**ROYAL SOCIETY OF CHEMISTRY**

## PAPER

Check for updates

# AI-driven robotic crystal explorer for rapid polymorph identification

Edward C. Lee,† Daniel Salley,† Abhishek Sharma and Leroy Cronin ⓘ *

Crystallisation is central to purification and to determining structure and material properties, yet small changes in conditions can produce many different polymorphs with distinct behaviours. Because crystallisation depends on multiple variables including solvent, temperature, pressure, and atmosphere and often proceeds unpredictably, mapping these outcomes is slow and expensive. Here we introduce a robotic crystal search engine that explores crystallisation space efficiently and autonomously. The platform couples high-throughput liquid handling with a closed-loop computer-vision system combined with human supervision that uses machine learning to detect crystals, distinguish polymorphs, and identify previously unseen forms. Using a benchmark polymorphic compound, we show that the robot can rapidly navigate a high-dimensional solvent space, quantify relative polymorph yields directly from images, and build a phase diagram without recourse to crystallography. This approach reveals the full set of polymorphs accessible under given conditions and identifies the optimal conditions for producing each one.

## Introduction

Crystal polymorphism occurs when a compound can form multiple distinct crystal structures.[1] Polymorphs exhibit different physical,[2] spectroscopic,[3] surface,[4] mechanical[5] and chemical properties,[6] and as such, identification and reliable separation of different crystal polymorphs is vital in many fields.[7] Additionally, for active pharmaceutical ingredients, patenting laws protect the polymorph, not the molecule.[3] Polymorphism arises due to the effects of specific crystallisation conditions on inter- and intra-molecular interactions, such as hydrogen bonding, pi-stacking, and van der Waals forces, which affect the molecular orientation of and between nucleating molecules.[8–10] This gives rise to three possible features which may or may not be desired: concomitant crystallisation of different polymorphs, preferential crystallisation of polymorphs with sub-optimal properties, and spontaneous conversion to a more thermodynamically stable polymorph.[9]

For these reasons, a comprehensive knowledge of a compound's polymorphs and polymorph formation conditions is important for reproducibility, scaling, and yield optimisation. However, any exhaustive search to obtain all of a compound's polymorphs is both resource and time-expensive due to the high dimensionality of parameters that affect crystallisation that must be explored. Previously, there have been studies of crystallisation using automated platforms such as robotic handling of crystals such as sample scraping,[11] RAPID

platform for perovskites,[12] solid state chemistry including PXRD,[13] protein crystallisation,[14] exploring polymorphic landscape of organic cage.[15]

While high-throughput automation has helped, inefficient strategies such as Grid Search (GS)[16] are often applied with the aim of checking all conditions to a finite resolution, which is determined by available resources. GS is a sub-optimal strategy due to an excess of sampling points in regions where there is high outcome certainty, and a deficit of sampling points where outcome uncertainty is low, and this results in an inefficient use of any experimental budget. A better approach is to use an exploration/optimisation strategy combined with active learning, where sampling points are determined by a function operating on some measurable feature of the system being investigated, a method that has shown promise in many areas of chemistry.[17–19] However, applying this strategy to polymorph exploration presents two main challenges. Firstly, the only relevant observable feature is the relative yield of each polymorph, which is hard to quantify without expensive manual methods such as crystallography, and secondly, the response surfaces for polymorph yields are typically flat, making exploration and optimisation by yield alone impossible.

To resolve both of these problems, we have developed a method that uses an automated high-throughput closed-loop with human-in-the-loop control approach to automatically quantify relative polymorphic yields using computer vision, and optimise the search strategy to prioritise regions of crystallisation space with the highest uncertainty using Bayesian exploration, see Fig. 1. In addition to being able to quantify the amount of each known polymorph in a sample, this approach is able to identify the presence of any new polymorph discovered,

*School of Chemistry, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK. E-mail: Lee.Cronin@glasgow.ac.uk*

