

Rapid Data-Efficient Optimization of Perovskite Nanocrystal Syntheses through Machine Learning Algorithm Fusion

Carola Lampe, Ioannis Kouroudis, Milan Harth, Stefan Martin, Alessio Gagliardi, and Alexander S. Urban*

With the demand for renewable energy and efficient devices rapidly increasing, a need arises to find and optimize novel (nano)materials. With sheer limitless possibilities for material combinations and synthetic procedures, obtaining novel, highly functional materials has been a tedious trial and error process. Recently, machine learning has emerged as a powerful tool to help optimize syntheses; however, most approaches require a substantial amount of input data, limiting their pertinence. Here, three well-known machine-learning models are merged with Bayesian optimization into one to optimize the synthesis of CsPbBr₃ nanoplatelets with limited data demand. The algorithm can accurately predict the photoluminescence emission maxima of nanoplatelet dispersions using only the three precursor ratios as input parameters. This allows us to fabricate previously unobtainable seven and eight monolayer-thick nanoplatelets. Moreover, the algorithm dramatically improves the homogeneity of 2–6-monolayer-thick nanoplatelet dispersions, as evidenced by narrower and more symmetric photoluminescence spectra. Decisively, only 200 total syntheses are required to achieve this vast improvement, highlighting how rapidly material properties can be optimized. The algorithm is highly versatile and can incorporate additional synthetic parameters. Accordingly, it is readily applicable to other less-explored nanocrystal syntheses and can help rapidly identify and improve exciting compositions' quality.

1. Introduction


Halide perovskite nanocrystals (PNCs), first demonstrated in 2014, have been rapidly improved, yielding tunability throughout the visible spectrum, quantum yields approaching 100%, and diverse geometries and sizes.^[1] Due to their exceptional properties, PNCs have already been incorporated into diverse applications, focusing on optoelectronics such as LEDs, solar cells, and photodetectors, but also in field-effect transistors and, even more recently, photocatalysis.^[2–6] Despite these impressive improvements, several issues impede widespread commercialization, such as stability, lead toxicity, and spectral efficiency in the blue region of the visible spectrum.^[7–9] This latter effect is due to the chloride-perovskites being far from defect tolerant, resulting in extremely poor efficiencies compared to bromide- and iodide-based perovskites.^[10] Another way to tune the spectral response in PNCs is through quantum confinement. Especially, 2D nanoplatelets (NPLs) are ideal in this regard, as they exhibit no inhomogeneous broadening in the confined dimension,

with only incremental thickness values possible—currently between two and six monolayers (MLs).^[11] Analogous to the bulk-like Ruddlesden–Popper perovskites,^[12] their strong confinement can enable directional emission, boosting maximum theoretical external quantum efficiencies to 28%.^[13] The quality of these colloidal quantum wells has improved significantly; however, their photoluminescence quantum yields are still far from unity, and reproducibility is an issue. Improving the NPL quality or that of any nanocrystal (NC) dispersion is an arduous task, involving a vast possible number of parameters relating to composition and fabrication. Synthesizing all of these is both infeasible and unnecessary, as it is possible to create robust and data-efficient predictors to describe the outcome of changes in fabrication parameters.^[14]

The process of synthesis optimization, however is a daunting task if undertaken purely with experimental intuition. To this end, it is imperative to incorporate artificial intelligence into the pipeline. Artificial intelligence and machine learning have

C. Lampe, S. Martin, A. S. Urban
Nanospectroscopy Group and Center for NanoScience
Nano-Institute Munich
Faculty of Physics
Ludwig-Maximilians-Universität München
80539 Munich, Germany
E-mail: urban@lmu.de

I. Kouroudis, M. Harth, A. Gagliardi
Department of Electrical and Computer Engineering
Technical University of Munich
Hans-Piloty-Straße 1, 85748 Garching bei München, Germany

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.202208772>.

© 2023 The Authors. Advanced Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

DOI: 10.1002/adma.202208772

become substantially more prominent in the field of materials science, both for discovering new material combinations and also for optimizing their fabrication.^[15,16] Various approaches have emerged, such as artificial neural networks^[17,18] or random forests and support vector machines.^[19,20] However, no algorithm is universally superior, with certain ones more suited for specific applications.^[21] Moreover, combining different algorithms can merge their individual capabilities. One such example involves Gaussian processes (GPs) and Bayesian optimization. GPs, with their inbuilt uncertainty quantification functionalities, are especially attractive given their data efficiency and robustness against overfitting.^[22–24] Even if the predictor is robust, however, the optimization algorithm must be of equal strength for a viable pipeline to be constructed. Bayesian optimization has shown its value in determining the optimal values of predictors, which elucidate the effect that different fabrication and chemical parameters have on the resulting material properties. This is accomplished by balancing the exploration of new areas in input parameter space against the

exploitation of already acquired information.^[16,25] However, this has required a massive experimental effort to achieve impressive results, limiting the applicability of this method for rapid material discovery.

