

Distributed Multi-Objective Bayesian Optimization for the Intelligent Navigation of Energy Structure Function Maps For Efficient Property Discovery

Edward O. Pyzer-Knapp,^{a*} Linjiang Chen^b, Graeme M. Day^c and Andrew I. Cooper^b

AUTHOR ADDRESS ^a IBM Research Europe, Hartree Centre, Sci-Tech Daresbury, UK,

^b Leverhulme Research Centre for Functional Materials Design, Department of Chemistry and Materials Innovation Factory, University of Liverpool, Liverpool, UK

^c School of Chemistry, University of Southampton, Southampton, UK

KEYWORDS. *Materials Discovery, Machine Learning, Simulation*

ABSTRACT: Energy-structure-function (ESF) maps have emerged as a powerful tool for in silico materials design, coupling crystal structure prediction techniques with property simulations to assess the potential for new candidate materials to display desirable properties. Despite continuing increases to accessible computational power, however, the computational cost of acquiring an ESF map often remains too high to allow integration into true high-throughput virtual screening techniques. In this paper, we propose the next evolution of the ESF map, which uses parallel Bayesian optimization to selectively acquire energy and property data, generating the same levels of insight at a fraction of the computational cost by limiting the expensive property calculations to a small fraction of the predicted crystal structures associated with a molecule. We utilize this approach to obtain a two orders of magnitude speedup on a previous ESF study that focused on methane capture materials, saving over 500,000 CPUh from the original protocol. Through acceleration of the acquisition of ESF-type insight, we pave the way for the use of ESF maps in automated ultra-high throughput screening pipelines. This greatly reduce the opportunity risk associated with the choice of system to calculate. For example, it will allow researchers to use ESF maps in the search for physical properties where the computational costs are currently just intractable, or to investigate orders of magnitude more systems for a given computational cost.

Introduction

In principle, the combination of machine learning and virtual computational screening is a powerful method for the discovery of new functional organic materials.^{1,2} Computational techniques show great promise for the calculation of both the thermodynamic stability and the associated functional properties of candidate materials, but it is often difficult in practice to exploit these breakthroughs. A key challenge is the prohibitive computational expense of accurately calculating energies and properties for every candidate material that is to be screened, and it is here where machine learning may be able to provide significant benefit.

One of the most challenging cases is the *a priori* design of functional molecular organic crystals with desirable materials properties. Unlike their framework-based counterparts, such as zeolites and MOFs,³⁻⁵ molecular crystals rarely obey simple geometric principles that can be exploited for rational design. Indeed, small changes to the molecular structure can have dramatic effects on the crystal packing and, hence, the resultant solid-state properties. Molecular crystal packing is often dictated by weak, competing intermolecular interactions: hence, the *a priori* design of materials with pre-determined, desirable properties requires a more subtle approach than for materials where structure and therefore function can be ‘designed in’ through the use of intuitive bonding rules, the ex-

ploitation of known framework topologies, or other geometric principles.

ESF Maps

Energy structure function (ESF) maps are a combination of crystal structure prediction (CSP) with per-structure-property calculation which have been shown to be a powerful tool for the virtual screening of candidate organic molecules for desirable properties such as deliverable gas capacity and charge carrier mobility.⁶ In an ESF map, candidate crystal structures are generated using CSP methodologies, which are then screened virtually for a desired property, such as gas adsorption capacity⁶ or charge mobility.⁷ The resulting pairing of energy and function is then used as an indicative tool for the propensity of the molecule to express the desired properties. This information can be used to guide an experimental campaign, which has been used to validate this ESF map approach.^{6,8} However, while this strategy can be effective, generation of the ESF map can be computationally intensive: for example, for the methane storage predictions, it took around 800,000 CPU hours to compute an ESP map for just one of the molecules in the study (T2E), which was distributed approximately equally between the cost of CSP and property calculations.⁶ The cost of computing ESF maps grows as the property of interest becomes more computationally expensive, and

also when the ESF maps contain larger numbers of candidate structures; this is particularly the case for porous materials, where the energy range that includes all observable crystal structures is extended by solvent templating. Multiple components (e.g., cocrystals) and multiple stable molecular conformers also increase the dimensionality of the energy landscape and, so, often lead to more possible crystal structures.^{8,9}

Bayesian Optimization

Bayesian optimization¹⁰ is a technique for evaluating a so-called black-box function; that is to say, a function for which there is not access to the analytical, closed form.

$$\min_{x \in X} f(x)$$

Bayesian optimization has become popular recently in the machine-learning community for the efficient tuning of the hyperparameters of deep learning models,¹¹ but given its strengths as a global optimizer, and its powerful theoretical guarantees,¹² it has also started to find applications in an increasingly diverse set of domains.^{13–16} The core application area of Bayesian optimization is when each sample of the function f is expensive to acquire, either in financial cost, acquisition time, or both, thus making this approach very attractive for our goal of more efficiently navigating large ESF maps.