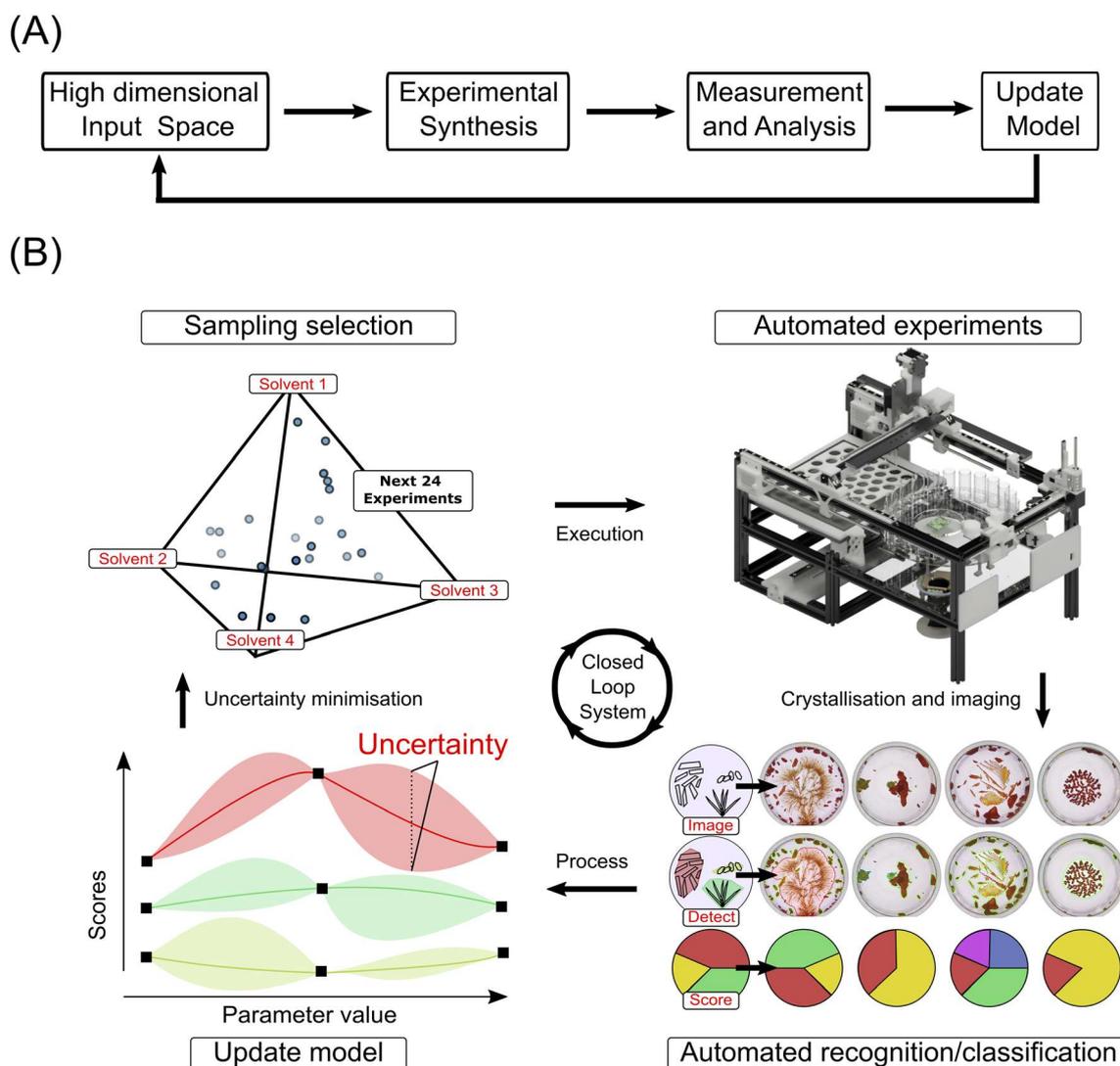† These authors contributed equally.

**Fig. 1** Polymorph exploration discovery loop. (A) A workflow of generalised closed-loop for the exploration of large combinatorial space. (B) Closed loop experimental platform together with human-in-the-loop supervision. Top left: crystallisation conditions are chosen from a parameter space initially at random and subsequently based on previous findings. Top right: the automated platform then prepares the crystallisation solutions as specified. Bottom right: experimental samples of crystallised material are imaged using a high-definition camera. Images are segmented into crystal/non-crystal using computer vision, as well as into different polymorph classes. New polymorphs are identified as the areas where the difference between the set of "crystal" pixels and the union of all "polymorph" pixels is greater than 0. Bottom left: these data are then used to create a Bayesian surrogate model from which the points with the highest uncertainty are selected in subsequent generations.

which may be characterised outside of the closed loop. Once characterised, the conditions required to form new discovery undergo Bayesian optimisation, whereby, the relative amount (as determined by computer vision) of the new polymorph is optimised. It is important to note that exploration and exploitation strategy is closed loop, however, once an unknown or novel polymorph is identified, it does require human intervention for offline XRD characterisation for phase confirmation and to introduce it in the algorithmic framework. Hence, the identification of new polymorph requires human-in-the-loop, but in principle, the crystallographic identification can be performed at later stage, and a new label can be assigned for a novel polymorph during a closed-loop process.

Here, we show how the combination of polymorph identification and quantification computer vision can be used together with a Bayesian exploration/optimisation strategy can efficiently explore a crystallisation space comprised of the relative amounts of four solvents for the compound 5-methyl-2-[(2-nitrophenyl)amino]thiophene-3-carbonitrile, build a phase diagram and discover the conditions to form a rare polymorph in the crystallisation space.

## Results and discussion

### Platform

The automated platform was inspired by previous designs[20] and consisted of a liquid handling and crystallisation robot with

a capacity of 72 parallel crystallisations from a total solvent volume of 2.5 ml, as well as automated imaging using a camera positioned below the crystallisation vials. The platform itself and its significant components/assemblies can be seen in Fig. 2 and details of its construction and operation are described in SI Section 2.

**Crystal-vision**

A computer-vision based library, named crystal-vision, was developed for automated detection, classification, and segmentation of *in situ* crystal images based on Mask RCNN[21] computer vision by loading a pre-trained image classification model and retraining the final layers on images of crystals. This enabled the creation of image analysers which could be used to infer the presence, type, location, and size of crystals in a sample vial. The detection capability was used to distinguish between crystal presence and absence, whereas classification was used to identify what subclass of crystal had been identified. The classifier was trained with two subclasses: morphology and identity. The morphology classifier could distinguish between amorphous/non-crystalline material, powder crystalline samples, overlapping crystals/crystal clusters, single/isolated crystals, and particular features of crystals such as tracht. The identity classifier could identify the compound of the crystal, provided that it had been previously incorporated into the classifier training database. A limitation to this method is the possibility of type I and type II errors. A type I error can quickly be rectified by performing X-ray analysis on the crystal, and the dataset being updated. However, a type II error would result in a new polymorph being missed. While this method cannot self-correct this type of error, it should be noted that this is also a problem faced by manually examining samples. Training details are provided in SI Section 3.1 and 3.2 together, these capabilities allowed the detection of crystallisation onset time, and crystalline quality of a specific compound or polymorph.

