate model was developed using an Artificial Neural Network (ANN). This is a feedforward neural network, where information only flows forwards, instead of backwards. Artificial Neural Network Reference Needed Here!

3.4 Model Predictive Control

A section on using the surrogate models to enable MPC to control for desired characteristics.

3.4.1 Optimisation

A section on how both the supersaturation controller and CSD controller will be optimised, such as the minimisation and maximisation functions and the way in which these will be applied.

3.5 In-Silico Validation

How will this system be validated in-silico? What is the full picture of the control and optimisation loop?

Present the theory behind the specific methods being used, equations, how the data will be transformed, what the theory and key litertature behind how this is done is.... Show the mathematics behind the models that we are trying, the pros and cons, how they will help achieve what we are researching.

Then present experimental setups, procedures to be used with them, how data will be captured etc...

4 Results and Discussion

4.1 Population Balance Modelling/Simulation

A section that describes the simulations used, if gPROMs, or the population balance modelling behind it, in terms of the results they gave.

4.2 Surrogate Modelling

A section that describes the surrogate modelling done so far, and the methods behind these models, in terms of the results that we got.

4.3 Model Predictive Control

4.4 Summary

What can we conclude from these results? What's missing? Etc....

Here discuss any current results of modelling or experimentation., how it links back to literature. What this proves or doesn't prove.

End on what this means for future work and how this will continue to be developed.

5 Future Work

What is the future work that will be done as a result of all this? Where are the research aims going forward, and why? How is it planned that these will be achieved?

To achieve the KPIs designated in Section 1.3, a initial stepwise project plan has been developed for the main areas of the project:

There are two main areas of optimisation that can therefore be carried out, one on the control method itself, and its ability to return to a setpoint, improving how quickly and efficiently the setpoint can be restored. Another to find the optimum response that balances several parameters, and from such inform the setpoint that should be targetted. As an example, a CSD may be set as a target setpoint, from an optimal created by exploration of a simulated environment. The MPC attempts to reach this setpoint and maintain it, by predicting what conditions will achieve this. The parameters of this MPC could be optimised, say by restimating parameters as in [47], or by retraining the surrogate modelling during operation with the new data generated, based on how much error the prediction has from the actual. As this occurs the system optimisation can explore, and update achievable CSDs, continuously refining the optimal parameters according to the weighting given to the important KPIs. This is represented briefly in Figure 1.

5.1 Work Breakdown Structure

For individual sections of the PhD work - a work breakdown structure will be used to plan out work packages, or tasks. This will mainly be used to plan out critical information and to visualise what is required for each section, to ensure nothing is missed. This will also assist with removing excess or unneeded work to ensure targets are met. See Figure 27.