Bayesian optimization has two fundamental principles. First, it promotes the use of a surrogate function \hat{f} to represent the true (unknown) function f that is being optimized. Since each data point is likely to be expensive to acquire, it is important that this surrogate function has robust and well-defined uncertainties associated with its evaluation. In this study, this model is a Gaussian process,¹⁷ although other models have been used.^{18,19}

A Gaussian process is a non-parametric machine-learning model, which can be described by a mean function, μ , and a kernel function, $K(x, x')$:

$$p(\mathbf{f}|X) = \mathcal{N}(\mu, K(x, x'))$$

Where \mathbf{f} is the vector of function values $[f(x_1), f(x_2) \dots f(x_N)]$ evaluated at input points $x_1, x_2 \dots x_N$. There are many potential choices for the kernel function $K(x, x')$ and for this study we used a Matern kernel²⁰:

$$C_{\frac{3}{2}} = \sigma^2 \left(1 + \frac{\sqrt{3}}{p} \right) \exp \left(-\frac{\sqrt{3}}{p} \right)$$

where the length-scale p is determined on a per-feature basis using the automatic relevance determination (ARD)²¹ protocol.

We also introduce a white noise kernel, whose scale is determined as a hyperparameter of the overall Gaussian process, and tuned to maximize the log-likelihood of the model with respect to the data.

The second major principle of Bayesian optimization is to balance exploration (the acquisition of new knowledge) and exploitation (the reliance on existing knowledge) when deciding which data points to acquire.²² This takes advantage of the existence of the uncertainties associated with the evaluations of the surrogate function \hat{f} and is controlled through a construct known as the acquisition function. There are a number of potential acquisition functions, with the most popular being Expected Improvement (EI),²³ which aims to maximize the expected improvement to the optimization of collecting a data point. While EI is seemingly a serial methodology, there have been strategies implemented recently that generalize to the parallel setting.^{18,24–26} Typically, these do not scale well with the number of dimensions, and those which do require sparsity and incoherence properties of the feature space that are not present in this problem.²⁶ Thompson sampling²⁸ solves this problem by approximating the predictive distribution as follows:

$$p(y_j|x_j, \mathcal{D}_{\mathcal{X}}) = \int p(y_j|x_j, \theta) p(\theta, \mathcal{D}_{\mathcal{X}}) d\theta$$

- where $p(\theta)$ represents the prior distribution given a set of data $\mathcal{D}_{\mathcal{X}}$ - thus approximating the posterior distribution using Monte Carlo, based upon a single sample from $p(\theta, \mathcal{D}_{\mathcal{X}})$. This method thus scales significantly better with the scale and dimensionality of the problem.

The use of Thompson sampling for parallel Bayesian optimization requires an adaptation of this methodology known as Parallel and Distributed Thompson Sampling (PDTs),¹⁸ which is described visually in the inset of Figure 1. PDTs extends the Thompson sampling framework to a parallel case, exploiting the fact that PDTs with batch size S is the same as running sequential Thompson Sampling S times without updating the current posterior. This allows the parallel and distributed calculation of the acquisition function, ensuring that this method is highly scalable with increasing batch size. This is particularly important in the case described here, since it allows for evaluations to be distributed over a cluster computer system, or even over a completely distributed system, such as IBM’s World Community Grid,²⁹ which harnesses the power of volunteer compute by harvesting ‘idle’ cycles from volunteer devices such as laptops, small computational systems, or even mobile devices.

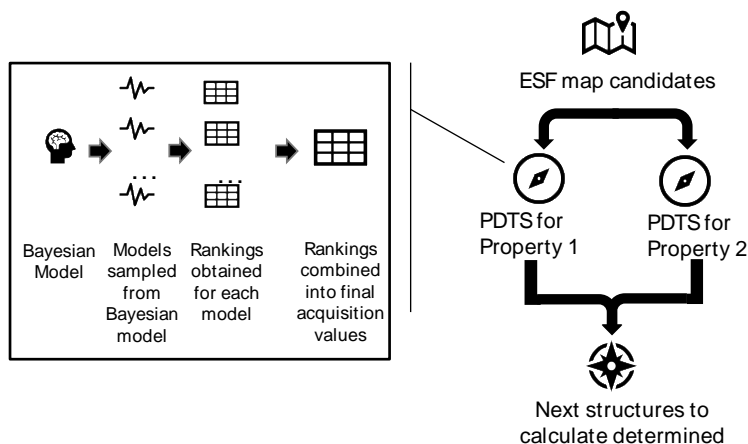


Figure 1 A graphical illustration of the intelligent navigation framework

In this study, we further extend PDTs to the multi-objective case (MO-PDTS) without harming the scalability and thus the parallel performance. To achieve this, we assign a separate PDTs sampler to each objective, the acquisition functions of which are then combined in a single step, determining the final acquisition function for the overall optimization process. Under a Gaussian process prior, this combination is equivalent to optimizing a single objective consisting of a weighted combination of objective values, with one significant advantage. Since the acquisition values are distinct from the models used to predict them in our MO-PDTS setting, each sampler can be built from a completely different set of descriptors.