With the inclusion of image segmentation, the size and location of each crystal could be detected, and thus growth rates could be monitored for multiple single crystals simultaneously in one reaction vial. Alternatively, the system could be used to identify the presence of a previously unobserved crystal. This method involved using a classifier which had been trained on many types of crystals (classifier A) to recognise whether something is or is not a crystal. Another classifier which had been trained to recognise the specific identity of a crystal (classifier B) could then be applied to attain the identity of crystals in the sample. A positive response in a region from classifier A but a negative response in the same region from classifier B implies that the region corresponds to a crystal
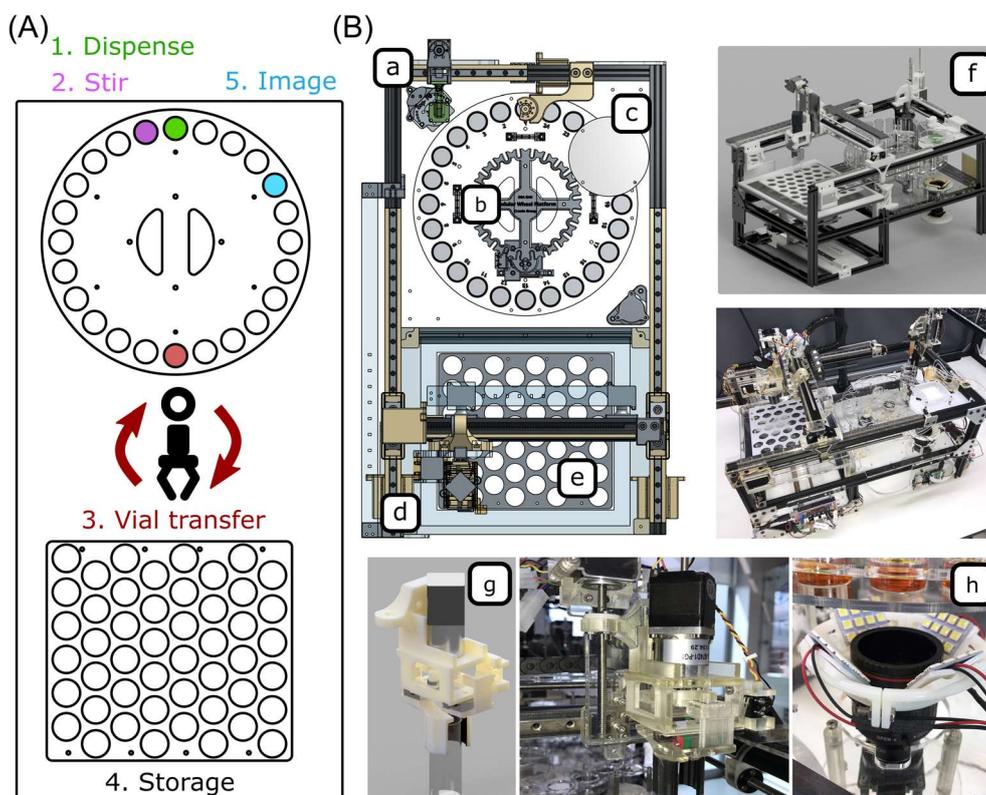


Fig. 2 Fully automated robotic platform. (A) Shows a brief overview of the significant steps the platform performs for each sample generation, including dispensing, stirring, vial transfer, storage, and imaging. (B) Experimental details of the platform including CAD design and built assembly. (a) Shows the position of the overhead stirring assembly, (b) 24 vial sample wheel, (c) location of the imaging setup, (d) X and Y-axis motors for the gripping assembly, (e) vial storage area, (f) complete platform as built (with a rendered CAD image), (g) rendered CAD and actual image of vial transfer gripper. (h) imaging set-up with C-mount raspberry Pi camera lens on a HQ raspberry Pi camera.
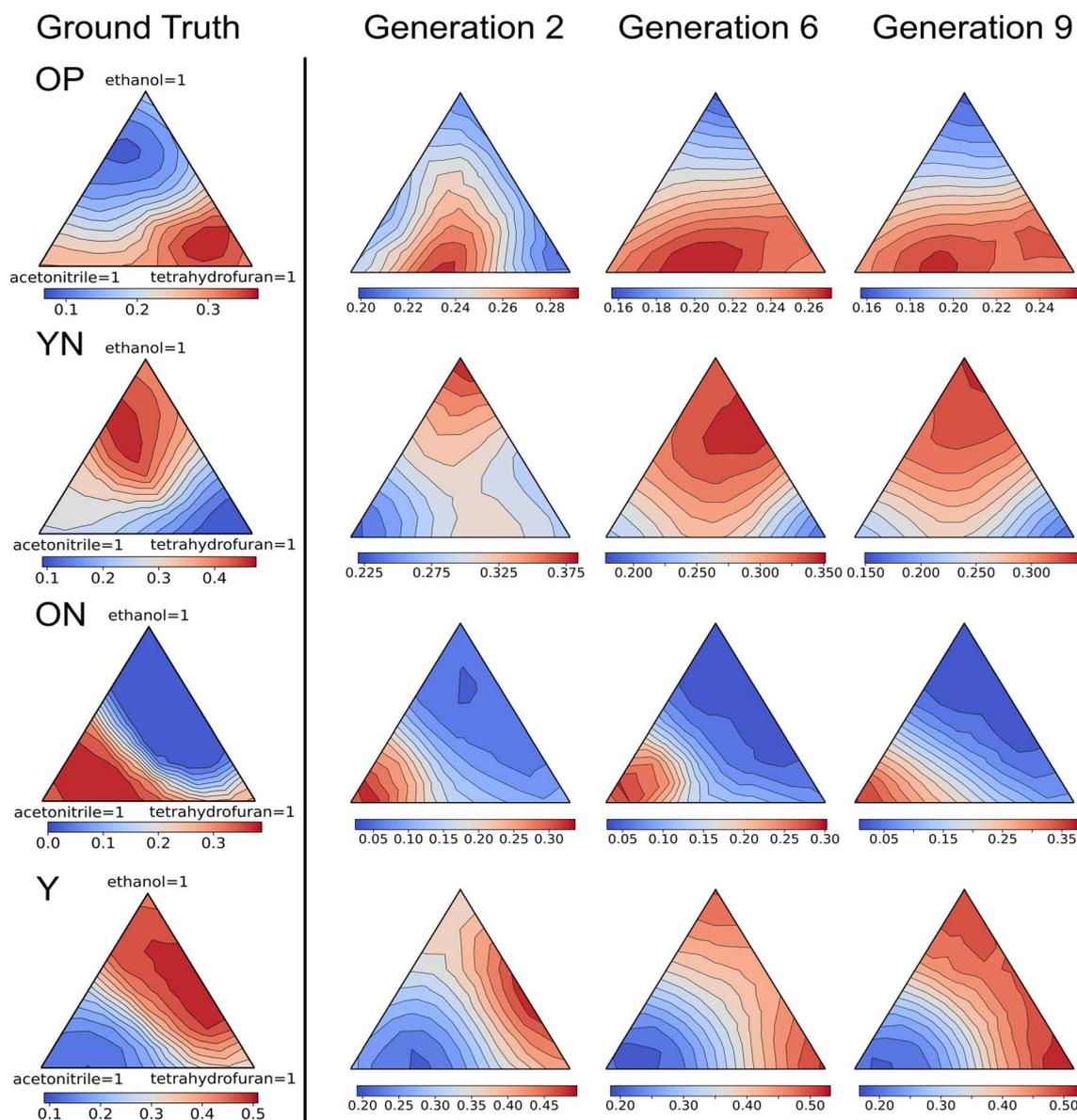
**Fig. 3** Simulation of polymorph crystallization experiment. (Left) Simulated ground truth likelihood heatmap for each polymorph in different solvent conditions (*e.g.* OP is much more likely to crystallize in THF than EtOH. MeCN is intermediate). (Right) Simulated likelihood heatmap for each polymorph under different solvent conditions across generations. Three generations are shown here, 2, 6, and 9. Over multiple generations, the likelihood approximates the ground truth for each polymorph shown in (left). For additional plots including scores, uncertainties, and error, see SI Section 3.