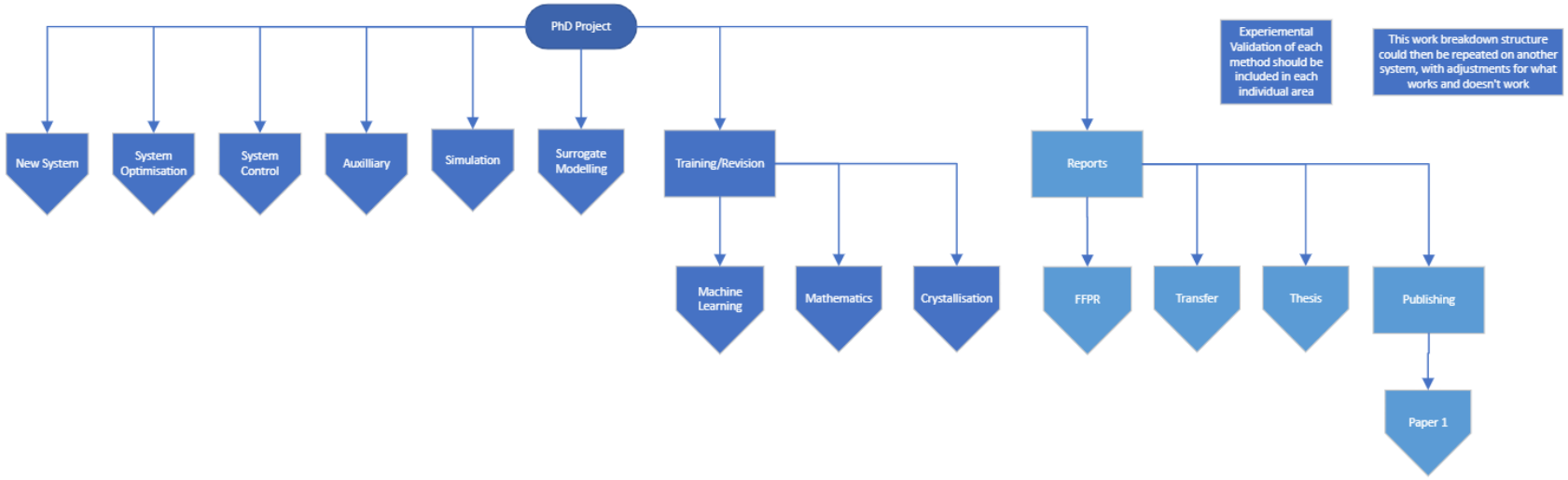


Figure 27: Initial Work Breakdown Structure

Individual sections will be broken down further into achievable tasks as in Figure 28.

5.2 Gantt Chart

As part of planning this project a provisional Gantt Chart has been created to demonstrate how the PhD may progress. As part of the Project Management aspect of this research this will be updated monthly and progress reviewed against planned to check progress is on track.

If progress is found to be unsatisfactory scope may be refined, depending on the reason for slower than expected completion. Please refer to Figure 29 to see the Gantt Chart.