Under the reasonable assumption that a model built from specifically chosen descriptors is more likely to have strong predictive ability than one built from a general set of descriptors, the ability to separate the predictors affords the user a framework that is significantly more transferable across a range of property types. Since the domain over which this optimization is being performed is a discrete set of structures, the ability to fully distribute this calculation is maintained.

METHODOLOGY

We test our methodology on three ESF maps that were calculated to evaluate the potential of three molecular materials for methane storage and delivery. We demonstrate how this new navigation workflow would have reduced the necessary computation and resulting *time to insight* for three for three systems, T2, T2E and P2 (Figure 3) recently predicted to have stable crystal structures with desirable methane deliverable capacities. These molecules were originally chosen because they represent a set of awkwardly-shaped molecules. Hence, they have the potential to form porous structures with high methane capacities, but intuitive packing arguments alone cannot provide sufficient insight to make *a priori* arguments about the relative potential of these three molecules to perform well in this application. ESF maps for this application are often very expensive, due to the fact that a large energy range of predicted crystal structures needs to be included to take into account the effects of solvent stabilization during crystal growth. Therefore, in the study of T2, P2 and T2E, crystal structures up to 100 kJ/mol were considered, compared to a more usual energy range of 10 or 15 kJ/mol for crystal structure landscapes when porous structures are not the target. Since the number of structures on the landscape increases rapidly as we move away from the global minimum, this 7-10 fold increased energy range leads to a much larger increase in the number of crystal structures that must be considered.

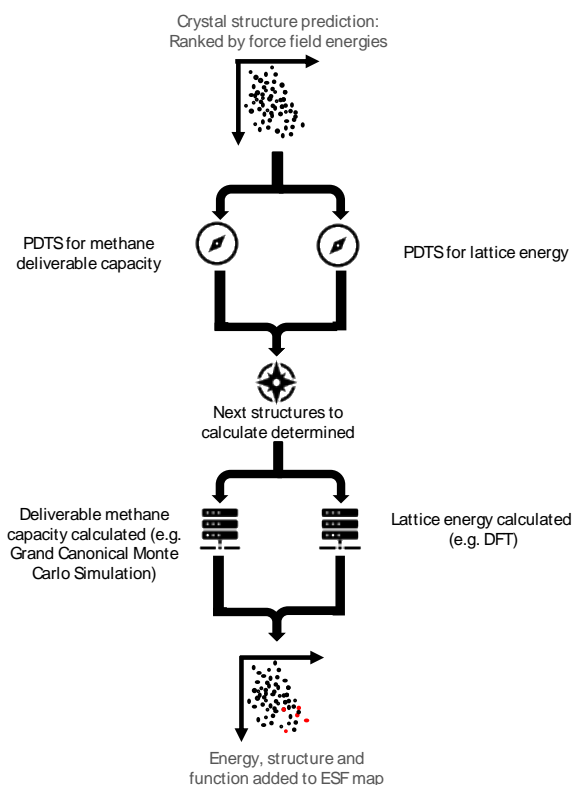
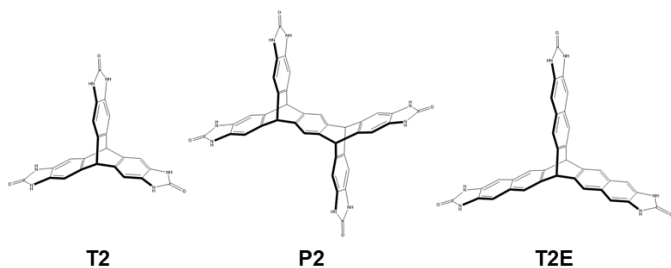


Figure 2 A flow-chart representing the utilization of MO-PDTS for accelerating ESF map construction. It should be noted that in some cases, a sufficiently accurate value for lattice energy is calculated at the initial generation stage, and in these cases, calculation of lattice energy is not necessary, but a second calculation at a higher level may be optionally employed.



MO-PDTS Optimization Details

Figure 4 Chemical structures of the three molecules in this study

We posed the problem as a multi-objective optimization over both energy and methane deliverable capacity, thus searching for the ESF maps for low-energy, highly porous, crystalline forms. For the purposes of this study, we are testing the methodology as if we do not have the final energies that we require, mimicking the case where higher level energies are required than were used in the structure search itself. We use the calculated force field energies as a proxy for these higher level energies. We do not consider the expense for these energy calculations when calculating savings, and so this study represents a lower bound on the potential for this method.

To demonstrate the modular nature of this approach, the two considered properties were modelled with different features – 28 geometrically defined features for porosity, and the NIST JARVIS³⁰ descriptor set for the lattice energy. Since JARVIS is a very high dimensional set of features with significant information redundancy, we use a principle component analysis to reduce the number of features, while retaining 99% of the variance. This resulted in the feature dimensions for the systems shown in Table 1.