whose identity is unknown, and therefore potentially a new discovery. Initially, classifier B would suffer from overfitting due to the lack of data at the start of the experiment. However, this problem could be mitigated by incorporating more data as they were collected (see SI Section 3).

### Exploration

The strategy for polymorph search and discovery involved a Bayesian methodology, where data from previous experiments were used to construct a model of the polymorphic system from Gaussian processes. The investigated parameters were the proportions of each solvent in the crystallising system, and the

observed parameters were the proportions of each polymorph detected by the crystal-vision detector. In this way, separate surrogate models could be constructed for each polymorph to predict the likelihood of each polymorph's formation at any point, as well as the uncertainty (variance) associated with this. The aim of exploration was to obtain as accurate a model as possible across all investigated parameters in as few experiments as possible, increasing the likelihood that evaluation points which result in novel or rare outcomes are located faster than by methods such as grid search.

Specifically, crystallisation conditions over multiple generations were to maximise the negative integrated posterior
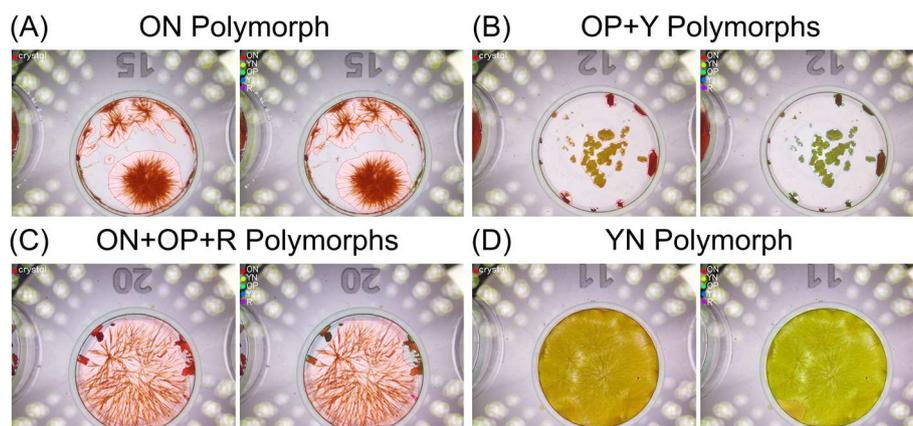
**Fig. 4** Crystal polymorphs of ROY detected, segmented, and labelled using mask RCNN image detection using two detectors. (1) Identified the presence of crystalline material and (2) identified the presence of each type of crystal from the set [ON, YN, OP, Y, R] times. (A) Detection of ON polymorph only, (B) detection of OP + Y polymorphs, (C) detection of ON + OP + R polymorphs, and (D) detection of YN polymorph only.

variance of the surrogate model. This was undertaken in an automated semi-closed loop process using an automated platform and an active learning search algorithm using the BoTorch package[22] (See Fig. 1). A set of simultaneous crystallisations were performed on the automated platform by mixing samples of the dissolved compound with other solvents at 25 °C and 1 atm pressure, then waiting 12 hours for slow evaporation and crystallisation to occur.

At this point, the vials were robotically transferred to an imaging platform, and the images were analysed using two classifiers created using the crystal-vision library to detect and identify known crystals and to alert the user to any unknown crystals, as described above. This process was simulated *in silico* (SI Section 3.3) with a random acquisition function being compared to NIPV.

Upon detection of an unknown crystal, the closed loop would be interrupted to allow for verification of the new polymorph by X-ray diffractometry (XRD). This introduces human-in-the-loop supervision for detailed analysis using XRD. On confirmation of a new polymorph, the producing conditions were repeated multiple times in order to obtain a larger training dataset of the polymorph, and the classifier was retrained incorporating these.