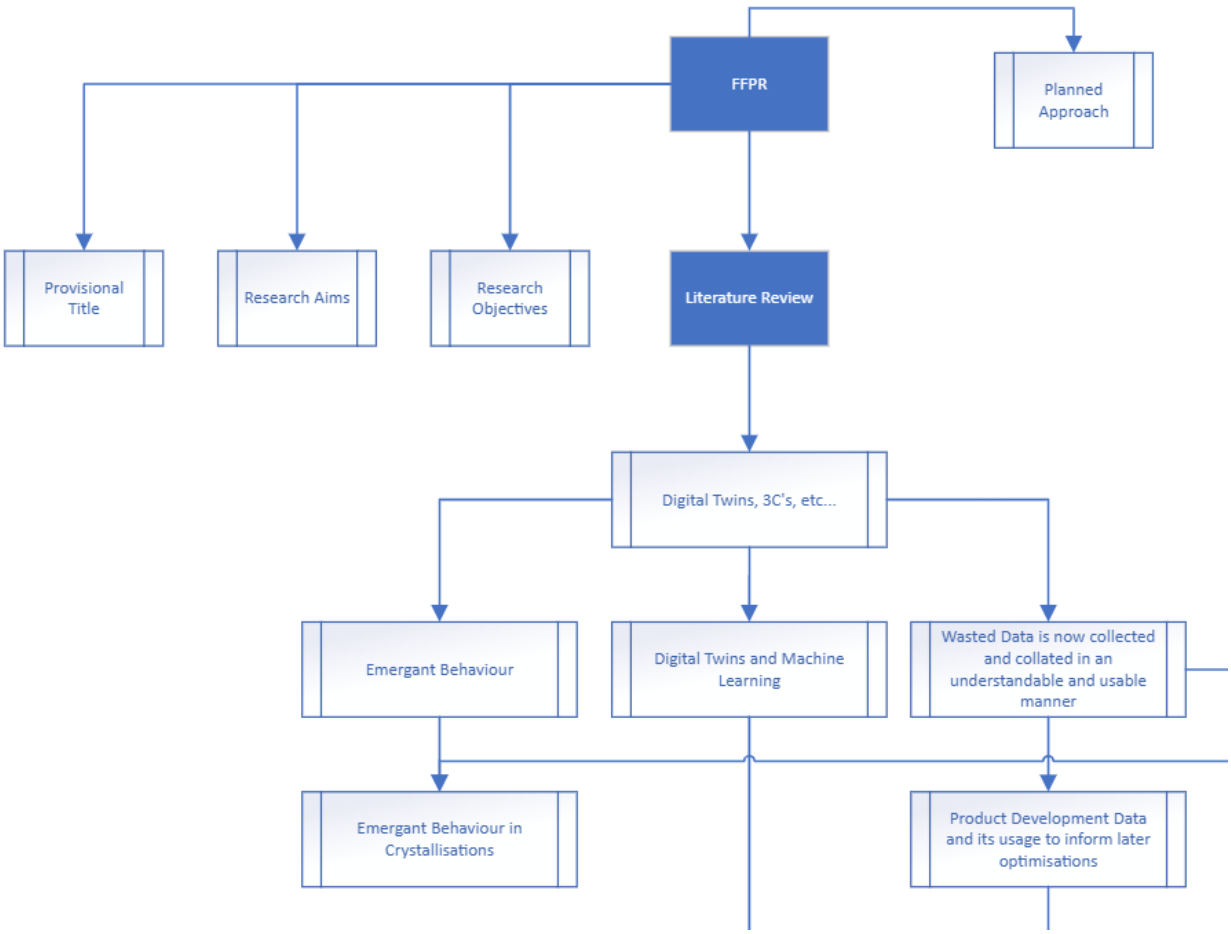


Figure 28: Example of an individual Work Breakdown Structure - Just a section of the full visio file

