

Mapping Pareto Fronts for Efficient Multi-Objective Materials Discovery

Andre Low¹, Yee-Fun Lim¹, Kedar Hippalgaonkar¹, and Eleonore Vissol-Gaudin¹

¹Affiliation not available

December 12, 2022

Abstract

With advancements in automation and high-throughput techniques, complex materials discovery with multiple conflicting objectives can now be tackled in experimental labs. Given that physical experimentation is greatly limited by evaluation budget, maximizing efficiency of optimization becomes crucial. We discuss the limitations of using hypervolume as a performance indicator for desired optimality across the entire multi-objective optimization run and propose new metrics specific to experimentation: ability to perform well for complex high-dimensional problems, minimizing wastage of evaluations, consistency/robustness of optimization, and ability to scale well to high throughputs. With these metrics, we perform a comparison of two conceptually different and state-of-the-art algorithms (Bayesian and Evolutionary) on synthetic and real-world datasets. We discuss the merits of both approaches with respect to exploration and exploitation, where fully resolving the Pareto Front could be the main aim for greater scientific value in understanding materials space, and thus provide a perspective for materials scientists to implement optimization in their platforms.

Corresponding author(s) Email: kedar@ntu.edu.sg

Introduction

Materials science as a field is being disrupted with advances in machine learning and automation, (Lookman et al., 2015; Liu et al., 2017; Correa-Baena et al., 2018) where high-throughput experimentation (HTE) capabilities accelerate discovery of materials in more complex search spaces. Users not only save time on experimentation by virtue of automated workflows with faster processing, but also leveraging on equipment with larger batches of experiments to increase throughput and thus minimise experimental time. (Zhang and Block, 2009; Mennen et al., 2019) There have been many successful applications of HTE, particularly in the single objective problem space alongside machine learning-assisted optimisation strategies. (Sun et al., 2018, 2019; Burger et al., 2020; Dave et al., 2020; Gongora et al., 2020; Langner et al., 2020; Li et al., 2020; Shimizu et al., 2020; Wang et al., 2020; Bash et al., 2022; Deneault et al., 2021; Mekki-Berrada et al., 2020) However, many real-world problems are more complex, specifically with multiple conflicting properties to be optimized, for example: strength vs ductility in metal alloys, (Li et al., 2016) device thickness vs fill factor in photovoltaics, (Ramirez et al., 2018) or selectivity vs current density in catalysts. (Ren et al., 2019) In addition, such problems may include constraints that restrict the space of feasible solution. This motivates the need for multi-objective optimisation strategies with constraint handling capabilities to be integrated in HTE setups. (Alsharif et al., 2020; Bash et al., 2020; Grizou et al., 2020; Abdel-Latif et al., 2021) The first step could consist of formulating complex material science problems as constrained multi-objective optimisation problems (CMOPs).

A CMOP with m objectives and $(q+k)$ constraints, can be defined as:

$$\begin{aligned} \min [?] F(x) &= (f_1(x), \dots, f_m(x))^T \\ \text{st } g_i(x) &\geq 0, \quad i = 1, \dots, q \\ h_j(x) &= 0, \quad j = 1, \dots, k \\ x [?] &R^n \end{aligned}$$

where $F(\mathbf{x})$ defines the multi-dimensional objectives to be optimised, and $g_i(\mathbf{x})$ and $h_j(\mathbf{x})$ define the inequality and equality constraints, respectively. A solution is an n -dimensional vector of decision variables, \mathbf{x} . To determine the objective value of a solution, a Pareto-optimal solution x^1 dominates another solution x^2 if $F(x^1) \leq F(x^2)$ where they are feasible. A total set of all feasible and Pareto-optimal solutions can then be defined as the Pareto Set, or Pareto Front (PF) when mapped onto the objective space. This PF represents all solutions with the optimal trade-off between objectives.

A commonly defined materials discovery problem is usually of combinatorial nature with unexplored regions of objective space, given some mixture of chemicals, precursors, and other process parameters. This problem can be formulated as a CMOP with an unknown PF to be extrapolated to, with minimal evaluation budget. (Yong et al., 2022; Lim et al., 2021; Sabharwal et al., 2016; Klein et al., 2015) This is achieved through selection and evaluation of available solutions, where each solution represents the set of experimental input parameters (chemicals, temperature settings etc.) used in the screening. The number of data points is typically low, with most works generally limited to around 10^2 - 10^3 data points due to practical bottlenecks such as time taken to synthesize and characterize, or simply due to a limited time/cost budget.

In addition, the PF can be discontinuous with multiple infeasible regions due to underlying property limitations such as phase boundaries/solubility limits, or engineering rules, for example summing mixtures to 100%. (Gopakumar et al., 2018) Such constraints can also be knowledge-based, where a domain expert with prior knowledge sets them to pre-emptively ‘avoid’ poor results and converge faster. (Niculescu et al., 2006; Asvatourian et al., 2020; Liu et al., 2022) **Figure 1** illustrates an example of such a problem.

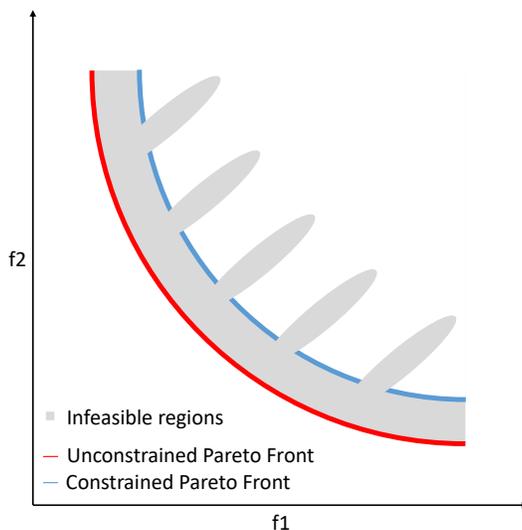


Figure 1: Illustration of constrained multi-objective (f_1 and f_2) space for a convex minimization problem in bi-objective space. The addition of infeasible regions in grey shifts the original PF from solid red to blue.