System	Number of Dimensions
T2	45
P2	59
T2E	28

Table 1 Dimensionality of JARVIS descriptors, once reduced using PCA to retain 99% of original variance

To quantify the acceleration achieved, we compare our results here to the calculation of full ESF maps, previously reported by some of the authors⁶; that is, we accurately computed both lattice energies and methane deliverable capacities for all structures on the three associated ESF maps. Since ESF maps are used as indicators of the potential for a molecule to behave in a desirable way, we based our metric of success on the first encounter time for the global minimum on the ESF landscape; that is to say, the structure that has the best combination of low energy and high methane deliverable capacity. For this study, we weighted the contribu-

tion to this score from the energy term and the property term equally:

$$S = aE_i + bP_i$$

Where a, b are weighting coefficients to energy and property respectively, and in this study are equal and normalized to remove units and ensure the scales of the two properties are comparable. We note that for a more conservative approach, it is possible to weight the energy term more highly — that is, to increase the likelihood that the identified structure is thermodynamically accessible in the laboratory. Figure 4 shows that, in general, structures with high deliverable methane capacity have a high lattice energy. Thus, we expect that the number of structures which have both desirable methane deliverable capacity and low lattice energy to be small; further emphasising the need for an efficient, accelerated approach, and also the importance of the multi-objective nature of our search strategy.

The MO-PDTS was seeded with a novel initialization strategy based upon k-means, inspired by the generation of inducing points for sparse Gaussian processes. In

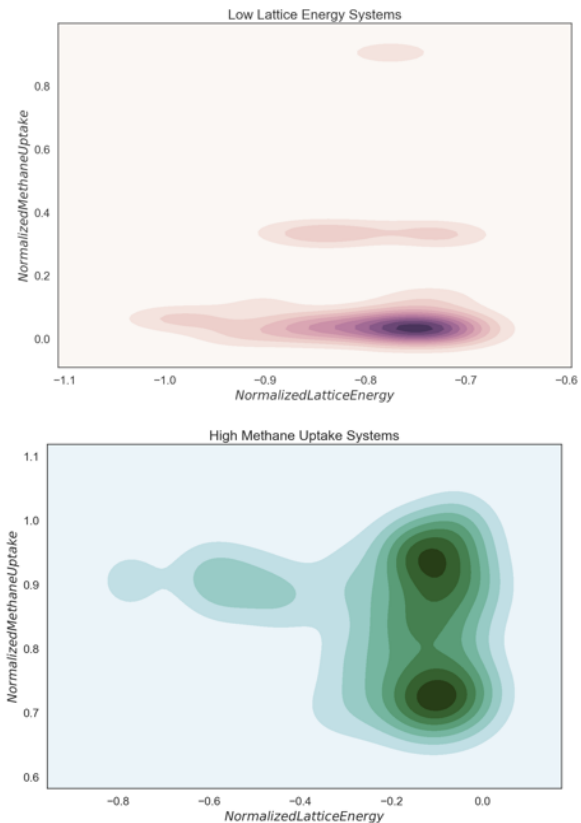


Figure 3 Estimated density plot for the normalized lattice energy and the normalized methane deliverable capacity for the T2 system, with values normalized over the entire dataset between 0,1 for methane capacity (-1,0 for energy). Upper plot focuses on crystal packings with high methane deliverable capacity; lower plot highlights systems with low lattice energies. Most of the low-energy systems have poor methane deliverable capacity, and the largest high methane capacity systems are relatively high in lattice energy; that is, these two properties are, broadly speaking, orthogonal

this methodology, k centroids were determined over input descriptor (feature) space using the k -means algorithm. The structures that minimized the distance to these centroids were chosen to initialize the search; that is, we selected the nearest structure to each of the k -centroids. Under a uniform distribution, this is equivalent to a Latin Hypercube due to the spherical repulsion of k -means. However, under a non-uniform distribution, we believe that this initialization captures the underlying data structure better, leading to increased model stability throughout the optimization process. MO-PDTS was then run for 10 epochs, at each of which 10 structures were selected, and properties calculated. In order to account for the difference in magnitudes of the two objectives, the values for each were scaled for each objective based upon the twenty selected structures from which the search was seeded.

Simulation Details

For each ESF map, candidate crystal structures were generated using a quasi-random sampling procedure, as implemented in the Global Lattice Energy Explorer software.³¹ Molecules were first sketched in ChemDraw, followed by an initial molecular-geometry optimization with the COMPASS force field,³² as implemented in the Materials Studio software package.³³ Force-field-optimized molecular geometries were further refined by re-optimization using density functional theory (DFT) with the M06-2X exchange-correlation functional and 6-311G** basis set. Molecular DFT calculations were performed with the Gaussian-09 software.³⁵ These molecular geometries were held rigid throughout crystal structure generation and lattice energy minimization.

Lattice energy calculations were performed with an anisotropic atom-atom potential using DMACRYS.³⁶ Electrostatic interactions were modelled using an atomic multipole description of the molecular charge distribution (up to hexadecapole on all atoms) from the B3LYP/6-31G**-calculated charge density using a distributed multipole analysis.³⁷ Atom-atom repulsion and dispersion interactions were modelled using a revised Williams intermolecular potential.³⁸

Methane deliverable capacity was predicted for each structure at a temperature of 298 K and a pressure of 65 bar. All of the adsorption predictions were performed using grand-canonical Monte Carlo (GCMC) simulations involving a 50,000-cycle equilibration period and a 50,000-cycle production run.