In this study, we develop an algorithm to rapidly optimize material synthesis as illustrated in **Figure 1**. We demonstrate the capabilities of the algorithm, comprising Gaussian processing with a neural network and a random forest classifier, by applying it to a synthesis of 2D CsPbBr₃-based NPLs with which we already have significant experience. The algorithm uses three precursor amounts to predict the emission wavelength of the resulting NPLs and the quality, that is, the homogeneity of the photoluminescence (PL) spectra. Starting from a pool of 100 initial syntheses, we carried out seven rounds of optimization. The algorithm produced 14 new precursor combinations, which were then used for synthesis and the PL of the resulting dispersions measured. For all previously synthesized NPL thicknesses (2–6 ML), we significantly reduced the width and asymmetry of the PL emission, signifying higher

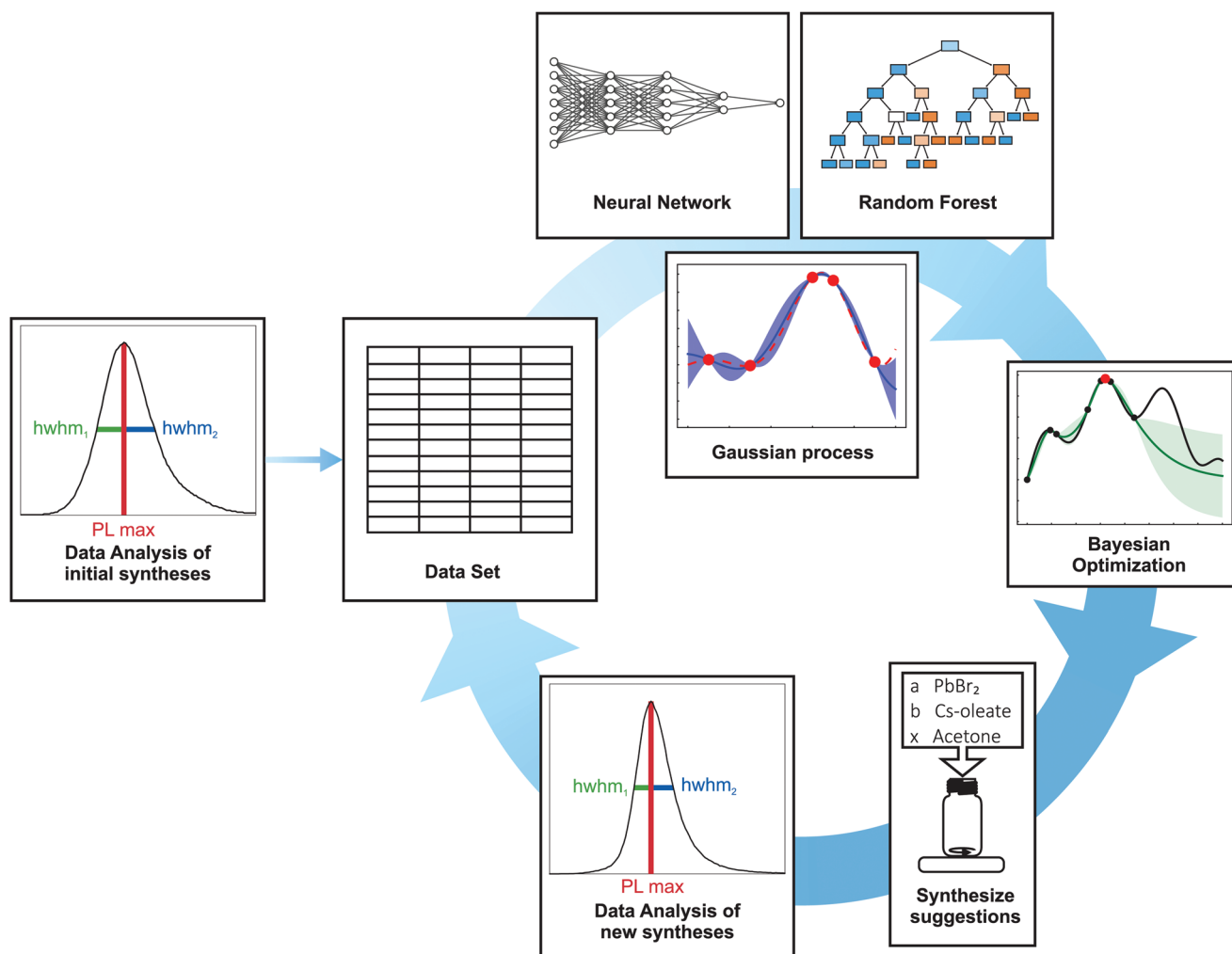


Figure 1. Scheme of the optimization process: Initially existing data points (syntheses) were analyzed and used to predict a spectral figure of merit (FoM) based on the narrowness and symmetry of their PL spectra using Gaussian processes in combination with a random forest and a neural network. Constrained Bayesian optimization subsequently leverages the information of the artificial intelligence step to provide suggestions for new compositions. For each cycle, 14 new syntheses were carried out and characterized. The amended dataset is subjected to a new optimization cycle.

homogeneity and thus improved emission quality. Additionally, the algorithm effectively predicted precursor combinations leading to hitherto unobtained, thicker NPLs (seven and eight MLs). The algorithm's performance was strong, especially considering the small amount of necessary experimental synthesizing. Importantly, it can easily be adapted to other less-known syntheses and incorporate additional synthetic input parameters, such as temperature and humidity.

2. Results and Discussion

2.1. Synthesis and Optical Characterization

In contrast to typical inorganic semiconductor NC syntheses and halide perovskite quantum dots, which rely on the hot-injection method, our synthesis is based on the ligand-assisted reprecipitation (LARP) method.^[2,11,26] Importantly, it is conducted at room temperature in ambient atmospheres.^[11,27,28] Briefly, a cesium-oleate precursor is injected into a PbBr₂-ligand (comprising oleylamine and oleic acid) solution. After ≈10 s, acetone, which acts as an antisolvent, is injected into the solution to induce NPL formation. After 60 s of vigorous stirring, the reaction is terminated by centrifugation at 1800 g for 3 min. The supernatant is discarded, and the precipitate is redispersed in hexane. Our previous studies determined that NPLs form when the amount of the A-site cation, in this case, cesium, is restricted.^[27] By fine-tuning the volumes of the cesium and lead-precursors and the anti-solvent, acetone, we were able to obtain nearly homogeneous dispersions of NPLs from two to six MLs.^[11] Accordingly, we chose these three parameters as the input parameters for the learning algorithm while keeping all others, for example, synthesis time, temperature, and humidity, essentially constant. To quantify the quality of the synthesis, we acquired and analyzed PL spectra of the resulting dispersions. We chose not to focus on the PL intensity but rather on the narrowness and the spectral position of the PL emission. Accordingly, we determined the narrowness f_{nar} by the usual half width at half maximum (HWHM) metric,^[29] while the symmetry, f_{sym}