CMOPs can be solved in various ways, but recently, two classes of algorithms have shown promises in solving such problems with a high level of success, namely: multi-objective evolutionary algorithms (MOEA) and

multi-objective Bayesian optimisation (MOBO).

MOEAs (Deb, 2011) work by maintaining and evolving a population of solutions across an optimisation run. For example, Genetic Algorithms (GA) are a specific subset that utilise ‘operations’ alike biological processes: members of the population are selected to become parents based on a specific selection criterion, and then undergo crossover and mutation to form a children population. (Mitchell, 1998) Within the field of MOEAs, various constraint handling techniques have been proposed (Fan et al., 2019; Xu and Zhang, 2020; Tian et al., 2022) as well as extensions of MOEAs to many-objective ($m > 2$) problems. (Li et al., 2015) MOEAs are well suited to implementations where solutions can be tested in parallel, given their population-based approach, where each generation’s population can be treated as a batch. MOEAs have been successfully applied in materials-specific multi-objective problems: experimental data is used to construct a machine learning model which is then treated as a computation optimisation problem to be solved, and the results evaluated physically. (Zhang et al., 2021b; Patra et al., 2017; Menou et al., 2016; Coello and Becerra, 2009; Ganguly et al., 2007; Mahfouf et al., 2005) The use of MOEAs relevant to materials science has seen computational and inverse design problems. (Wu et al., 2020; Avery et al., 2017; Berardo et al., 2018; Pakhnova et al., 2020; Carvalho et al., 2020; Jennings et al., 2019; Salley et al., 2020)

MOBOs leverage on surrogate models to cheaply predict some black-box function, and then utilise an acquisition function to probabilistically compute a predictive function and return the best possible candidate where gain is maximised. (Shahriari et al., 2016) The choice of surrogate model can depend on the user, but in recent literature, it has become synonymous with ‘kriging’ which refers specifically to the use of Gaussian Processes (GP) as the surrogate model, taking advantage of its flexibility and robustness. (Rasmussen, 2003) The extension of MOBOs to CMOPs is less mature, with relatively new implementations that cover parallelization, multi-objective and constraints. (Garrido-Merchán and Hernández-Lobato, 2019; Belakaria et al., 2020; Daulton et al., 2020; Suzuki et al., 2020) On top of these, there are also hybrid variants such as TSEMO (Bradford et al., 2018) or MOEA/D-EGO (Zhang et al., 2010) which integrate the use of MOEAs to improve the prediction quality of the underlying surrogate models. In general, BO as an overarching optimisation strategy has already been established as an attractive strategy for use in both computational design problems, (Mannodi-Kanakithodi et al., 2016; Solomou et al., 2018; Yuan et al., 2018; Karasuyama et al., 2020; Janet et al., 2020; Hanaoka, 2021) as well as experimentation problems (MacLeod et al., 2022, 2020; Cao et al., 2021; Schweidtmann et al., 2018; Christensen et al., 2021; Epps et al., 2020; Erps et al., 2021) due to its sample efficient approach.

As previously discussed, the PF defines the set of optimal solutions of a CMOP. For optimisation of CMOPs, hypervolume (HV) is often used as a performance indicator. It defines the Euclidean distance bounded by a point, and the reference point in a single dimension, and a HV in multiple dimensions. It directly shows the quality of the solutions since a solution set with high HV is closer to the true PF and is diverse as it effectively dominates more objective space. An illustration of the HV measure for a multi-objective (two dimensions for illustration) convex minimization problem is presented in **Figure 2**, where HV is computed by finding the area of non-dominated solutions, i.e. the solutions closest to PF without any competitor, bounded by a reference point.

Aside from being a performance metric to compare optimisation strategies, HV can also be directly evaluated to guide convergence of various algorithms. Hanaoka et al showed that scalarization-based MOBOs may be best suited for clear exploitation and/or preferential optimisation trajectory of objectives, whereas HV-based MOBOs are better for exploration of the entire search space. (Hanaoka, 2022) Indeed, HV-based approaches empirically show a preference in proposed solutions towards the extrema of a PF, (Auger et al., 2012; Guerreiro et al., 2020) and thus can better showcase extrapolation. In contrast, scalarization approaches to reduce multi-objective problems to a single-objective such as hierarchically in Chimera (Häse et al., 2018) or any user-defined function (Zhang et al., 2021a) have limitations: i) it is difficult to determine how to properly scalarize objectives; ii) single objective optimisation methods cannot propose a set of solutions that balance trade-off.

Within the context of multi-objective optimisation and material science implementation, two state-of-the-

art algorithms were compared in the present work: q-Noisy Expected Hypervolume Improvement (qNE-HVI) (Daulton et al., 2021) and Unified Non-dominated Sorting Genetic Algorithm III (U-NSGA-III). (Seada and Deb, 2016) They are MOBO and MOEA-based algorithms, respectively, and were chosen based on their reported performance in solving complex CMOPs (with respect to HV score), and the fact that they are capable of highly parallel sampling, making them suitable for integration within an HTE framework. Furthermore, both algorithms are chosen from open-source Python libraries, making them easy to implement and enabling reproducibility of results presented.