Results

The ESF maps for methane deliverable capacity for T2, P2 and T2E contained $\sim 5.5k$, $\sim 10k$, and $\sim 30k$ structures, respectively. To ensure reproducibility, and to display the robustness of our approach, the intelligent navigation workflow was tested for each system with 10 replicate experiments, each of which were seeded with different initial structures chosen from the landscape. Using these replicate experiments, we were able to use the bootstrap methodology to calculate confidence intervals for the convergence of each of the three systems with respect to ideal behaviour. All of the samplers converge on an ideal solution before 100 samples, or 10 epochs, have been completed. Since the executions are completed in parallel, when we calculate the first encounter time, we must only base this evaluation on the epoch in which the global minimum was discovered; that is, there is no advantage

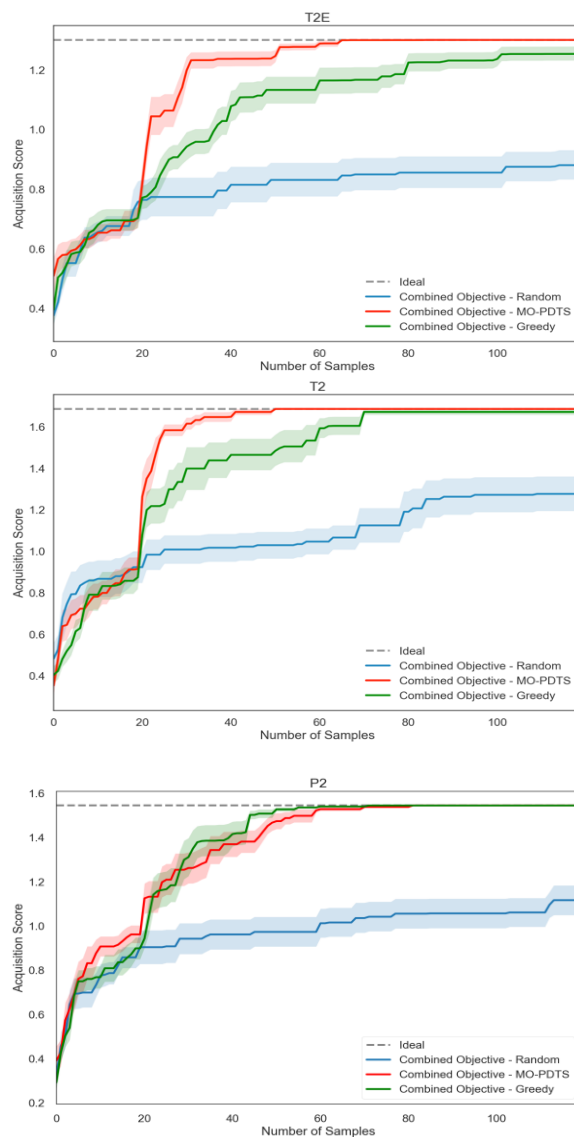


Figure 5 Performance of the MO-PDTS sampler for the three systems studied. Confidence intervals are generated using the bootstrap methodology from 10 replicate experiments seeded with different candidate structures

to being discovered halfway through a batch.

Table 2 shows the distribution of performance over the 10 repeats, with the best performance being achieved by the T2 system, which shows a mean first encounter time of 14.3 samples, or within 2 completed epochs. Both P2 and T2E have a mean first encounter time of around 4 epochs. This can be rationalized by the full ESF maps for these systems: when normalized lattice energy (range -1,0) is plotted against normalized methane deliverable capacity (range 0,1) –(Error! Reference source not found.), there are more structures in the bottom right of the map for T2 than for other systems, relating to structures that have both a high methane deliverable capacity and a low lattice energy. T2 also exhibits superior performance in our normalized objectives, with a score of circa 1.6, as compared to P2’s score of circa 1.5 indicating that there is a more favourable trade-off between low energy and high methane capacity structures.

Structure	Mean Encounter Time	Mean Epochs Required
T2E	39.0	4
T2	14.3	2
P2	34.0	4

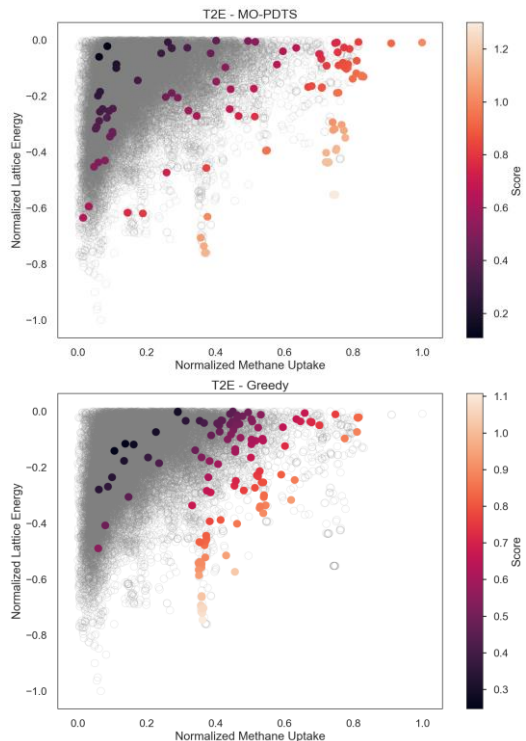
Table 2 The average performance achieved over 10 replicates for the three systems studied. Mean Encounter Time is the mean sample number at which the minimum is discovered, and Mean Epochs Required is the sampling epoch in which this sample fell.