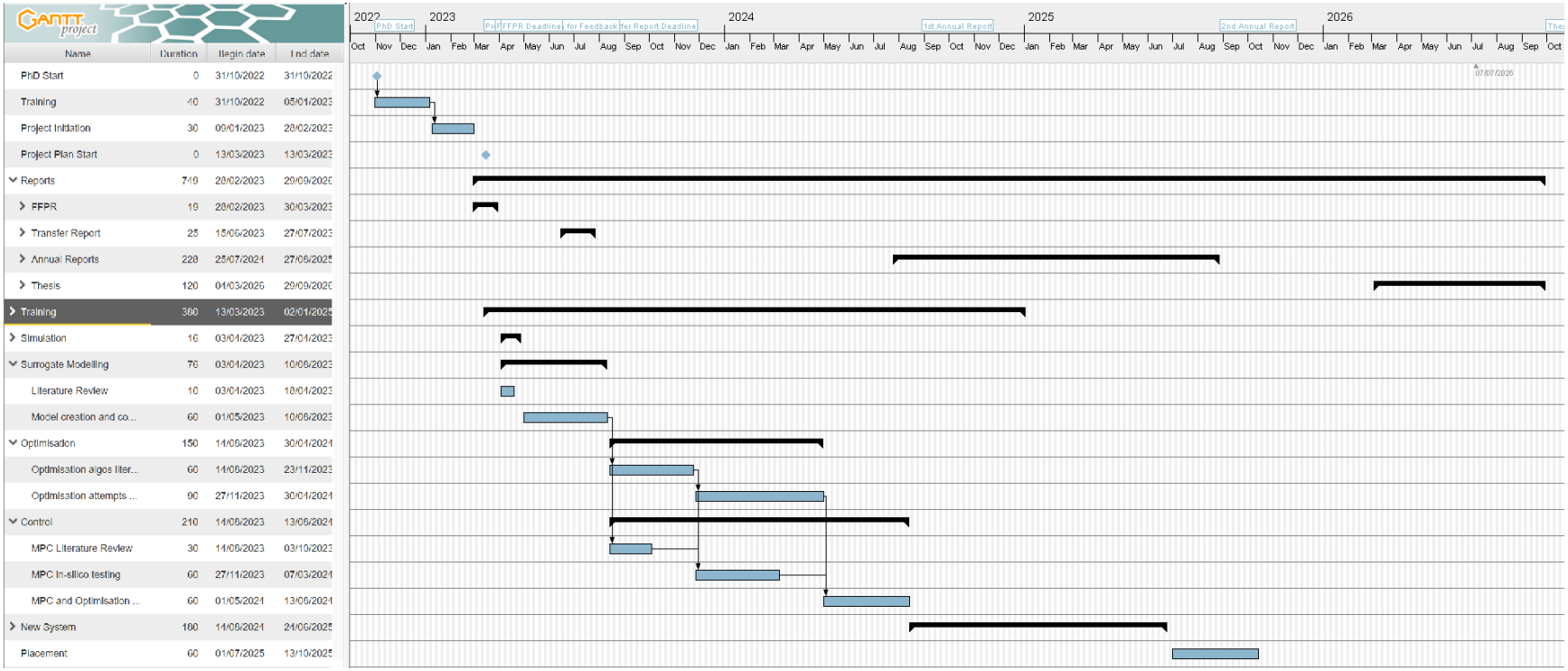


Figure 29: High-level Gantt Chart of Project

5.3 Training

Section on training planned to achieve what is needed, discussion on why and when this will happen.

Full details to be found in Section B of the Appendix.

References

[1] F. Tao and M. Zhang, "Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing," *IEEE Access*, vol. 5, pp. 20418–20427, 2017.

[2] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the Digital Twin: A systematic literature review," *CIRP Journal of Manufacturing Science and Technology*, vol. 29, pp. 36–52, May 2020.

[3] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital Twin in manufacturing: A categorical literature review and classification," *IFAC-PapersOnLine*, vol. 51, pp. 1016–1022, Jan. 2018.

[4] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology*, vol. 94, pp. 3563–3576, Feb. 2018.

[5] J. H. ter Horst, C. Schmidt, and J. Ulrich, "32 - Fundamentals of Industrial Crystallization," in *Handbook of Crystal Growth (Second Edition)* (P. Rudolph, ed.), Handbook of Crystal Growth, pp. 1317–1349, Boston: Elsevier, Jan. 2015.

[6] A. S. Myerson, D. Erdemir, and A. Y. Lee, eds., *Handbook of Industrial Crystallization*. Cambridge: Cambridge University Press, third ed., 2019.

[7] T. L. Threlfall and S. J. Coles, "A perspective on the growth-only zone, the secondary nucleation threshold and crystal size distribution in solution crystallisation," *CrystEngComm*, vol. 18, pp. 369–378, Jan. 2016.

[8] P. Gajjar, T. T. H. Nguyen, J. Sun, I. D. Styliari, H. Bale, S. A. McDonald, T. L. Burnett, B. Tordoff, E. Lauridsen, R. B. Hammond, D. Murnane, P. J. Withers, and K. J. Roberts, "Crystallographic tomography and molecular modelling of structured organic polycrystalline powders," *CrystEngComm*, vol. 23, pp. 2520–2531, Apr. 2021.

[9] Y. Gao, T. Zhang, Y. Ma, F. Xue, Z. Gao, B. Hou, and J. Gong, "Application of PAT-Based Feedback Control Approaches in Pharmaceutical Crystallization," *Crystals*, vol. 11, p. 221, Mar. 2021.

[10] K. McBride and K. Sundmacher, "Overview of Surrogate Modeling in Chemical Process Engineering," *Chemie Ingenieur Technik*, vol. 91, no. 3, pp. 228–239, 2019.

[11] S. S. Haykin, *Neural Networks and Learning Machines*. Prentice Hall, 2009.

[12] K. Prantikos, L. H. Tsoukalas, and A. Heifetz, "Physics-Informed Neural Network Solution of Point Kinetics Equations for a Nuclear Reactor Digital Twin," *Energies*, vol. 15, p. 7697, Jan. 2022.