was given by interpreting the spectrum as a distribution and calculating its skewness.^[30] As these two metrics have different ranges, they were normalized separately with the min–max method between zero and one and subsequently added. A more detailed derivation can be found in Supporting Information.

2.2. Designing the Algorithm

Accordingly, small values of f_{nar} and f_{sym} indicate good samples with a resulting FoM of one being a perfect spectrum and zero a very poor one. Examples of poor spectra are shown in Figure 2a,b, and the best spectrum according to the metric obtained in Figure 2c. The goal of the machine learning-powered process is thus the maximization of the FoM. Several aspects must be considered when developing the full framework for guiding the NPL synthesis. Starting from a limited number of syntheses (100), we have to be sure to prevent overfitting. Therefore, to predict which precursor ratios yield the highest FoM values, we adopted a Gaussian process predictor using the python package GPy.^[31] This choice was made because Gaussian processes are generally more robust than others with a limited number of data. Our goal was to obtain ideal spectra for a given NPL thickness; hence, we needed a way to determine how many MLs the resulting NPLs would have. Due to the strong confinement of the NPLs, there is a robust correlation between thickness and PL emission wavelength.^[11] We implemented a neural network through the tensorflow library^[32] to predict the PL peak position based on the precursor ratios, starting from the pool of initial syntheses. We incorporated this into the overall FoM optimization pipeline as soft or hard Lagrange multipliers, constraining the PL to a given spectral window (see Table 1). In the soft case, we only incentivize and do not fully constrain PL emission near the mean of the spectral range. In contrast, in the hard case, the peak position is forced to fall on the mean, under the rationale that the suggestions will be least affected by the peak position prediction error. The hard constraints gave the best results in the initial stages of the optimization, when the peak position was still inaccurate,

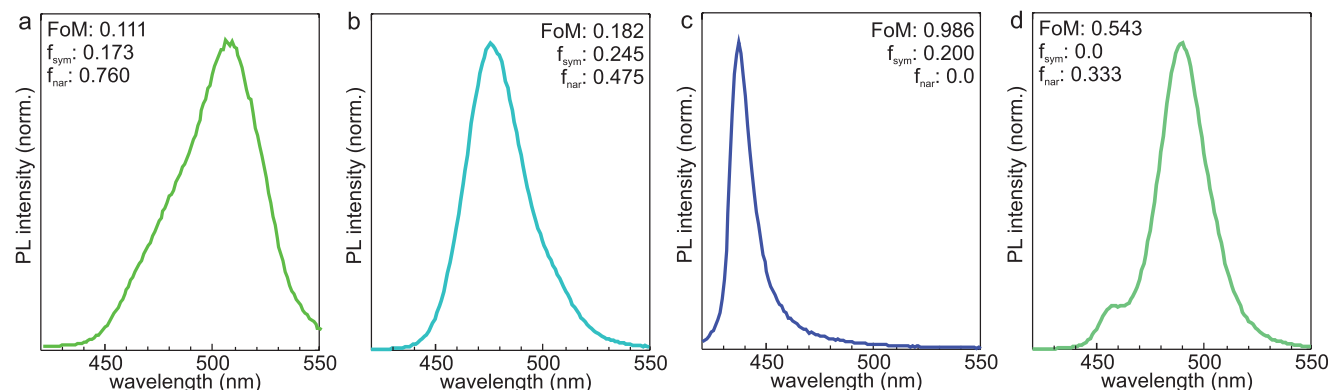


Figure 2. Examples of synthesized spectra with the classification of their merit according to narrowness (f_{nar}), symmetry (f_{sym}), and overall quality (FoM). a,b) Examples of poor spectra with (a) being extremely broad and (b) being quite asymmetric. c) PL spectrum from the synthesis with the highest obtained FoM value. An extremely narrow spectrum compensates for the slight asymmetry. d) The FoM calculation classifies this PL spectrum as very good due to a near-perfect symmetry. However, the short wavelength shoulder is a clear indicator of multiple thicknesses. This feature is predicted and preempted with the random forest classifier.

Table 1. NPL thickness and constraints for upper and lower limits.

NPL thickness [ML]	2	3	4	5	6	7	8
Lower limit [nm]	427	455	472	484	491	499	504
Upper limit [nm]	438	467	479	489	498	503	507

with the soft constraints giving better flexibility and therefore results when the peak position could be more accurately predicted. More details are given in the Supporting Information.

In some syntheses, the PL spectrum revealed secondary peaks, which were not identified through the FWHM method, yet signify polydispersity, see Figure 2d. For this synthesis, the FoM is relatively high at 0.543; however, the shoulder at 455 nm is a clear sign of an inhomogeneous NPL population. To increase the data efficiency of the optimization pipeline, we trained a random forest classifier with the aid of the library scikit-learn to identify the conditions leading to polydisperse NPL distributions (see Experimental Section for details). Importantly, while the classifier could identify roughly half of the combinations resulting in multiple peaks, it only erroneously excluded less than 5% of the combinations yielding single peaks (see Supporting Information, for more details). Consequently, viable compositions are hardly prevented from being evaluated while large areas of infeasible compositional space are cordoned off, significantly improving the algorithm's efficiency. The final set of considerations results from heuristic laboratory rules garnered from experience. Among these are minimum precursor volumes, the condition that the concentration of the Cs-precursor must be larger than the lead precursor to ensure NPL formation, and that the acetone volume is at least 30% of the total precursor volume to ensure the function as an antisolvent.