Comparison to Greedy Sampling

An alternative approach to the reduction in computational cost for the exploration of large ESF maps (or compound libraries) is to use a greedy sampling method. For this class of search algorithm, a model is built from existing data, and used to predict values for data that has not yet been acquired. At each epoch of sampling, the candidate that has the largest predicted value is selected (or the smallest value, for minimisation purposes) and added to the training set, from which the model is then refitted. Indeed, most traditional QSPR methods use this methodology, either implicitly or explicitly, for accelerated materials discovery.

It can be seen from Figure 5 that in all cases the MO-PDTS sampler locates the ideal solution, and significantly outperforms the baseline random sampler. For T2E and T2 systems, there is a clear advantage over the greedy sampler, indicating that these are systems where there are competing local maxima, and demonstrating the advantage of the more sophisticated MO-PDTS method. In the case of P2, the performances are similar, indicating that

there is a single clear structure-property relation-



ship which can be exploited by the greedy sampler.

Figure 6 A comparison of the MO-PDTS and greedy sampling strategies. Candidate structures are colored by their combined energy-structure score, and no color indicates the structure was not sampled. It can be seen for the T2E case, that the greedy sampler gets stuck in a local maxima, but that PDTS is able to locate the global maximum

The dangers of a greedy sampler are clearly illustrated in the case of T2E, and shown diagrammatically in Figure 6. It can be seen that the greedy sampler identifies a reasonably well performing structure property relationship, and concentrates its sampling in this area. Unfortunately, this structure-property relationship does not indicate the existence of a second ‘peak’ of activity, of higher value. The balance of exploration and exploitation in MO-PDTS avoids this situation, and samples in a more intelligent and robust manner. The performance curves in Figure 5 indicate that there is little ‘cost’ to adopting this more sophisticated strategy over a more traditional, greedy approach when the structure-property landscape is simple, but significant benefits when it is not.

Computational Savings

We have seen that the proposed intelligent navigation approach to ESF maps yields considerable computational savings. The exact details are shown in Table 3. In all cases, we see greater than 2 orders of magnitude improvement in the ‘time to insight’,

which results in hundreds of thousands of saved CPU hours.

Structure	Number of structures in ESF map	% of ESF sampled	CPU h to generate full ESF map	Computational Saving* (CPU h)
T2E	29848	0.14	392,213	391,427
T2	5403	0.32	74,469	73,945
P2	9817	0.39	96,369	95,583

Table 3 Computational Savings as a fraction of the potential ESF Map for the systems T2, T2E and P2. *Computational saving is based upon averaging the cost of each calculation over the entire set

For the entire campaign, 544,955 hours were saved using this technique; for context, this saving is similar in magnitude to a small grant on a supercomputer system. For many functional properties, this high level of computational acceleration could transform ESF maps from a proof-of-concept demonstration to an important, routinely-used practical tool for in silico high throughput screening, particularly for physical properties that are expensive to compute. It is important to note that we assessed these savings based solely on the savings in property evaluations. In the case where higher level energies are required for lattice energy rankings, e.g., by DFT, the savings would be even greater. Even only considering the property calculation savings demonstrated in this study, these benchmark figures suggest that this technique might allow a user to screen orders of magnitude more candidates for the same computational expense. As with all accelerations of this kind, there is not the same completeness guarantee that is possible by calculating the entire ESF map. However, we believe that this is more than compensated by the huge increase in throughput and the ability to evaluate a much broader range of candidate molecular structures. In many cases, the use of this technique may be the difference between an ESF map for a particular property being calculated and being deemed too expensive. This represents a significant practical advance in the ESF methodology, allowing us to tackle new functional properties that have hitherto been deemed impossible because of their high computational cost.

Conclusions

We present an important evolution in the ESF mapping technique for the a priori prediction of materials properties: a smart navigator for ESF maps based upon multi-objective, parallel and distributed Thompson sampling (MO-PDTS). The scalability of this method adds negligible overhead to the computation of the ESF-map; by selectively sampling the map, and only requiring the use of expensive function calculations for a fraction of the structures,

we are able to make significant computational savings. For the three structures here studied, we were able to save well over half a million CPU hours. This has two key advantages – first, we significantly reduce the opportunity risk for the selection of systems for ESF map calculation; that is, did I choose the right molecule to spend this resource on? Second, through the reduction of computational requirements, we extend the power of the ESF map approach both to researchers who are not able to access the necessary computational resources and also to expensive property calculations for large, complex ESF maps that are simply intractable today.