[13] H. Jiang, S. Wang, L. Xing, V. J. Pinfield, and J. Xuan, "Machine learning based techno-economic process optimisation for CO2 capture via enhanced weathering," *Energy and AI*, vol. 12, p. 100234, Apr. 2023.

[14] Z. K. Nagy, "Model based robust control approach for batch crystallization product design," *Computers & Chemical Engineering*, vol. 33, pp. 1685–1691, Oct. 2009.

[15] R. von Schomberg, "Prospects for technology assessment in a framework of responsible research and innovation," in *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden* (M. Dusseldorp and R. Beecroft, eds.), pp. 39–61, Wiesbaden: VS Verlag für Sozialwissenschaften, 2012.

[16] H. L. Needleman, C. Gunnoe, A. Leviton, R. Reed, H. Peresie, C. Maher, and P. Barrett, "Deficits in Psychologic and Classroom Performance of Children with Elevated Dentine Lead Levels," *New England Journal of Medicine*, vol. 300, pp. 689–695, Mar. 1979.

[17] M. J. Molina and F. S. Rowland, "Stratospheric sink for chlorofluoromethanes: Chlorine atom-catalysed destruction of ozone," *Nature*, vol. 249, pp. 810–812, June 1974.

[18] "Research integrity." https://www.ukri.org/what-we-offer/supporting-healthy-research-and-innovation-culture/research-integrity/, May 2023.

[19] "What research integrity is." https://ukcori.org/what-research-integrity-is/, May 2023.

[20] S. L. Titus, J. A. Wells, and L. J. Rhoades, "Repairing research integrity," *Nature*, vol. 453, pp. 980–982, June 2008.

[21] J. Mehlich, F. Moser, B. Van Tiggelen, L. Campanella, and H. Hopf, "The Ethical and Social Dimensions of Chemistry: Reflections, Considerations, and Clarifications," *Chemistry – A European Journal*, vol. 23, no. 6, pp. 1210–1218, 2017.

[22] "Our core principles." https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/framework-for-research-ethics/our-core-principles/, May 2023.

[23] J. Stilgoe, R. Owen, and P. Macnaghten, "Developing a framework for responsible innovation," *Research Policy*, vol. 42, pp. 1568–1580, Nov. 2013.

[24] "Framework for responsible research and innovation." https://www.ukri.org/about-us/epsrc/our-policies-and-standards/framework-for-responsible-innovation/, May 2023.

[25] M. Grieves, "Digital Twin: Manufacturing Excellence through Virtual Factory Replication," Mar. 2015.

[26] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," pp. 85–113, Aug. 2017.

[27] C. Perrow, *Normal Accidents: Living with High Risk Technologies - Updated Edition*. Princeton University Press, rev - revised ed., 1999.

[28] J. D. Seader, E. J. Henley, and D. K. Roper, *Separation Process Principles: Chemical and Biochemical Operations*. Hoboken, NJ: Wiley, 3rd ed ed., 2011.

[29] H. J. M. Kramer and G. M. van Rosmalen, "CRYSTALLIZATION," in *Encyclopedia of Separation Science* (I. D. Wilson, ed.), pp. 64–84, Oxford: Academic Press, Jan. 2000.

[30] R. Davey, *From Molecules to Crystallizers*. No. 86 in Oxford Chemistry Primers, Oxford: University Press, 2000.

[31] I. Sunagawa, *Crystals: Growth, Morphology, & Perfection*. Cambridge University Press, Aug. 2007.

[32] A. Chianese and H. J. M. Kramer, eds., *Industrial Crystallization Process Monitoring and Control*. Wiley, first ed., Mar. 2012.

[33] Z. Q. Yu, R. B. H. Tan, and P. S. Chow, "Effects of operating conditions on agglomeration and habit of paracetamol crystals in anti-solvent crystallization," *Journal of Crystal Growth*, vol. 279, pp. 477–488, June 2005.