Once scores for each polymorph in each reaction had been calculated, they were incorporated into a Bayesian network and a surrogate model was built from the posterior probabilities using Gaussian processes. Since we were interested in prioritising exploration of uncertain regions in the reaction space, subsequent reaction conditions were then chosen to minimise the uncertainty in the surrogate models of each polymorph,
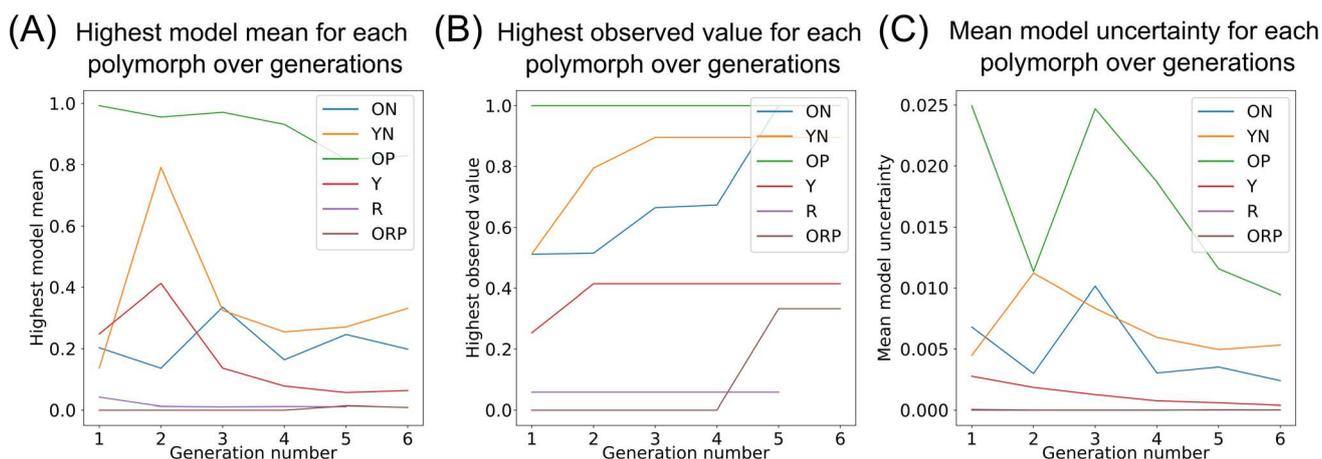


**Fig. 5** Model state from experimental observations from generations of crystallisation experiments. (A) The highest likelihood predicted by the model over each generation. A model mean value of 1 expresses that the model has found a location where it believes the likelihood of observing a particular polymorph is 100%. This can increase or decrease over generations as new observations update the model's predictions. These values are initially noisy, but are updated each generation and eventually converge as the model is exposed to more observations. (B) The maximum value obtained in a single experiment for each polymorph across each generation. (C) The model uncertainty for each polymorph which is updated each generation. Initially the results are noisy due to the probabilistic nature of crystallisation, where sampling of the same data point can result in a different outcome. However, over subsequent generations, the model uncertainty decreases for each polymorph until the uncertainties stabilise.
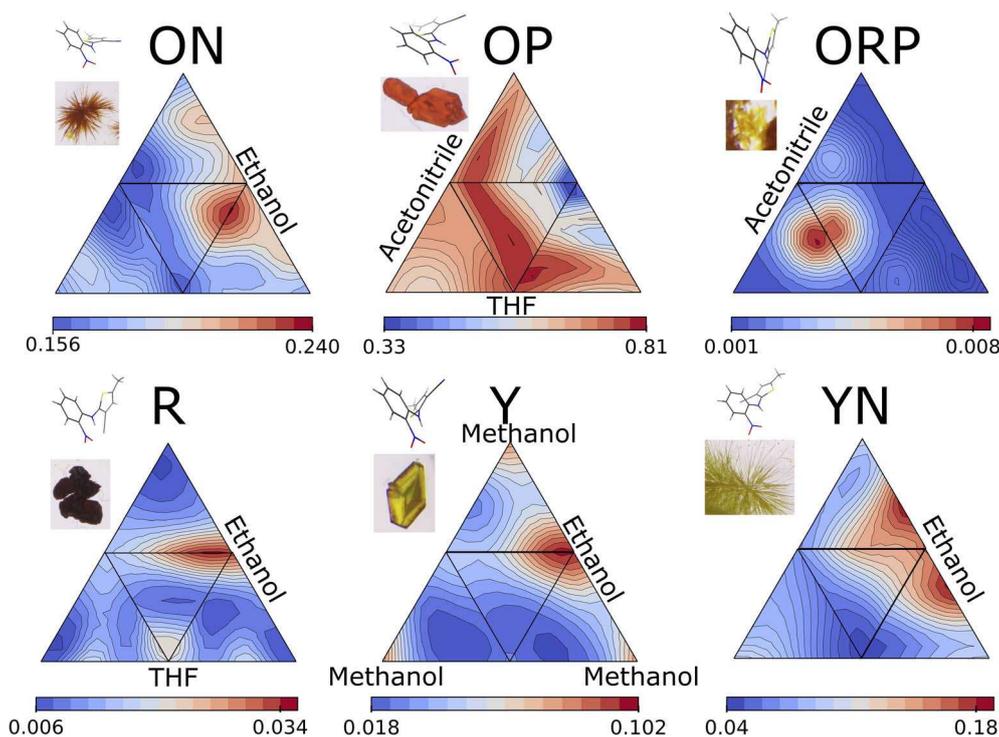
**Fig. 6** Generation 7 experiments and observations. Top left-stacked bar chart showing the ratios solvent composition for each reaction. Top right-compositional scatter plot of solvents used in each reaction. Vertices correspond to 100% of one solvent, opposite faces correspond to 0% of that solvent. Middle left-stacked bar chart showing the polymorph proportions observed for each reaction. Middle right-polymorph proportions represented as pie charts located appropriately in a compositional scatter plot of solvents. Bottom-Plots showing four ternary heatmaps of the surrogate model where the sum of 3 components = 1, and the fourth is 0. Each heatmap corresponds to one surface of the tetrahedron above, and the whole can be viewed as a flattening out of the tetrahedron.