These individual contributions were merged into a Bayesian optimization framework to deliver suggestions for promising precursor ratios for new syntheses. This approach generates suggestions focusing initially on points with high uncertainty and gradually shifting toward known areas with a high payoff. The advantage of this method is that it exploits known space and explores potentially promising points, but it completely neglects areas with a low payoff and low uncertainty. This results in a very data-efficient scheme, as only a fraction of the parameter space is explored to reveal local maxima and minima. The tradeoff between exploitation and exploration is merged into a surrogate function which is then optimized. Not many data points are initially available, so the algorithm favors exploration at the cost of exploitation. However, over time, this balance shifts with further exploitation leading to a faster optimization of the surrogate function. In this phase, already acquired data points are exploited to refine suggestions and obtain optimized peaks. In this work, we used the expected improvement acquisition function to define the balance between exploration and exploitation (see Supporting Information for additional details).

2.3. Optimization Process

Specifically, we generated 100 000 random 3D vectors (representing the precursor ratios), adhering to the heuristic

constraints as mentioned earlier. For each NPL thickness, we determined the vectors with the highest overall optimization goal not predicted to exhibit a double peak, and synthesized NPLs with these precursor settings. Overall, we ran seven cycles with 14 syntheses per cycle for a total of 220 syntheses, including the initial ones. An overview of all the syntheses is given in Figure 3a, where each one is located in the 3D input parameter space (given by the volumetric ratios of the precursor solutions with the total volume normalized to 1) with the color signifying the emission wavelength. The overall trend is similar, with smaller Cs:Pb ratios leading to shorter emission wavelengths emanating from thinner NPLs. Note that the point density deviates significantly throughout the parameter space. This constitutes a visual representation of the algorithm initially exploring wide areas loosely and focusing on specific areas of high interest once it has developed a good understanding of the space. The quality of the PL spectra improved noticeably, as can be seen in Figure S1, Supporting Information, for the case of five ML NPLs. The FWHM narrows considerably from 38 to 19 nm (193 to 99 meV), while the f_{sym} is reduced by 1.607 to 0, achieving perfect symmetry. Additionally, the spectral position of the PL maximum shifts gradually from 490 to 486 nm. In terms of the FoM, this corresponds to an improvement from 0.429 to 0.916. Similar improvement was achieved for all previously established NPL thicknesses, as shown in Figure 3b, where the FWHM before (gray bars) and after optimization (colored) are shown for all thicknesses. The optimized spectra are displayed as colored curves in Figure 3c, where the dashed gray ones are from typical initial syntheses. All optimized spectra are significantly narrower and less asymmetric. All peak maxima exhibited a blueshift with the sole exception of the three ML NPLs. The improvement can also be seen in the FoM values, which increased substantially for all thicknesses. These are displayed in Table 2 along with the precursor amounts (both new and old). While there are only subtle shifts in the precursor amounts for some thicknesses, there are also significant deviations, such as for the four ML and six ML NPLs. This demonstrates the advantage of the employed approach, as it is unlikely these precursor ratios would have been heuristically selected. The quality of the spectra is noticeable even when compared to the best spectra obtained from the initial NPL synthesis (see Figure S2, Supporting Information).^[11]

2.4. Results of the Optimized NPL Syntheses

The algorithm performed very well, as shown in Figure 4. With every new synthesis added to the training sets, the error in predicting the quality factor decreased substantially (Figure 4a). As the downward trend is still steep at the end of the implemented training data, the algorithm would probably improve further with more syntheses. Nevertheless, even this level of accuracy was sufficient for the purposes of the application. In contrast, the predictor for the resulting PL emission wavelength improved rapidly for the first 30–40 syntheses, continued by a slower improvement (Figure 4b). This shows how well the precursor amounts define the PL emission of the resulting NPLs and that the algorithm could adapt quickly. The difference between predicted and measured PL emission maxima is