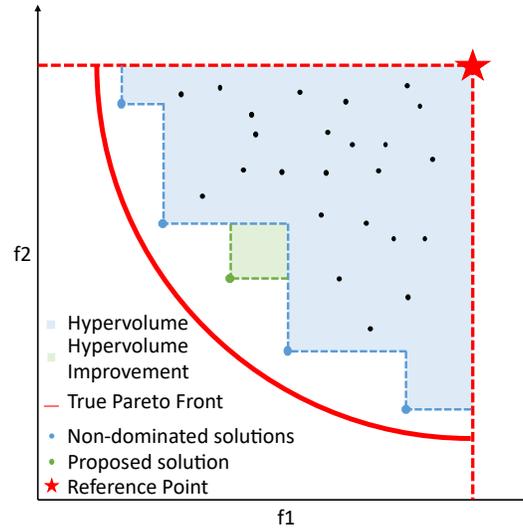


Figure 2: Illustration of hypervolume for a convex minimization problem in bi-objective space. The red line represents the ground truth PF, while the blue points and region reflect the best-known solutions and their associated hypervolume, respectively. The green point and region are then used to illustrate the contribution of a new evaluated solution. The computation of hypervolume in objective space is performed with respect to a lower bound with a reference point, shown by the red star.

Experimental Section/Methods

We thus propose 4 different metrics.

Firstly, dimensional contour plots – 10 runs at a relatively large evaluation budget (100 iterations x 8 points per batch) are plotted for number of dimensions versus total evaluations, colored by HV score. This is done for the scalable synthetic problems only, and allows us to illustrate performance when dimensionality is scaled up to represent more complex combinatorial problems.

Secondly, optimisation trajectory – a single optimisation run at high evaluation budget (100 iterations x 8 points per batch) is plotted in objective space to illustrate the trajectory of proposed solutions at each iteration towards the PF. This allows us to graphically analyze how either algorithm traverses the objective space, and provides a different perspective in understanding the exploration-exploitation trade-off.

Thirdly, probability density map – 10 runs at a lower evaluation budget (24 iterations x 8 points per batch) are plotted together in objective space, and colored with a Gaussian kernel density estimate to illustrate the probability distribution of solutions. This is an alternative to optimization trajectory, where instead we consider the consistency and robustness during optimization for different random starts.

Lastly, batch sizing – various batch sizes are compared using log HV difference to illustrate their HV improvement, and thus illustrates the performance of both algorithms when considering different throughputs, as well as whether more gradual optimization (smaller batches but higher iterations) or vice versa is appropriate.

In all cases, we initialised with a Sobol sampling of $2^*(\text{variables}+1)$.

Results

Synthetic Problems

For synthetic benchmarks, we select two-objective scalable problems for comparison as described in **Table 1**. The ZDT test suite (Zitzler et al., 2000) provides a range of PF shapes, while the MW test suite (Ma and Wang, 2019) provides constraints and uniquely shaped PFs to challenge the optimisation algorithms. Both test suites rely on a similar construction method for minimization problems: taking a single variable function f_1 against a shape function f_2 as such:

$$\begin{aligned} \min [?] f_1(x) &= x_1 \\ \min [?] f_2(x) &= g(x)h(f_1(x), g(x)) \end{aligned}$$

The single variable function closely resembles certain real-life multi-objective problems where an input is to be minimised against some other objective, for example minimizing process temperature, while achieving a target output. (MacLeod et al., 2022) U-NSGA-III in **Figure 3** a) and c) shows a more gradual change in

Name	PF Geometry	n_var	n_obj	n_constr	ref_pt
ZDT1	Convex	Scalable	2	0	[11, 11]
ZDT2	Concave				
ZDT3	Disconnected				
MW7	Disconnected (mixed)			2	[1.2, 1.2]

Table 1: List and details of synthetic problems.

colour and did not reach the maximum values for higher dimensions, indicating a slower rate of convergence and poorer HV improvement, respectively, which scale with dimensions. In contrast, results presented in Figure 3 b) and d) for ZDT1 and ZDT2, respectively, indicate that qNEHVI converges fast at a high HV improvement, as illustrated by the bright yellow coloration which appears early and maintains this up to dim=12 with little loss in initial performance. qNEHVI, while showing superiority in overall HV score for the ZDT3 and MW7 problem, had a lower rate of convergence and maximum HV improvement as dimensions increase, illustrated in Figure 3 f) and h) by the colour gradient. Although we note that in other literature, GP models tend to perform poorly at high dimensionalities, (Moriconi et al., 2020; Eriksson and Jankowiak, 2021) this was not observed here, to the limit of 12 dimensions. We believe that the underlying stochastic QMC sampling used is what drives the optimisation and hence the performance remains robust.

It should be noted that in Figure 3 e), U-NSGA-III’s HV score on the ZDT3 problem scales inconsistently with dimensionality: dim=5 shows better HV improvement (brighter colour) compared to dim=2 to 4. We attribute this to the disconnected PF being strongly affected by differences in initialisation, where entire regions can be lost as the evolutionary process fails to extrapolate and explore sufficiently. Lastly, we observe in Figure 3 g) for MW7 that U-NSGA-III performs significantly worst as compared to qNEHVI, regardless of dimensionality. The presence of more complex constraints in the problem means that many solutions are likely to be infeasible and require more iterations to evolve to feasibility according to the evolution mechanism. Infeasible solutions do not contribute to HV improvement at all, and we note that this is one of the limitations of plotting using HV as a metric, where feasibility management is not clearly reflected.