specifically by applying an algorithm which would choose a reaction that would minimise negative integrated posterior variance (NIPV). This allowed an efficient reaction selection routine to maximise exploration away from known/expected outcomes in a system with high input and output dimensionality. Once solvent mixtures were selected, the automated platform was able to perform a subsequent generation of crystallisations and continue the loop.

The compound 5-methyl-2-[(2-nitrophenyl)amino] thiophene-3-carbonitrile (also known as ROY) is one of the most polymorphic compounds known (12 forms reported to date)[23] was selected as a potential candidate for the discovery of further polymorphic forms. It also has the property that several of its differently coloured polymorphs can form under similar conditions, resulting in concurrent polymorph crystallisation and intrinsically noisy data. As such, finding trends in high dimensional spaces, and therefore rational search and purification strategies is difficult. Because of this, performing a grid search in a large crystallisation space is an inefficient strategy, making this a good system on which to apply Bayesian exploration.

### Simulated experiments

Simulations (see methods) were performed in order to define expected behaviour of the system over successive reaction generations and determine what the consequence of searching in space with high dimensionality of both deterministic inputs and stochastic outputs. The simulations show that there is initially a large degree of uncertainty in the models, and likelihood is not consistent with the ground truth data. However, over successive generations the predicted likelihood closely resembles that of the ground truth data, indicating that the model had found the correct crystallisation parameters.

### Crystal detection

Every generation resulted in 72 images of crystals which needed to be classified and scored according to the relative amounts of each polymorph present in the sample. A general crystal classifier had been trained on data from the MARCO dataset, together with a set of samples of ROY crystals obtained from initial screening and was not retrained for the duration of the experiment. This could only classify one object, labelled as "crystal". For the identification and scoring of each polymorph, we began by assuming no knowledge *a priori* about polymorph appearance, abundance or identity. As such, images obtained in the first generation had to be manually labelled and the crystals undergo diffractometry where the polymorph identity after microscopic visual inspection was uncertain. Due to growth specific conditions, visually dissimilar crystals were occasionally found after XRD analysis to be the same polymorph, visually similar could be mislabelled as another polymorph, (*e.g.* some instances of R and OP). However, incorporating these disparate

samples into the training data allowed the model to generalise and correctly classify similar samples, see Fig. 4.

The first generation was trained on 50 images of the polymorphs OP, ON, R, Y and YN. The classifier was retrained after generation 2 on a further 30 images in order to increase its generalisability and after generation 5 when polymorph ORP was confirmed by XRD. To assign a score for each polymorph in a reaction, the fractional area of each polymorph was calculated from the size of its pixel mask.

### Physical experiments

In the experiments, 7 generations of 24 triplicate crystallisations were performed. Each consisted of a 1 ml solution of ROY in acetone (3.85 mmol L$^{-1}$) being mixed with a 1.5 ml mixture of four other solvents (a: ethanol, b: methanol, c: acetonitrile, and d: tetrahydrofuran) in different ratios, where $a + b + c + d = 1$. The values of each of the solvent variables were chosen at random for the first generation from a Dirichlet distribution. The image classifiers were those as described in the section on crystal detection, and the generation acquisition function used those as described in the simulation section, except that 4 solvent dimensions were investigated, and the simulated scores for each polymorph were replaced by the scores obtained from the image detection procedure outlined above.

Initially there was a large uncertainty in the models for each polymorph due to the large volume of unexplored space, but this decreased over successive generations. The veracity of the model can be checked when comparing the expected error between evaluated points and the model prediction at each of those points as seen in Fig. 5a. However, the mean error difference between subsequent generations decreased over time, indicating that the model was converging on an overall solution. Fig. 5b shows the largest abundance of each polymorph detected across the entirety of all generation, which can be used as an expected baseline yield for optimisation experiments: a region with 100% polymorph does not need optimised. Fig. 5c shows the uncertainty of each prediction reported by the model. This trends down over time due to successive additional data points. The data is noisy due to the probabilistic nature of crystallisation, which may cause in an increase in uncertainty if the initial dataset of low chance observations occur in early generations.

This process was repeated for 7 generations, and the final surrogate models are shown in Fig. 6. The first 5 generations saw only the first 5 initial polymorphs (ON, YN, OP, Y and R), however, in generation 6, a sample was classified as containing crystals but not as containing any known polymorph. The crystals in the sample were identified by X-ray crystallography as another polymorphic form of ROY, ORP, and the polymorph classifier was retrained to incorporate this, with the inclusion of an additional class. One crystallisation generation could be performed per day, accounting for manual instantiation, automated addition, evaporation and clean-up processes.

However, since the number of instances of this polymorph was only one, a generation was then created to confirm and establish the conditions required to produce ORP, as well as to gain additional training data for image detection. This was done by optimising for increased Expected Improvement (EI) of ORP over all the input parameters instead of decreased NIPV. Each of the Bayesian models was then updated to accommodate the new polymorph and the experiment was continued.