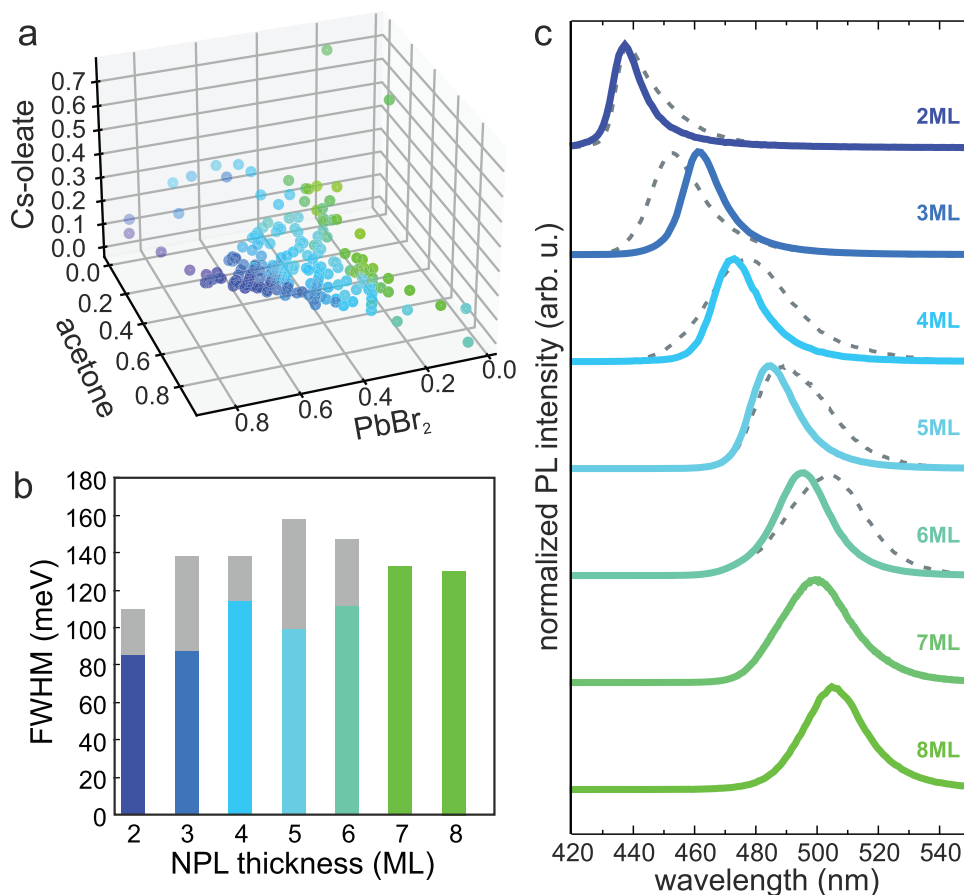


Figure 3. a) Results of all synthesized NPL dispersions displayed in the 3D precursor parameter space of precursor ratios, normalized, so that all volumes add to 1. The points are colored according to their respective PL emission maximum. b) The narrowness of NPL spectra was measured by FWHM values (in meV) before optimization (gray bars) and after optimization (colored bars). A clear improvement is observed for all samples. c) Optimized PL spectra for all NPL thicknesses (colored lines) compared to initial, typical PL spectra (dashed gray). As the seven and eight ML NPLs were only obtained through optimization, there are no initial spectra.

highlighted in Figure 4c. Here, there is a very strong grouping of the points along the diagonal (indicated as the gray line), indicative of very high accuracy. Interestingly, the algorithm predicted several precursor combinations to result in spectra with emission maxima at wavelengths longer than 494 nm, which corresponds to six ML NPLs. While not all of these resulted in NPLs, we were able to obtain two separate dispersions with distinct emission maxima at 501 and 505 nm, which we could assign to seven ML and eight ML NPLs, respectively

and confirm through TEM imaging (see Figure S3, Supporting Information). We must note that the eight ML NPL dispersion also contains small amounts of slightly thicker NPLs. Nevertheless, the algorithm enabled us to obtain specific NPL thicknesses for the first time, which also helped determine important excitonic properties of the strongly confined asymmetric nanocrystals.^[33] This is very impressive considering: a) the significantly low amount of data (i.e., syntheses) being fed into the algorithm and b) the fact that no data points in this

Table 2. Precursor amounts and FoMs for the optimized and conventional syntheses of different NPL thicknesses are listed.

NPL	Conventional ratios				Optimized ratios				% Improvement
	PbBr ₂	Cs-oleate	Acetone	FoM	PbBr ₂	Cs-oleate	Acetone	FoM	
2 ML	0.58	0.029	0.38	0.330	0.680	0.016	0.304	0.987	198
3 ML	0.411	0.041	0.548	0.727	0.488	0.024	0.488	0.560	No improvement
4 ML	0.358	0.045	0.597	0.711	0.472	0.127	0.401	0.794	11
5 ML	0.294	0.059	0.647	0.721	0.299	0.084	0.617	0.439	No improvement
6 ML	0.247	0.062	0.691	0.394	0.218	0.020	0.762	0.917	132
7 ML	–	–	–	–	0.302	0.132	0.566	0.459	–
8 ML	–	–	–	–	0.096	0.0030	0.814	0.469	–

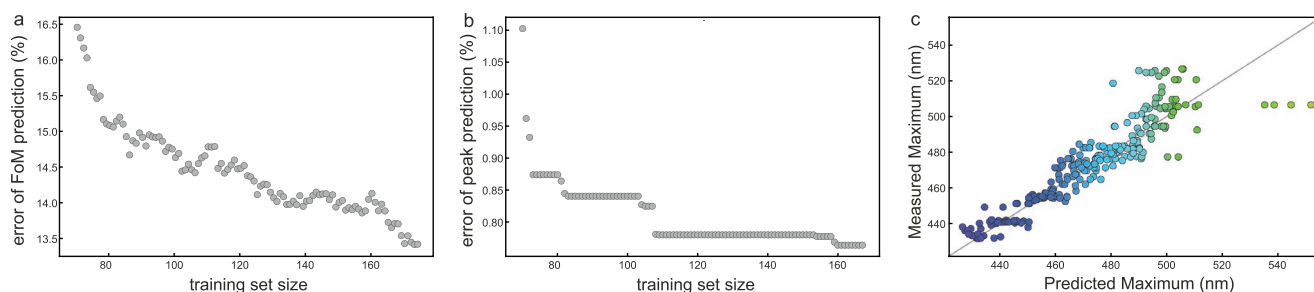


Figure 4. a) Evolution of error for the FoM prediction with an increasing number of training samples. The error would likely decrease even further with additional syntheses. b) Evolution of error for the peak position predictor with an increasing number of training samples. The initially low error has almost reached a plateau, confirming the high efficiency of the algorithm in predicting the resulting PL emission maxima. c) The final accuracy of the peak position predictor. The outlier points correspond to rare occurrences in the dataset and hardly decrease the overall evaluation. All results were generated with a 90–10 training validation split and tenfold validation.