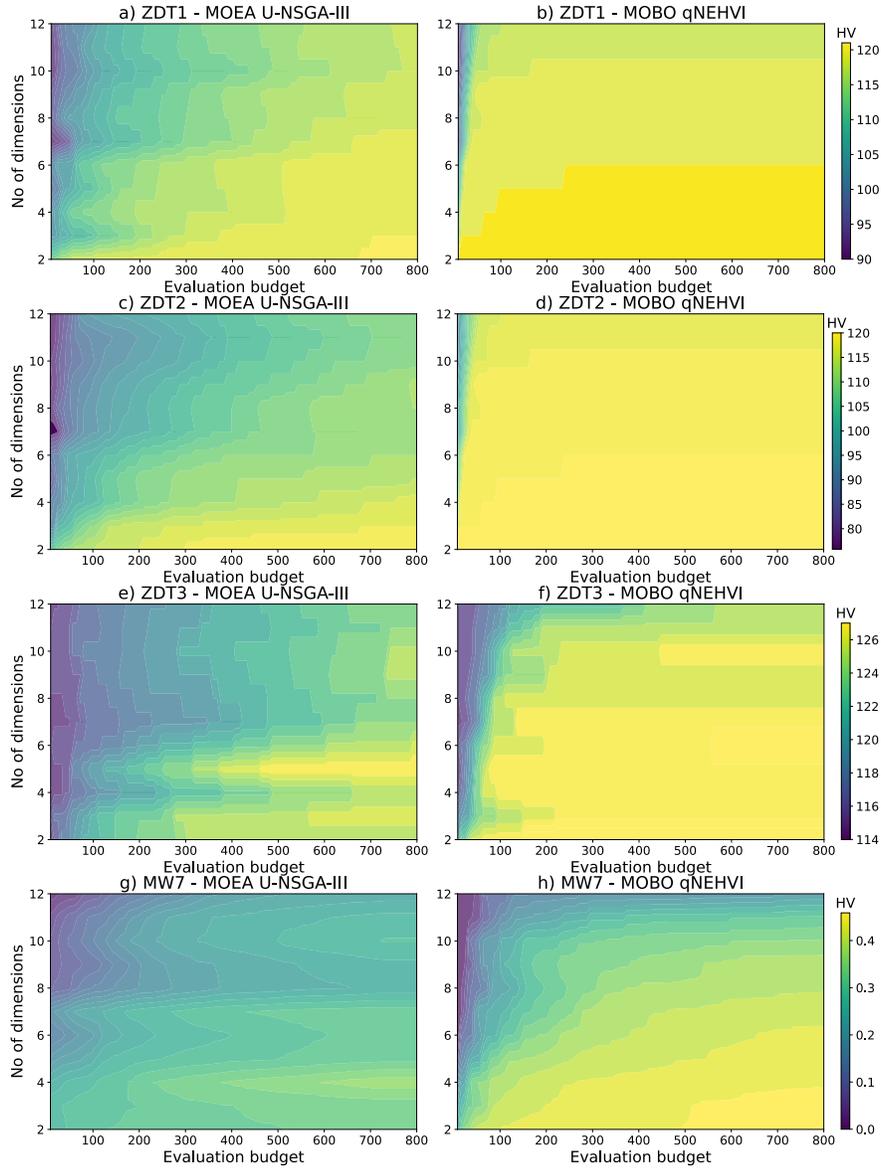


Figure 3: Contour plots for dimension vs evaluation budget. a-b) ZDT1. c-d) ZDT2. e-f) ZDT3. g-h) MW7. The colour bar illustrates the mean cumulative HV score with respect to cumulative evaluations, over a total evaluation budget of 100 iterations x 8 points per batch. Results are averaged over only 5 runs due to high computational cost of searching over many dimensions. The results here show that qNEHVI is a far superior method when looking at only HV as a performance metric.

In order to investigate why qNEHVI presented a higher HV improvement for qNEHVI, we then proceed to plot the optimization trajectory to observe solutions in objective space, as shown in **Figure 4**. We set the number of dimensions to 8. This is representative of a range of experimental parameters that materials scientists would consider practical. We first performed a single optimisation run of 100 iterations x 8 points per batch. The evaluated solutions are plotted onto the objective space and coloured by their respective iteration from dark to bright.

The general observations in Figure 4 a)-h) comparing qNEHVI to U-NSGA-III are consistent with results

previously reported in Figure 3, specifically in terms of HV scores and convergence rate. In all sub figures, qNEHVI was able to propose solutions at the PF within the first 20 iterations, as shown by the darker colour of points along the red line (true PF). This suggests that it is very sample efficient. However, it was unable to fully exploit the region of objective space close to the PF, and solutions in later iterations are non-optimal. In fact, in Figure 4 b) and d), ZDT1 and ZDT2 respectively, a large portion of solutions lie along the $f_1=x_1=0$ line. This is explained by the choice of reference point, which we explore in more detail in SI 1.

We hypothesize that qNEHVI is unable to identify multiple bi-objective points along the PF because the underlying GP surrogate model did not accurately model the PF for ZDT1-3. As for MW7, despite the algorithm being able to propose many solutions near the unconstrained PF, it failed to overcome the constraints, as seen by the failure to adjust to the new dotted red line. We observed that qNEHVI’s superior HV score (Figure 3) could be attributed to the stochastic nature of QMC sampling, which is used to provide a pool of candidates for the surrogate model and acquisition function to determine the next ‘best’ batch of points to evaluate. This hypothesis is supported by results reported in SI 2, where it can be observed that the GP model did not fully learn the objective function.

In contrast, U-NSGA-III, while requiring a significantly larger number of iterations to reach the PF, had a more consistent optimisation trajectory towards the PF, as seen by the gradual colour gradient in Figure 4 a), c), e), g). This suggests that there are less wasted evaluations for MOEAs, as the latter iterations are targeted towards the PF. However, despite having more solutions near the PF, the HV score is lower for U-NSGA-III than qNEHVI. This is a limitation of using HV as a performance metric: it strictly rewards non-dominated solutions across the entire search space, i.e. a handful of solutions at the PF extrema are preferred, as shown previously in Figure 3 where U-NSGA-III showed poorer HV improvement compared to qNEHVI for ZDT1, ZDT3 and MW7.