ASSOCIATED CONTENT

AUTHOR INFORMATION

Corresponding Author

* Dr Edward O. Pyzer-Knapp epyzerk3@uk.ibm.com, IBM Research Europe, Hartree Centre, Daresbury, UK, WA4 4AD, .

Author Contributions

The manuscript was written through contributions of all authors.

Funding Sources

EPK – Innovation Return on Research Programme, funded by Department for Business, Energy and Industrial Strategy AIC, LC - Leverhulme Trust for supporting the Leverhulme Re-search Centre for Functional Materials Design. GMD - European Research Council for funding under the Eu-ropean Union’s Seventh Framework Programme (FP/2007-2013) through grant agreement number 307358 (ERC-stG-2012-ANGLE).

ACKNOWLEDGMENT

EPK acknowledges supported from the STFC Hartree Centre’s Innovation Return on Research programme, funded by the Department for Business, Energy & Industrial Strategy.

A.I.C. and L.C. acknowledge the Leverhulme Trust for supporting the Leverhulme Research Centre for Functional Materials Design.

GMD thanks the European Research Council for funding under the European Union’s Seventh Framework Programme (FP/2007-2013) through grant agreement number 307358 (ERC-stG-2012-ANGLE).

REFERENCES

- (1) Pyzer-Knapp, E. O.; Suh, C.; Gomez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Aspuru-Guzik, A. What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery. *Annual Review of Materials Research* **2015**, *45* (1), null. <https://doi.org/10.1146/annurev-matsci-070214-020823>.
- (2) Suh, C.; Fare, C.; Warren, J. A.; Pyzer-Knapp, E. O. Evolving the Materials Genome: How Machine Learning Is Fueling the