spectral region were provided, proving the strong extrapolative abilities of our pipeline. We carried out only 220 syntheses, which is more than an order of magnitude less than in similar studies^[16] and our previous experience with the synthesis certainly helped streamline the process. Accordingly, in the next step, we will apply the algorithm to a novel synthesis, which we have not yet had experience with, and test how well it performs in this case and which parts might need further optimization. As is, our code and data are freely available for method evaluation, reproduction, and repurposing (see Experimental Section). Also, to further increase turnaround time and enhance reproducibility, we will look into merging the algorithm with high throughput, automated synthesis via microfluidics with temperature control.^[34,35]

3. Conclusion

We have implemented a novel machine-learning method to obtain CsPbBr₃ NPLs of a given thickness, with high optical quality and a vastly reduced amount of data. The algorithm, comprising three machine-learning models merged with Bayesian optimization, can optimize the spectral quality, thus the monodispersity of the samples and the desired emission wavelength, that is, the thickness of the resulting NPLs. Additionally, predicting and taking into account the purity of the resulting spectrum and incorporating additional heuristic constraints helped decrease the data pressure, meaning that the algorithm was able to rapidly improve the quality of NPL dispersions while requiring only at most three optimization cycles per NPL. In total, only 220 samples were required to obtain these results, which are orders of magnitude less than other, more complex approaches. Moreover, the model was able to identify precursor ratios leading to hitherto unobtainable NPL thicknesses, proving that the models learned correlations beyond the training set bounds, which is of high significance to the prediction certainty. Further, we acquired spectra that hint strongly at the formation of even thicker NCs, some of which are potentially unconfined; however, further investigation will be necessary to confirm these findings. The model can easily be expanded to account for additional synthetic parameters, such as temperature and humidity, and optimize quantum yields or stability. Able to improve a more-known synthesis so substan-

tially, the algorithm should be able to rapidly optimize the synthesis and quality of less-known materials. This also constitutes one of the next steps for the algorithm, besides implementing a high throughput automated synthesis via microfluidics to further speed up the process and ensure higher reliability.

4. Experimental Section

Synthesis: Cs₂CO₃ (cesium carbonate, 99%), PbBr₂ (lead(II)bromide, >98%), oleic acid (technical grade 90%), oleylamine (technical grade 70%), acetone (for HPLC, >99.9%), toluene (for HPLC, >99.9%), and hexane (for HPLC, >97.0%, GC) were purchased from Sigma-Aldrich and used without further purification.

The PbBr₂ ligand solution and Cs-oleate were prepared according to Bohn et al.^[11] The synthesis procedure remained the same for all used synthesis parameters. It is presented in the following: The synthesis was carried out under an ambient atmosphere at room temperature. The synthesis ratios were multiplied by three to get reasonable volumes to work with in the laboratory. A reaction glass was loaded with PbBr₂-precursor solution and Cs-oleate was immediately added under vigorous stirring. After 10 s, acetone was added quickly, and the reaction mixture was stirred for 1 min. Afterward, the mixture was centrifuged at 4000 rpm for 3 min, and the precipitate was redispersed in hexane (2 mL). Immediately after synthesis, the samples were optically characterized with a commercial spectrometer (FLOUROMAX-Plus, HORIBA).

Algorithm Details: In this chapter, a brief overview of each algorithmic component of the pipeline in order to facilitate its reproduction and reuse is presented. For the purposes of reproducibility of results, the code and data are available in the following repository https://github.com/benkour/Quantum_Dot_optimization.git. All of the information on the required package versions is available in the readme of the repository. Algorithm parameters that are not explicitly mentioned have default values that can be readily found in the documentation. All algorithms were tested with different random seeds and parameters, and without exception, the methods were found to be robust to hyperparameter changes.

Peak Position Prediction: A neural network was implemented through the tensorflow library^[32] and trained to identify the peak position based on the ratios of the precursor solutions. It consisted of three hidden layers with five neurons and an output layer with one neuron. The layer activations were respectively elu, relu, relu, elu, determined by statistical space search of the available activations, neuron number and layer number. The uniform variance scaling initializer was chosen as the weight initializer. The optimizer used was root mean squared propagation (RMSDrop) with exponential decaying learning rate and the training took 1000 epochs. The network was evaluated employing 30 random 98–2% splits and training for 1000 epochs. The reason for

the unusually small test set is the limited amount of data and was compensated by the multiple models trained.

Quality of Peak Prediction: The quality of the peak was assessed by combining two metrics, one for symmetry and one for width. The symmetry was defined according to Equation 1 and is reflecting the skewness metric drawn from the realm of statistics.

$$f_{\text{sym}} = 100 \frac{\text{mean} - \text{median}}{\text{mean}} \quad (1)$$

The narrowness f_{nar} was defined by the full width of the distribution at half minimum (FWHM), as defined routinely, for example ref. [36]

The total objective function was defined by

$$f_{\text{obj}} = \text{normalization} \left(\frac{1}{\text{normalization}(f_{\text{sym}}) + \text{normalization}(f_{\text{nar}})} \right) \quad (2)$$

where normalization() denotes the 0–1 min max normalization that brings both metrics to the same magnitude so that they can be taken into account equally.