Notably, we observe in Figure 4 e) and g) that the disconnected PFs for ZDT3 and MW7 can lead to entire regions of objective space being omitted. This is clearly seen in both sub-figures where solutions only have a single trajectory towards the nearest PF region. We previously made the statement, based on results reported in Figure 3 c) and d), for the same synthetic problems, that the disconnected spaces are strongly influenced by initialisation, where U-NSGA-III’s mechanism of tournament selection rewards immediate gain over coverage, i.e. exploitation over exploration. This is both a strength and weakness of U-NSGA-III in comparison to qNEHVI, where the stochastic QMC sampling enables greater exploration of the overall search space, but not the PF.

Results reported in **Figure 5** further reinforce the observation that qNEHVI produces a large pool of non-optimal solutions for all benchmarks problems, where many points exist away from the PF. Additionally, the darker coloration for qNEHVI in Figure 5 b), d), f) and h) indicates a much lower probability of occurrence, which reinforces our hypothesis, that HV improvement can be partially attributed to the stochastic nature of QMC sampling. Additionally, Figure 5 b) and d) for ZDT1 and ZDT2 respectively also show that there were many solutions being proposed at the extrema of $f_1=x_1$.

This is the same behaviour as that observed for a single run in Figure 4 b) and d), and we further elaborate upon it in SI 1. In contrast, the heuristic nature of U-NSGA-III provides more consistency between optimisation runs, which is shown by the brighter regions of points near the PF in Figure 5a), c), e) and g). indicating a higher probability density. Notably, the bright regions are not spread across objective space evenly. There is a preference for the lower range of $f_1=x_1$ since it is easily tunable, i.e. it is simple to derive improvement by simply decreasing x_1 . This is in line with our previous discussions based on results reported in Figure 4, where U-NSGA-III prefers solutions with immediate improvement. Furthermore, we observe that the bright regions are concentrated near the PF, which indicates that U-NSGA-III was able to consistently approach the PF and maintain a larger pool of near-Pareto solutions over the optimisation runs, despite the limited evaluation budget.

In contrast, qNEHVI had relatively few points, although they are lying directly on the PF, which is then

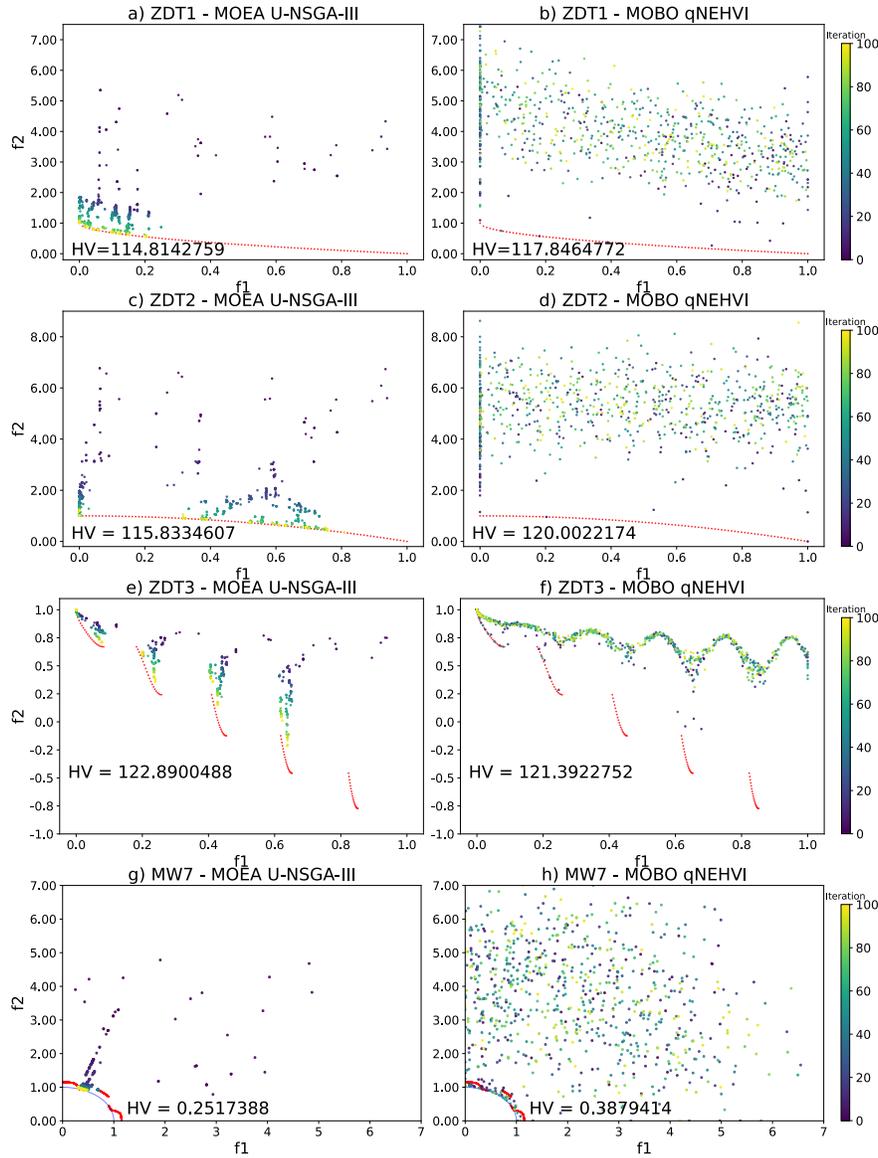


Figure 4: Optimisation trajectory in objective space for a single optimisation run of 100 iterations x 8 points per batch. a-b) ZDT1. c-d) ZDT2. e-f) ZDT3. g-h) MW7. The red line represents the true PF, while MW7 being a constrained problem has an additional blue line to show the unconstrained PF. The colour of each experiment refers to the number of iterations. All problems clearly show a more gradual evolution of results as the number of iterations progress in U-NSGA-III whereas qNEHVI rapidly approaches PF and then fails to converge further.