[34] E. Simone, R. Othman, G. T. Vladisavljević, and Z. K. Nagy, "Preventing Crystal Agglomeration of Pharmaceutical Crystals Using Temperature Cycling and a Novel Membrane Crystallization Procedure for Seed Crystal Generation," *Pharmaceutics*, vol. 10, p. 17, Jan. 2018.

[35] Z. K. Nagy, J. W. Chew, M. Fujiwara, and R. D. Braatz, "Advances in the Modeling and Control of Batch Crystallizers," *IFAC Proceedings Volumes*, vol. 37, pp. 83–90, Jan. 2004.

[36] T. Rosenbaum, L. Tan, M. Dummeldinger, N. Mitchell, and J. Engstrom, "Population Balance Modeling To Predict Particle Size Distribution upon Scale-Up of a Combined Antisolvent and Cooling Crystallization of an Active Pharmaceutical Ingredient," *Organic Process Research & Development*, vol. 23, pp. 2666–2677, Dec. 2019.

[37] Å. C. Rasmuson, "Crystallization Process Analysis by Population Balance Modeling," in *Handbook of Industrial Crystallization* (A. Y. Lee, A. S. Myerson, and D. Erdemir, eds.), pp. 172–196, Cambridge: Cambridge University Press, third ed., 2019.

[38] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, and I. K. Zhitkov, "Towards crystal structure prediction of complex organic compounds – a report on the fifth blind test," *Acta Crystallographica Section B*, vol. 67, no. 6, pp. 535–551, 2011.

[39] G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. Van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. Van De Streek, A. K. Wolf, and B. Schweizer, "Significant progress in predicting the crystal structures of small organic molecules – a report on the fourth blind test," *Acta Crystallographica Section B*, vol. 65, no. 2, pp. 107–125, 2009.

[40] P. Barrett, B. Smith, J. Worlitschek, V. Bracken, B. O'Sullivan, and D. O'Grady, "A Review of the Use of Process Analytical Technology for the Understanding and Optimization of Production Batch Crystallization Processes," *Organic Process Research & Development*, vol. 9, pp. 348–355, May 2005.

[41] Z. K. Nagy, E. Aamir, and C. D. Rielly, "Internal Fines Removal Using Population Balance Model Based Control of Crystal Size Distribution under Dissolution, Growth and Nucleation Mechanisms," *Crystal Growth & Design*, vol. 11, pp. 2205–2219, June 2011.

[42] A. Saleemi, C. Rielly, and Z. K. Nagy, "Automated direct nucleation control for in situ dynamic fines removal in batch cooling crystallization," *CrystEngComm*, vol. 14, pp. 2196–2203, Feb. 2012.

[43] M. Fujiwara, Z. K. Nagy, J. W. Chew, and R. D. Braatz, "First-principles and direct design approaches for the control of pharmaceutical crystallization," *Journal of Process Control*, vol. 15, pp. 493–504, Aug. 2005.

[44] J. B. Rawlings, S. M. Miller, and W. R. Witkowski, "Model identification and control of solution crystallization processes: A review," *Industrial & Engineering Chemistry Research*, vol. 32, pp. 1275–1296, July 1993.

[45] B. Szilágyi, P. Ş. Agachi, and Z. K. Nagy, "Chord Length Distribution Based Modeling and Adaptive Model Predictive Control of Batch Crystallization Processes Using High Fidelity Full Population Balance Models," *Industrial & Engineering Chemistry Research*, vol. 57, pp. 3320–3332, Mar. 2018.

[46] "Integrated Synthesis, Crystallization, Filtration, and Drying of Active Pharmaceutical Ingredients: A Model-Based Digital Design Framework for Process Optimization and Control — SpringerLink." https://link.springer.com/chapter/10.1007/978-3-030-90924-6_10.