A Gaussian process predictor was implemented through the Gpy library and trained to predict the quality of the peak using the ratios of the precursor solutions as inputs. The kernel used was the sum of a Matern32 and RBF kernels and optimized by maximization of the maximum likelihood estimation. The predictor was evaluated employing 30 random 98–2% splits and training for 1000 epochs. The reason for the unusually small test set is the limited amount of data and is compensated by the multiple iterations of the training-test cycle. It must be noted that the intuition behind this algorithm is the modeling of the predictions as normal distributions, whose parameters are optimized. It is therefore a natural result of this method that uncertainty is quantified simultaneously with the prediction and is reflected in the standard deviation of the resulting distributions.

Formability Prediction: The prediction of whether a composition will lead to a feasible, pure perovskite was undertaken by a random forest classifier with the maximum depth per tree being 150 and the minimum samples per split 10. The results were evaluated by performing a stratified fivefold validation. Both the stratified fold scheme and the classifier were implemented through the sklearn library.^[37]

A brief overview of the methods is given in **Table 3**.

Bayesian Optimization: Multiple acquisition functions were tested and the expected improvement was found to perform best for the given configuration. The initial value of the tradeoff between exploration and exploitation was 1 and it slowly decreased until in the final iteration it had the value 0.01. Further, it was suitably constrained with the constraints described in Section 4.

Confusion Matrix for the Multi-Peak Classifier: This double peak random forest classifier has an error of ≈15% and the confusion matrix shown in **Table 4**.

Constraints: The peak position was introduced through two separate constraining strategies. The first applied hard Lagrange multipliers of the form shown in Equation 3 on the upper and lower limit of the wavelength for each nanoplatelet number.

$$-e^{-\Xi[V_{\text{pred}} - V_{\text{min}}]} \text{ and } -e^{-\Xi[V_{\text{max}} - V_{\text{pred}}]} \quad (3)$$

with Ξ being a very large number, V_{max} is the maximum peak wavelength of the respective NPL thickness, and V_{min} the minimum wavelength of the respective NPL thickness. This gave more freedom to the algorithm but made it sensitive to peak prediction errors close to the boundaries.

Table 3. Method overview.

Method	Inputs	Outputs
Deep neural network	Precursor solution ratios	Peak position
Gaussian processes	Precursor solution ratios	Narrowness and symmetry metric
Random forest	Precursor solution ratios	0–1 for viable-non viable compositions

Table 4. Confusion matrix for number of peaks prediction.

Predicted behavior		True behavior		Total
		One peak	Multiple peaks	
One peak	One peak	167	25	192
	Multiple peaks	8	20	28
Total		175	45	220

To counter that a second strategy was developed that gave a small advantage to peaks closer to the middle point of the wavelength range. This constraint took the form shown in Equation 4

$$\frac{V_{\text{mean}} - V_{\text{pred}}}{V_{\text{min}} - V_{\text{max}}} \times F_{\text{FoM}} \quad (4)$$

with V_{mean} being the mean wavelength of the respective NPL thickness and F_{FoM} the predicted value for the symmetry and narrowness. In this fashion, a penalty proportional to the objective function was implemented, incentivizing but not constraining the optimizer to search closer to the middle point.

The formability and purity constraint was implemented with a single hard Lagrange multiplier:

$$-e^{-\Xi[1 - F_{\text{class}}]} - 1 \quad (5)$$

where $F_{\text{class}} = 1$ signifying a normal peak and $F_{\text{class}} = 0$ signifying multiple peaks present.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

C.L., I.K., and M.H. contributed equally to this work. This project was funded by: the European Research Council Horizon 2020 through the ERC Grant Agreement PINNACLE (759744), by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC 2089/1-390776260, by the Bavarian State Ministry of Science, Research, and Arts through the grant "Solar Technologies go Hybrid (SolTech)," the Project ProperPhotoMile, supported under the umbrella of SOLAR-ERA.NET Cofund 2 by The Spanish Ministry of Science and Education and the AEI under the project PCI2020-112185 and CDTI project number IDI-20210171; the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag project number FKZ 03EE1070B and FKZ 03EE1070A and the Israel Ministry of Energy with project number 220-11-031. SOLAR-ERA.NET is supported by the European Commission within the EU Framework Programme for Research and Innovation HORIZON 2020 (Cofund ERA-NET Action, N 786483), and by the "EXC 2089: e-conversion" DFG-cluster of excellence [project number: 390776260], <https://www.e-conversion.de/>. "ARTEMIS" - TUM innovation network, Technical University of Munich funded through the German Excellence Initiative and the state of Bavaria.

Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

Bayesian optimization, data-efficient optimization, Gaussian processes, halide perovskites, machine learning, nanocrystals, photoluminescence