shown as a higher mean HV compared to U-NSGA-III. In a real-world context, the larger pool of near-Pareto solutions could have scientific value, especially for users looking to build a materials library and further understand the PF. However, this is not reflected by the HV performance indicator.

Optimisation trajectory in objective space for a single optimisation run of 100 iterations x 8 points per batch. a-b) ZDT1. c-d) ZDT2. e-f) ZDT3. g-h) MW7. The red line represents the true PF, while MW7 being a constrained problem has an additional blue line to show the unconstrained PF. The colour of each

experiment refers to the number of iterations. All problems clearly show a more gradual evolution of results as the number of iterations progress in U-NSGA-III whereas qNEHVI rapidly approaches PF and then fails to converge further.

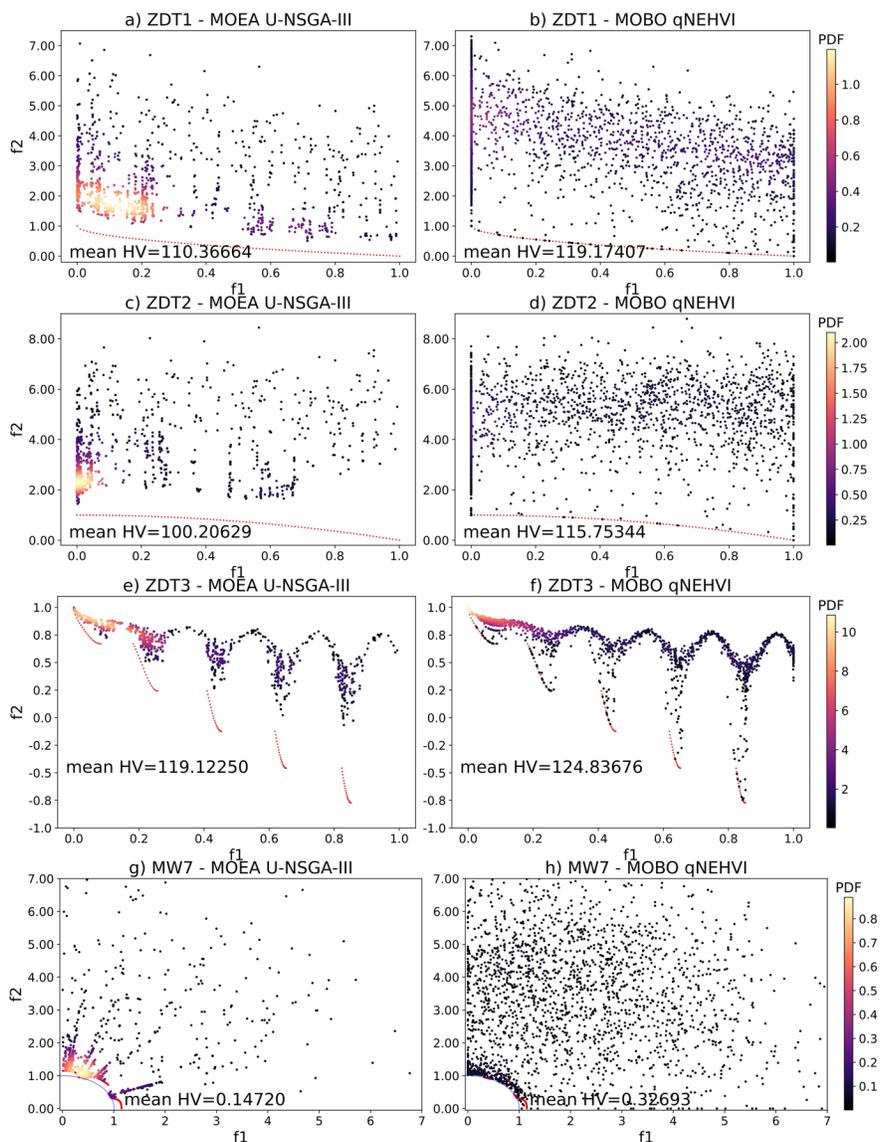


Figure 5: Probability density maps in objective space for 10 runs of 24 iterations x 8 points per batch. a-b) ZDT1. c-d) ZDT2. e-f) ZDT3. g-h) MW7. The evaluated data points are plotted with a Gaussian kernel density estimate using SciPy to illustrate the distribution of points across objective space. The colour bar represents the numerical value of probability density. Results are averaged over the 10 runs and highlight the lower diversity of points and consistency in optimisation trajectory for qNEHVI compared to U-NSGA-III.

The choice of batch size is another important parameter to consider for materials scientists. It can be tuned when attempting to scale up for HTE. A larger batch size is usually ideal since it provides higher throughput, and thus more time savings since lesser iterations are required. However, batch size affects the performance of optimisation strategies, potentially reducing the number of iterations needed in a run. We thus perform optimisation on the same synthetic problems for different batch sizes, keeping dimensionality at $\text{dim}=8$ and

with the same evaluation budget of 192 points and 10 runs as mentioned earlier.