- Next Generation of Materials Discovery. *Annual Review of Materials Research* **2020**, *50* (1), 1–25. <https://doi.org/10.1146/annurev-matsci-082019-105100>.
- (3) Farha, O. K.; Yazaydin, A. Ö.; Eryazici, I.; Malliakas, C. D.; Hauser, B. G.; Kanatzidis, M. G.; Nguyen, S. T.; Snurr, R. Q.; Hupp, J. T. De Novo Synthesis of a Metal–Organic Framework Material Featuring Ultrahigh Surface Area and Gas Storage Capacities. *Nat Chem* **2010**, *2* (11), 944–948. <https://doi.org/10.1038/nchem.834>.
 - (4) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal–Organic Frameworks. *Nature Chemistry* **2012**, *4* (2), 83–89. <https://doi.org/10.1038/nchem.1192>.
 - (5) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The Materials Genome in Action: Identifying the Performance Limits for Methane Storage. *Energy Environ. Sci.* **2015**, *8* (4), 1190–1199. <https://doi.org/10.1039/C4EE03515A>.
 - (6) Pulido, A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B. J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; Stephenson, A.; Kane, C. M.; Clowes, R.; Hasell, T.; Cooper, A. I.; Day, G. M. Functional Materials Discovery Using Energy–Structure–Function Maps. *Nature* **2017**, *543* (7647), 657–664. <https://doi.org/10.1038/nature21419>.
 - (7) Yang, J.; De, S.; Campbell, J. E.; Li, S.; Ceriotti, M.; Day, G. M. Large-Scale Computational Screening of Molecular Organic Semiconductors Using Crystal Structure Prediction. *Chemistry of Materials* **2018**, *30* (13), 4361–4371.
 - (8) Cui, P.; McMahon, D. P.; Spackman, P. R.; Alston, B. M.; Little, M. A.; Day, G. M.; Cooper, A. I. Mining Predicted Crystal Structure Landscapes with High Throughput Crystallization: Old Molecules, New Insights. *Chemical Science* **2019**, *10* (43), 9988–9997.
 - (9) McMahon, D. P.; Stephenson, A.; Chong, S. Y.; Little, M. A.; Jones, J. T.; Cooper, A. I.; Day, G. M. Computational Modeling of Solvent Effects in a Prolific Solvatomorphic Porous Organic Cage. *Faraday discussions* **2018**, *211*, 383–399.
 - (10) Brochu, E.; Cora, V. M.; de Freitas, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *ArXiv10122599 Cs* **2010**.
 - (11) Snoek, J.; Larochelle, H.; Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. *arXiv:1206.2944 [cs, stat]* **2012**.
 - (12) Bull, A. D. Convergence Rates of Efficient Global Optimization Algorithms. *Journal of Machine Learning Research* **2011**, *12* (Oct), 2879–2904.
 - (13) Schultz, L.; Sokolov, V. Bayesian Optimization for Transportation Simulators. *Procedia Computer Science* **2018**, *130*, 973–978. <https://doi.org/10.1016/j.procs.2018.04.098>.
 - (14) Park, J.; Law, K. H. A Bayesian Optimization Approach for Wind Farm Power Maximization. In *Smart Sensor Phenomena, Technology, Networks, and Systems Integration 2015*; International Society for Optics and Photonics, 2015; Vol. 9436, p 943608. <https://doi.org/10.1117/12.2084184>.
 - (15) Pyzer-Knapp, E. Bayesian Optimization for Accelerated Drug Discovery. *IBM Journal of Research and Development* **2018**, *62* (6), 2–1.
 - (16) McDonagh, J. L.; Shkurti, A.; Bray, D. J.; Anderson, R. L.; Pyzer-Knapp, E. O. Utilizing Machine Learning for Efficient Parameterization of Coarse Grained Molecular Force Fields. *Journal of chemical information and modeling* **2019**, *59* (10), 4278–4288.
 - (17) Rasmussen, C. E. *Gaussian Processes for Machine Learning*; MIT Press, 2006.
 - (18) Hernández-Lobato, J. M.; Requeima, J.; Pyzer-Knapp, E. O.; Aspuru-Guzik, A. Parallel and Distributed Thompson Sampling for Large-Scale Accelerated Exploration of Chemical Space. In *International Conference on Machine Learning*; 2017; pp 1470–1479.
 - (19) Snoek, J.; Rippel, O.; Swersky, K.; Kiros, R.; Satish, N.; Sundaram, N.; Patwary, M.; Ali, M.; Adams, R. P.; others. Scalable Bayesian Optimization Using Deep Neural Networks. *arXiv preprint arXiv:1502.05700* **2015**.
 - (20) Matérn, B. *Spatial Variation. Medd. Fr. St. Skogsf. Inst.* **49** (5). Also Appeared as Number 36 of *Lecture Notes in Statistics*; Springer-Verlag, New York, 1960.
 - (21) Neal, R. M. *Bayesian Learning for Neural Networks*; Springer New York, 1996.
 - (22) Jasrasaria, D.; Pyzer-Knapp, E. O. Dynamic Control of Explore/Exploit Trade-Off In Bayesian Optimization. In *Science and Information Conference*; Springer, Cham, 2018; pp 1–15.
 - (23) Mockus, J. The Bayesian Approach to Global Optimization. In *System Modeling and Optimization*; Lecture Notes in Control and Information Sciences; Springer, Berlin, Heidelberg, 1982; pp 473–481.
 - (24) González, J.; Dai, Z.; Hennig, P.; Lawrence, N. D. Batch Bayesian Optimization via Local Penalization. *arXiv:1505.08052 [stat]* **2015**.
 - (25) Shah, A.; Ghahramani, Z. Parallel Predictive Entropy Search for Batch Global Optimization of Expensive Objective Functions. *arXiv:1511.07130 [cs, stat]* **2015**.
 - (26) Groves, M.; Pyzer-Knapp, E. O. Efficient and Scalable Batch Bayesian Optimization Using K-Means. *arXiv preprint arXiv:1806.01159* **2018**.
 - (27) Wang, J.; Clark, S. C.; Liu, E.; Frazier, P. I. Parallel Bayesian Global Optimization of Expensive Functions. *arXiv:1602.05149 [math, stat]* **2016**.
 - (28) Thompson, W. R. On the Likelihood That One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* **1933**, *25* (3/4), 285–294. <https://doi.org/10.2307/2332286>.
 - (29) Hinds, J.; Pyzer-Knapp, E. Accelerating Scientific Discovery through Crowdsourced Computing. In *ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY*; AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA, 2016; Vol. 252.
 - (30) Choudhary, K.; DeCost, B.; Tavazza, F. Machine Learning with Force-Field-Inspired Descriptors for Materials: Fast Screening and Mapping Energy Landscape. *Phys. Rev. Materials* **2018**, *2* (8), 083801. <https://doi.org/10.1103/PhysRevMaterials.2.083801>.
 - (31) Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M. Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling. *J. Chem. Theory Comput.* **2016**, *12* (2), 910–924. <https://doi.org/10.1021/acs.jctc.5b01112>.
 - (32) Sun, H. COMPASS: An Ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102* (38), 7338–7364. <https://doi.org/10.1021/jp980939v>.
 - (33) Accelrys Software Inc. *Materials Studio v6.10*; Accelrys Software Inc.; Accelrys Software Inc., 2012.
 - (34) *Materials Studio v6.10*; Accelrys Software Inc., 2012.
 - (35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö;

- Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian~09 Revision E.01.
- (36) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures Using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Physical Chemistry Chemical Physics* **2010**, *12* (30), 8478. <https://doi.org/10.1039/c004164e>.
- (37) Stone, A. J.; Alderton, M. Distributed Multipole Analysis. *Molecular Physics* **1985**, *56* (5), 1047–1064. <https://doi.org/10.1080/00268978500102891>.
- (38) Pyzer-Knapp, E. O.; Thompson, H. P. G.; Day, G. M. An Optimized Intermolecular Force Field for Hydrogen-Bonded Organic Molecular Crystals Using Atomic Multipole Electrostatics. *Acta Crystallographica Section B* **2016**, *72* (4), 477–487. <https://doi.org/10.1107/S2052520616007708>.