[47] B. Szilágyi, A. Eren, J. L. Quon, C. D. Papageorgiou, and Z. K. Nagy, "Digital Design of the Crystallization of an Active Pharmaceutical Ingredient Using a Population Balance Model with a Novel Size

Dependent Growth Rate Expression. From Development of a Digital Twin to In Silico Optimization and Experimental Validation," *Crystal Growth & Design*, vol. 22, pp. 497–512, Jan. 2022.

[48] I. Onaji, D. Tiwari, P. Soulatiantork, B. Song, and A. Tiwari, "Digital twin in manufacturing: Conceptual framework and case studies," *International Journal of Computer Integrated Manufacturing*, vol. 35, pp. 831–858, Aug. 2022.

[49] Z. Zhou, X. Li, and R. N. Zare, "Optimizing Chemical Reactions with Deep Reinforcement Learning," *ACS Central Science*, vol. 3, pp. 1337–1344, Dec. 2017.

[50] A. D. Randolph and M. A. Larson, "Transient and steady state size distributions in continuous mixed suspension crystallizers," *AIChE Journal*, vol. 8, no. 5, pp. 639–645, 1962.

[51] H. M. Hulburt and S. Katz, "Some problems in particle technology: A statistical mechanical formulation," *Chemical Engineering Science*, vol. 19, pp. 555–574, Aug. 1964.

[52] B. G. Palwe, M. R. Chivate, and N. S. Tavare, "Growth kinetics of potassium sulphate crystals in a DTB agitated crystallizer," *Chemical Engineering Science*, vol. 39, pp. 903–905, Jan. 1984.

[53] A. Pohar and B. Likozar, "Dissolution, Nucleation, Crystal Growth, Crystal Aggregation, and Particle Breakage of Amlodipine Salts: Modeling Crystallization Kinetics and Thermodynamic Equilibrium, Scale-up, and Optimization," *Industrial & Engineering Chemistry Research*, vol. 53, pp. 10762–10774, July 2014.

[54] N. Gherras and G. Fevotte, "On the use of process analytical technologies and population balance equations for the estimation of crystallization kinetics. A case study," *AIChE Journal*, vol. 58, no. 9, pp. 2650–2664, 2012.

[55] M. Sheikhzadeh, M. Trifkovic, and S. Rohani, "Adaptive MIMO neuro-fuzzy logic control of a seeded and an unseeded anti-solvent semi-batch crystallizer," *Chemical Engineering Science*, vol. 63, pp. 1261–1272, Mar. 2008.

[56] F. Boukouvala, F. J. Muzzio, and M. G. Ierapetritou, "Dynamic Data-Driven Modeling of Pharmaceutical Processes," *Industrial & Engineering Chemistry Research*, vol. 50, pp. 6743–6754, June 2011.

[57] J. J. Sikorski, G. Brownbridge, S. S. Garud, S. Mosbach, I. A. Karimi, and M. Kraft, "Parameterisation of a biodiesel plant process flow sheet model," *Computers & Chemical Engineering*, vol. 95, pp. 108–122, Dec. 2016.

[58] E. Sanchez, S. Pintos, and N. Queipo, "Toward an optimal ensemble of kernel-based approximations with engineering applications," in *IEEE International Conference on Neural Networks - Conference Proceedings*, pp. 2152–2158, 2006.

[59] H. C. Goel, N. Samanta, K. Kannan, I. P. Kumar, and M. Bala, "Protection of spermatogenesis in mice against gamma ray induced damage by Hippophae rhamnoides," *Andrologia*, vol. 38, no. 6, pp. 199–207, 2006.

[60] S. Ruder, "An overview of gradient descent optimization algorithms," June 2017.

[61] R. Dechter, "Learning While Searching in Constraint-Satisfaction-Problems.," in *AAAI*, pp. 178–185, Jan. 1986.