The authors of qNEHVI hypothesised that it operates better at small batch sizes by providing a smoother gradient descent in sequential optimisation. Results reported in **Figure 6** a), b) and d) for ZDT1, ZDT2 and MW7, respectively, support this hypothesis, and we clearly observe that the lowest batch size setting of 2, as represented by the pink line, has the best performance overall. Interestingly, this is also the case for U-NSGA-III where the lowest batch size of 2 tends to give better HV for ZDT1-3 as seen by the blue line. This is also empirically shown in literature where, given a total budget, higher populations may impede convergence as it effectively limits the number of iterations. (Wang et al., 2019; Hort and Sarro, 2021; Tanabe and Oyama, 2017)

It is suggested that the same did not apply for MW7 since the disconnected PF was often not fully explored due to differences in initialisation and how the heuristic search operated, which we discuss previously for Figure 4 and 5. Instead, a larger batch size i.e. larger population is beneficial in maintaining solutions across disconnected regions of objective space, as seen by the red line in Figure 6d). We also explain why this did not apply to ZDT3: since the initial sampling was generally able to cover the search space well, there are relatively little ‘lost’ regions as seen from Figure 4c). Additionally, we provide optimisation trajectory plots for U-NSGA-III at different batch sizes in the SI 3 to illustrate this.

Furthermore, we also observe that qNEHVI has greater variance in log HV difference, compared to U-NSGA-III. This further reinforces our hypothesis that the performance of qNEHVI is in part due to the stochastic QMC sampling, whilst the heuristic nature of U-NSGA-III means that the evolution of solutions is more consistent.

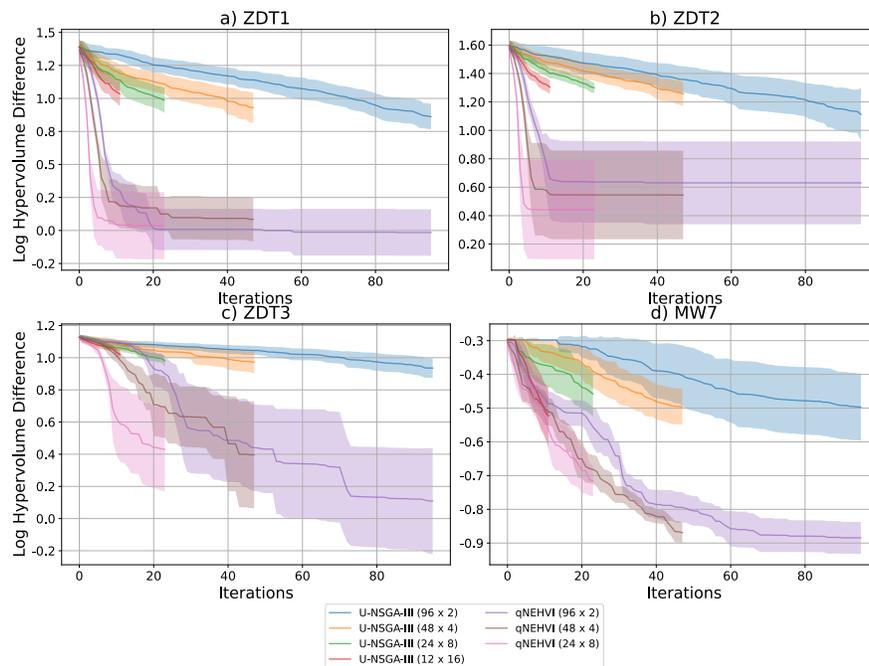


Figure 6: Convergence at different batch sizes with the same total evaluation budget of 24×8 . a) ZDT1. b) ZDT2. c) ZDT3. d) MW7. We omitted qNEHVI for batch of 16 due to prohibitively high computation cost when scaling up. Plots are taken with mean and 95% confidence interval of $\log_{10}(\text{HV}_{\text{max}} - \text{HV}_{\text{current}})$, with HV_{max} being computed from known PF in pymoo. We follow the same details as for Figure 5. Results suggest that qNEHVI works better with low batching on disconnected PF.

Real-world Problems

Based on results reported in Figures 4, 5, 6, we formulate the hypothesis that qNEHVI as a MOBO strategy is very sample efficient, i.e. able to arrive at the PF rapidly with few evaluations and is superior in maximizing hypervolume as a performance metric. In comparison, we found that U-NSGA-III provides a more consistent search due to its heuristic evolution nature over that of stochastic QMC sampling in qNEHVI, and furthermore maintain a larger pool of near-Pareto samples that is not reflected by the HV performance metric. We also report that smaller batch sizes are generally better in both strategies over the two-objective jobs used.

To test this hypothesis, we repeated our experiments on real-world multi-objective datasets. (Yeh, 2008; MacLeod et al., 2022) An unavoidable issue of empirically benchmarking optimisation strategies on real-world problems is that some surrogate model must be used in-lieu of a black-box where new data is experimentally validated. Alternatively, a candidate selection problem can be used where optimisation is limited to only proposing new candidates from a pre-labelled dataset until eventually the ‘pool’ of samples is exhausted. (Janet et al., 2020; Hanaoka, 2022; Gopakumar et al., 2018; Liang et al., 2021) The benefit of this method over surrogate-based methods is that only real data from the black-box is used, rather than data extrapolated from a model approximating its behaviour. However, the candidate selection approach assumes that the existing dataset contains all data points necessary to perfectly represent the search space and true PF. It is generally not possible to prove that this is the case, unless the exact function mapping input to output of the black box is known, or the dataset contains all possible combination of input/output pairs and is therefore a complete representation of the problem like that of inverse design.