## Conclusions

We have shown that by using a combination of robotic automation, computer vision and artificial intelligence it is possible to run open-ended search and exploration loops for new polymorph forms in a complex polymorphic system. Our system was able to perform fully automated crystallisation cycles of a solvent space in triplicate to produce high-resolution images of multiple polymorphs of 5-methyl-2-[(2-nitrophenyl)amino] thiophene-3-carbonitrile (ROY). By employing two image classifiers trained on different features, first to determine the presence of crystals and second to classify the crystal form, we have shown that by using computer vision, different polymorphs can be automatically distinguished, with minimal human intervention. We have been able to discover the crystallisation conditions that lead to the rapid localisation and formation of a polymorph that had not previously been reported when using this combination of crystallisation solvents. As such, we believe this method is a viable strategy for the discovery of novel polymorphs for less well-explored compounds.

## Methods

### Chemical reagents

All solvents used for crystallisation solutions and cleaning were HPLC grade from Sigma-Aldrich. 5-Methyl-2-[(2-nitrophenyl) amino]thiophene-3-carbonitrile (>97%) (ROY) was obtained from Tokyo Chemical Industry Ltd.

### Platform

The platform was constructed in-house from a range of 3D printed, laser-cut and commercially available components. Further details of the platform and an overview of the subassemblies can be seen in SI Section 2, and a STEP file of a complete experimental setup can be found on GitHub. The software control of the platform for basic operations was written in Python 3. The software for image analysis was written in Python using Detectron2 and OpenCV packages.

### Simulations

A system with 4 possible polymorphs was initialised, where the likelihood of a single crystallisation event of a particular polymorph at a particular time point in an experiment followed a Poisson distribution (eqn (1)), which is typical of crystallising systems:[24]

$$P_{(x;\lambda)} = \frac{\lambda^x e^{-\lambda}}{x!} \tag{1}$$

where $x = 1$ for a single crystallisation, and $\lambda$ is the polymorph crystallisation rate, which is determined by eqn (2)

$$\lambda = Ae^{-\Delta G*/kT} \quad (2)$$

where $A$ is a concentration independent pre-exponential factor based on molecular attachment rates and $\Delta G*$ is the (concentration-dependent) nucleation free energy barrier given by

$$\Delta G* = K\frac{\gamma^3 \nu_m{}^2}{\ln^2(\sigma/\sigma*)} \quad (3)$$

where $K$ is a constant to represent temperature and the Boltzmann constant, $\gamma$ is the polymorph interfacial energy, $\nu_m$ is the polymorph molar volume, $\sigma*$ is the polymorph solubility, and $\sigma$ is the concentration of the solution phase, which is determined by the number of moles, M, in solution divided by the solution volume $\nu$. The polymorph solubility at a point, $\sigma*$ was simulated as the magnitude of a probability function of a Gaussian distribution, with randomly assigned mean and covariance, at that point. From this, a heatmap can be created of the phase likelihood at each point, $(x)$ for each polymorph $p$ by comparing the nucleation free energy barrier of this phase *versus* other polymorph phases:

$$P(p,x) \propto \frac{\Delta G_p^*}{\sum_k \Delta G_k^*} \quad (4)$$

Simulated crystallisation conditions were generated using the same method as physical experiments, *i.e.* an initial set of solvents whose proportions were chosen at random from by a 3-dimensional Dirichlet distribution, as shown in Fig. 3. Each simulated crystallisation experiment was initialised with a volume, $\nu_0$ and solution moles, $M_0$, and then proceeded in a series of timesteps, where at every point the volume of solvent was reduced by a fixed amount to simulate evaporation. This caused an increase in the solution concentration, which altered the crystallisation probability, $P_{(x,\lambda)}$, of each polymorph at that timestep and ratio of solvents. A crystallisation event for a particular polymorph was then determined by comparison of the crystallisation likelihood with a pseudorandom number between 0 and 1 taken from a uniform distribution. On a successful crystallisation event, a new polymorph crystal was initialised with a number of moles of compound (dependent on polymorph), $M_c$ and this value was subtracted from the number of moles of compound remaining in the solution phase. In each subsequent timestep, every crystal then grew in number of moles based on the equation:

$$M_c(t + 1) = M_c(t)(1 + K_p c(t)) \quad (5)$$

$M_c(t)$ is the number of moles of a crystal at timestep $t$, $K_p$ is the growth constant for polymorph $p$, and $c(t)$ is the solution concentration at timestep $t$.

The total number of timesteps was set so that the final solution volume would be 0, and the crystallisation parameters of each polymorph were such that complete conversion of

solution compound moles to polymorph moles would be achieved in every experiment. These final values could then be compared against the theoretical proportions of each polymorph, which is determined by the ratio of initial likelihoods of each polymorph at timestep $t = 0$. Four methods of generation instantiation were simulated and compared: random, uncertainty minimisation, and estimated improvement. 10 generations of 24 triplicate simulation crystallisations were performed. It should noted that while in this system solvent evaporation is constant for each solvent, a physical system would exhibit variable evaporation rates of each solvent.

## Author contributions

L. C. conceived the idea and designed the project together with E. L., D. S, and A. S. D. S. designed and implemented the hardware of the platform and performed the experiments. G. K. and D. S. wrote the control software of the platform with help from AS. E. L performed the simulations with help from AS. E. L implemented the crystal vision software as well as the algorithms for uncertainty minimalisation with the assistance from A. S. A. S. coordinated the project together with L. C. L. C., E. L., A. S, D. S. wrote the paper.

## Conflicts of interest

The authors declare no conflict of interest.

## Data availability

The characteristic datasets supporting this article will be uploaded to the Zenodo repository https://doi.org/10.5281/zenodo.15531902. Due to the large volume, the complete dataset generated and used in this work are available upon request to the corresponding author *via* Lee.Cronin@glasgow.ac.uk. All the codes used in this work including hardware control, experimental data analysis, and machine learning models are available at https://github.com/croningp/polymorph.