Received: September 23, 2022

Revised: January 18, 2023

Published online: March 14, 2023

- [1] L. C. Schmidt, A. Pertegás, S. González-Carrero, O. Malinkiewicz, S. Agouram, G. Minguez Espallargas, H. J. Bolink, R. E. Galian, J. Pérez-Prieto, *J. Am. Chem. Soc.* **2014**, *136*, 850.
- [2] J. Shamsi, A. S. Urban, M. Imran, L. De Trizio, L. Manna, *Chem. Rev.* **2019**, *119*, 3296.
- [3] G. Abiram, M. Thanihaichelvan, P. Ravirajan, D. Velauthapillai, *Nanomaterials* **2022**, *12*, 2396.
- [4] J. Zhang, Q. Wang, X. Zhang, J. Jiang, Z. Gao, Z. Jin, S. F. Liu, *RSC Adv.* **2017**, *7*, 36722.
- [5] G. Gao, Q. Xi, H. Zhou, Y. Zhao, C. Wu, L. Wang, P. Guo, J. Xu, *Nanoscale* **2017**, *9*, 12032.
- [6] S. Shyamal, S. K. Dutta, T. Das, S. Sen, S. Chakraborty, N. Pradhan, *J. Phys. Chem. Lett.* **2020**, *11*, 3608.
- [7] J. Schoonman, *Chem. Phys. Lett.* **2015**, *619*, 193.
- [8] G. Schileo, G. Grancini, *J. Mater. Chem. C* **2021**, *9*, 67.
- [9] S. Weng, G. Yu, C. Zhou, F. Lin, Y. Han, H. Wang, X. Huang, X. Liu, H. Hu, W. Liu, Y. Wang, H. Lin, *Crystals* **2022**, *12*, 929.
- [10] J. Ye, M. M. Byranvand, C. O. Martínez, R. L. Hoye, M. Saliba, L. Polavarapu, *Angew. Chem. Int. Ed.* **2021**, *133*, 21804.
- [11] B. J. Bohn, Y. Tong, M. Gramlich, M. L. Lai, M. Dobliger, K. Wang, R. L. Z. Hoye, P. Muller-Buschbaum, S. D. Stranks, A. S. Urban, L. Polavarapu, J. Feldmann, *Nano Lett.* **2018**, *18*, 5231.
- [12] C. C. Stoumpos, D. H. Cao, D. J. Clark, J. Young, J. M. Rondinelli, J. I. Jang, J. T. Hupp, M. G. Kanatzidis, *Chem. Mater.* **2016**, *28*, 2852.
- [13] T. Morgenstern, C. Lampe, T. Naujoks, M. Jurow, Y. Liu, A. S. Urban, W. Brütting, *J. Lumin.* **2020**, *220*, 116939.
- [14] F. Mayr, M. Harth, I. Kouroudis, M. Rinderle, A. Gagliardi, *J. Phys. Chem. Lett.* **2022**, *13*, 1940.
- [15] Q. Tao, P. Xu, M. Li, W. Lu, *npj Comput. Mater.* **2021**, *7*, 23.
- [16] O. Voznyy, L. Levina, J. Z. Fan, M. Askerka, A. Jain, M.-J. Choi, O. Ouellette, P. Todorovic, L. K. Sagar, E. H. Sargent, *ACS Nano* **2019**, *13*, 11122.
- [17] H.-A. Chen, C.-W. Pao, *ACS Omega* **2019**, *4*, 10950.
- [18] P. R. Regonia, C. M. Pelicano, R. Tani, A. Ishizumi, H. Yanagi, K. Ikeda, *Optik* **2020**, *207*, 164469.
- [19] F. Baum, T. Pretto, A. Koche, M. J. L. Santos, *J. Phys. Chem. C* **2020**, *124*, 24298.
- [20] C. A. Rickert, E. N. Hayta, D. M. Selle, I. Kouroudis, M. Harth, A. Gagliardi, O. Lieleg, *ACS Biomater. Sci. Eng.* **2021**, *7*, 4614.
- [21] D. H. Wolpert, W. G. Macready, *IEEE Trans. Evol. Comput.* **1997**, *1*, 67.
- [22] Z. Li, L. E. Achenie, H. Xin, *ACS Catal.* **2020**, *10*, 4377.
- [23] Y. Zhang, X. Xu, *ChemistrySelect* **2020**, *5*, 9999.
- [24] A. Seko, S. Ishiwata, *Phys. Rev. B* **2020**, *101*, 134101.
- [25] P. V. Balachandran, B. Kowalski, A. Sehrioglu, T. Lookman, *Nature Commun.* **2018**, *9*, 1668.
- [26] J. Shamsi, P. Rastogi, V. Caligiuri, A. L. Abdelhady, D. Spirito, L. Manna, R. Krahne, *ACS Nano* **2017**, *11*, 10206.
- [27] J. A. Sichert, Y. Tong, N. Mutz, M. Vollmer, S. Fischer, K. Z. Milowska, R. García Cortadella, B. Nickel, C. Cardenas-Daw, J. K. Stolarczyk, A. S. Urban, J. Feldmann, *Nano Lett.* **2015**, *15*, 6521.
- [28] Y. Tong, E. Bladt, M. F. Aygüler, A. Manzi, K. Z. Milowska, V. A. Hintermayr, P. Docampo, S. Bals, A. S. Urban, L. Polavarapu, J. Feldmann, *Angew. Chem., Int. Ed.* **2016**, *55*, 13887.
- [29] A. N. Kumar Reddy, D. K. Sagar, *Pramana* **2015**, *84*, 117.
- [30] D. P. Doane, L. E. Seward, *J. Stat. Educ.* **2011**, *19*, 2.
- [31] GPy, GPy: A gaussian process framework in python, <http://github.com/SheffieldML/GPy> (accessed: November 2021).
- [32] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, **2015**, <https://www.tensorflow.org/>.
- [33] M. Gramlich, M. W. Swift, C. Lampe, J. L. Lyons, M. Dobliger, A. L. Efros, P. C. Sercel, A. S. Urban, *Adv. Sci.* **2022**, *9*, 2103013.
- [34] S. Sun, N. T. Hartono, Z. D. Ren, F. Oviedo, A. M. Buscemi, M. Layurova, D. X. Chen, T. Ogunfunmi, J. Thapa, S. Ramasamy, C. Settens, B. L. DeCost, A. G. Kusne, Z. Liu, S. I. P. Tian, I. M. Peters, J.-P. Correa-Baena, T. Buonassisi, *Joule* **2019**, *3*, 1437.
- [35] I. Lignos, R. M. Maceiczky, M. V. Kovalenko, S. Stavrakis, *Chem. Mater.* **2020**, *32*, 27.
- [36] M. Vashista, S. Paul, *Philos. Mag.* **2012**, *92*, 4194.
- [37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.