[62] I. N. Aizenberg, N. N. Aizenberg, and J. Vandewalle, *Multi-Valued and Universal Binary Neurons*. Boston, MA: Springer US, 2000.

[63] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, pp. 1527–1554, July 2006.

[64] G. E. Hinton, "Learning multiple layers of representation," *Trends in Cognitive Sciences*, vol. 11, pp. 428–434, Oct. 2007.

[65] M. Öner, F. C. C. Montes, T. Ståhlberg, S. M. Stocks, J. E. Bajtner, and G. Sin, "Comprehensive evaluation of a data driven control strategy: Experimental application to a pharmaceutical crystallization process," *Chemical Engineering Research and Design*, vol. 163, pp. 248–261, Nov. 2020.

[66] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, Feb. 2019.

[67] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, "Physics-informed neural networks (PINNs) for fluid mechanics: A review," *Acta Mechanica Sinica*, vol. 37, pp. 1727–1738, Dec. 2021.

[68] Y. Zheng and Z. Wu, "Physics-Informed Online Machine Learning and Predictive Control of Nonlinear Processes with Parameter Uncertainty," *Industrial & Engineering Chemistry Research*, vol. 62, pp. 2804–2818, Feb. 2023.

[69] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[70] D. Grechishnikova, "Transformer neural network for protein-specific de novo drug generation as a machine translation problem," *Scientific Reports*, vol. 11, p. 321, Jan. 2021.

[71] N. Geneva and N. Zabaras, "Transformers for modeling physical systems," *Neural Networks*, vol. 146, pp. 272–289, Feb. 2022.

[72] B. Srinivasan, S. Palanki, and D. Bonvin, "Dynamic optimization of batch processes: I. Characterization of the nominal solution," *Computers & Chemical Engineering*, vol. 27, pp. 1–26, Jan. 2003.

[73] D. L. Ma and R. D. Braatz, "Robust identification and control of batch processes," *Computers & Chemical Engineering*, vol. 27, pp. 1175–1184, Sept. 2003.

[74] Z. K. Nagy and R. D. Braatz, "Robust nonlinear model predictive control of batch processes," *AIChE Journal*, vol. 49, pp. 1776–1786, July 2003.

[75] B. Benyahia, P. Anandan, and C. Rielly, "Robust Model-Based Reinforcement Learning Control of a Batch Crystallization Process," in *2021 9th International Conference on Systems and Control (ICSC)*, pp. 89–94, Nov. 2021.

[76] M. E. Taylor and P. Stone, "Transfer Learning for Reinforcement Learning Domains: A Survey,"

[77] A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne, and A. A. Lapkin, "Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives," *Chemical Engineering Journal*, vol. 352, pp. 277–282, Nov. 2018.

[78] H.-J. Pan and J. D. Ward, "Optimization of Simple Batch Crystallization Systems Considering Crystal Shape and Nucleation," *Industrial & Engineering Chemistry Research*, vol. 59, pp. 9550–9561, May 2020.

[79] V. Manee, R. Baratti, and J. A. Romagnoli, "Learning to navigate a crystallization model with Deep Reinforcement Learning," *Chemical Engineering Research and Design*, vol. 178, pp. 111–123, Feb. 2022.

[80] E. Aamir, Z. Nagy, C. Rielly, T. Kleinert, and B. Judat, "Efficient crystal size distribution estimation approach for growth dominated crystallisation processes," Jan. 2008.

[81] G. X. Zhou, M. Fujiwara, X. Y. Woo, E. Rusli, H.-H. Tung, C. Starbuck, O. Davidson, Z. Ge, and R. D. Braatz, "Direct Design of Pharmaceutical Antisolvent Crystallization through Concentration Control," *Crystal Growth & Design*, vol. 6, pp. 892–898, Apr. 2006.

A ORBiT Methodology

This would be first appendix, which in this case is most likely to be on the ORBiT methodology from RRI.

B Training Plan

A section containing all excess details on the training plan and how it will be applied, what has been achieved, etc....