Supplementary information (SI) is available. See DOI: https://doi.org/10.1039/d5dd00203f.

## Acknowledgements

# References

1 D. Braga, F. Grepioni, L. Maini and M. Polito, in *Molecular Networks*, ed. M. W. Hosseini, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, vol. 132, pp. 87–95.

2 T. O. Ehiwe, B. D. Alexander, J. C. Mitchell, M. J. Snowden and L. J. Waters, Monitoring real time polymorphic transformation of sulfanilamide by diffuse reflectance visible spectroscopy, *J. Pharm. Anal.*, 2016, **6**, 179–183.

3 P. Láng, V. Kiss, R. Ambrus, G. Farkas, P. Szabó-Révész, Z. Aigner and E. Várkonyi, Polymorph screening of an active material, *J. Pharm. Biomed. Anal.*, 2013, **84**, 177–183.

4 H. H. Y. Tong, B. Yu. Shekunov, P. York and A. H. L. Chow, Influence of Polymorphism on the Surface Energetics of Salmeterol Xinafoate Crystallized from Supercritical Fluids, *Pharm. Res.*, 2002, **19**, 640–648.

5 M. Maździarz and T. Mościcki, New Zirconium Diboride Polymorphs—First-Principles Calculations, *Materials*, 2020, **13**, 3022.

6 A. Jeindl, L. Hörmann and O. T. Hofmann, How much does surface polymorphism influence the work function of organic/metal interfaces?, *Appl. Surf. Sci.*, 2022, **575**, 151687.

7 J. Bernstein, *Polymorphism in molecular crystals*, Oxford University Press, Oxford, New York, 2nd edn, 2020.

8 J. Anwar and D. Zahn, Polymorphic phase transitions: Macroscopic theory and molecular simulation, *Adv. Drug Delivery Rev.*, 2017, **117**, 47–70.

9 D. Bučar, R. W. Lancaster and J. Bernstein, Disappearing Polymorphs Revisited, *Angew. Chem., Int. Ed.*, 2015, **54**, 6972–6993.

10 J. Weng, Y. Huang, D. Hao and Y. Ji, Recent advances of pharmaceutical crystallization theories, *Chin. J. Chem. Eng.*, 2020, **28**, 935–948.

11 G. Pizzuto, H. Wang, H. Fakhruldeen, B. Peng, K. S. Luck and A. I. Cooper, in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, IEEE, Bari, Italy, 2024, pp. 2103–2110.

12 Z. Li, M. A. Najeeb, L. Alves, A. Z. Sherman, V. Shekar, P. Cruz Parrilla, I. M. Pendleton, W. Wang, P. W. Nega, M. Zeller, J. Schrier, A. J. Norquist and E. M. Chan, Robot-Accelerated Perovskite Investigation and Discovery, *Chem. Mater.*, 2020, **32**, 5650–5663.

13 A. M. Lunt, H. Fakhruldeen, G. Pizzuto, L. Longley, A. White, N. Rankin, R. Clowes, B. Alston, L. Gigli, G. M. Day, A. I. Cooper and S. Y. Chong, Modular, multi-robot

integration of laboratories: an autonomous workflow for solid-state chemistry, *Chem. Sci.*, 2024, **15**, 2456–2463.

14 N. E. Chayen, P. D. Shaw Stewart, D. L. Maeder and D. M. Blow, An automated system for micro-batch protein crystallization and screening, *J. Appl. Crystallogr.*, 1990, **23**, 297–302.

15 C. E. Shields, T. Fellowes, A. G. Slater, A. I. Cooper, K. G. Andrews and F. T. Szczypiński, Exploration of the polymorphic solid-state landscape of an amide-linked organic cage using computation and automation, *Chem. Commun.*, 2024, **60**, 6023–6026.

16 A. E. Gongora, B. Xu, W. Perry, C. Okoye, P. Riley, K. G. Reyes, E. F. Morgan and K. A. Brown, A Bayesian experimental autonomous researcher for mechanical design, *Sci. Adv.*, 2020, **6**, eaaz1708.

17 Y. Jiang, D. Salley, A. Sharma, G. Keenan, M. Mullin and L. Cronin, An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials, *Sci. Adv.*, 2022, **8**, eabo2626.

18 D. Salley, G. Keenan, J. Grizou, A. Sharma, S. Martín and L. Cronin, A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles, *Nat. Commun.*, 2020, **11**, 2771.

19 J. M. Granda, L. Donina, V. Dragone, D.-L. Long and L. Cronin, Controlling an organic synthesis robot with machine learning to search for new reactivity, *Nature*, 2018, **559**, 377–381.

20 D. S. Salley, G. A. Keenan, D.-L. Long, N. L. Bell and L. Cronin, A Modular Programmable Inorganic Cluster Discovery Robot for the Discovery and Synthesis of Polyoxometalates, *ACS Cent. Sci.*, 2020, **6**, 1587–1593.

21 K. He, G. Gkioxari, P. Dollar and R. Girshick, in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, 2017, pp. 2980–2988.

22 M. Balandat, B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson and E. Bakshy, in *Advances in Neural Information Processing Systems*, ed. H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, Curran Associates, Inc., 2020, vol. 33, pp. 21524–21538.

23 X. Li, X. Ou, H. Rong, S. Huang, J. Nyman, L. Yu and M. Lu, The Twelfth Solved Structure of ROY: Single Crystals of Y04 Grown from Melt Microdroplets, *Cryst. Growth Des.*, 2020, **20**, 7093–7097.

24 G. M. Maggioni and M. Mazzotti, Modelling the stochastic behaviour of primary nucleation, *Faraday Discuss.*, 2015, **179**, 359–382.