Here, due to the relatively small size of the datasets ($\sim 10^2$ data points), the candidate selection method was not implemented. Instead, we relied on training an appropriate regressor to model the dataset. The two real-world benchmarks used in this paper are presented in **Table 2**. Materials datasets with constraints are hard to find from available HTE literature, besides from simple combinatorial setups that need to sum to 100%. (Erps et al., 2021) Another example is Cao L. et al, (Cao et al., 2021) which included complex constraints in the form of solubility, although we were unable to attain their full dataset and solubility classifier. Similar to synthetic benchmark experiments, we compare both approaches based on 3 metrics,

Name	Problem	Model	n- var	n- obj	n- con- str	ref_pt
Thin film	Minimize process temperature and maximize conductivity of spray coated palladium films	GP regressor	4	2	0	[1.02, -0.05]
Concrete Slump	Maximize slump and compressive strength in concrete formulations	Neural network ensemble	7			[0, 0]

Table 2: List and details of real-world problems.

omitting dimensional contour plots:

1. Optimisation trajectory in objective space for 100 iterations x 8 points per batch
2. Probability density function in objective space for 24 iterations x 8 points per batch
3. Comparison of batch size for log hypervolume difference

Figure 7 further supports our conclusions drawn from results reported in Figure 4. As seen in Figure 7 b) and d), qNEHVI is highly sample efficient, with points at or near the PF within the first 20 iterations or so, indicated by the darker points lying on the red line. However, qNEHVI shows a large random distribution of non-optimal points away from PF across the entire optimisation as seen by both dark and bright points, which we attribute to the stochastic QMC sampling. U-NSGA-III performs a gradual evolution of points towards the PF as seen in Figure 7 a) and c), as well as maintaining a large pool of near-optimal solutions.

This is reflected by the lower HV scores for U-NSGA-III compared to those of qNEHVI.

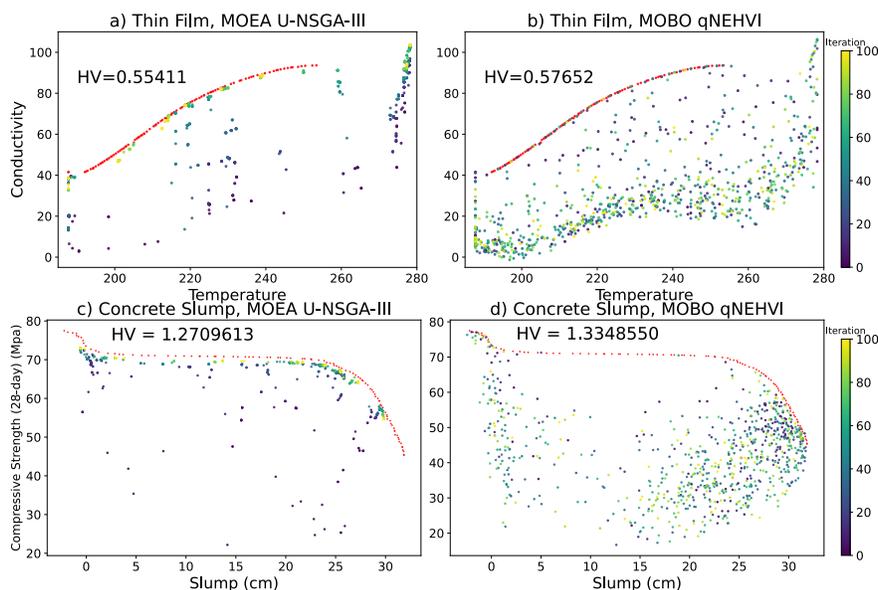


Figure 7: Optimisation trajectory in objective space for a single optimisation run of 100 iterations \times 8 points per batch. a-b) Thin Film. c-d) Concrete Slump. across objective space for a single run of 100 iterations \times 8 points per batch. The red line represents the PF. PFs for real-world datasets were virtually generated using NSGA-II for 500 generations with population size of 100. The colour of each experiment refers to the number of iterations. The results here corroborate the ‘wastage’ of solutions in qNEHVI, although which algorithm is superior appears to be problem dependent.

At a smaller evaluation budget, we observe that U-NSGA-III consistently maintains a large pool of near-optimal solutions, as the bright region is seen nearer to the PF, while reporting a lower mean HV compared to qNEHVI in **Figure 8** a) and e). **Figure 8** b) for the Thin Film problem also corroborates our findings that qNEHVI proposes many non-optimal solutions, as seen by the bright region away from PF, which indicates a higher probability of occurrence.

Interestingly, in **Figure 8** d) for Concrete Slump problem, we observe that qNEHVI is consistently converging to a specific region in objective space, while the U-NSGA-III search follows that of **Figure 8** b) with concentration of solutions at the near-optimal region close to PF. We hypothesize that qNEHVI’s performance for this problem is influenced by how the underlying GP surrogate model learns the function and strongly biases solutions to that specific region. We show further proof in SI 2, where we illustrate the expected PF given by the GP surrogate model.

In contrast, both problems here indicated that U-NSGA-III benefited more from larger batch sizes, as seen by the green line, which is different from what we observed in **Figure 6** for synthetic problems. Our hypothesis is that the modelled datasets present a more mathematically difficult optimisation problem, with various ‘obstacles’ that inhibit the evolution of solutions towards the PF. We support this by referring to our discussions for **Figure 7** c) and d) on Concrete Slump regarding local optima, as well as observing a notable blank region of objective space which U-NSGA-III fails to flesh out in **Figure 7** a) for Thin Film problem. Overall, results reported here suggest that given state-of-the-art implementations in HT experiments, a small batch-size with MOBO is the right strategy to converge rapidly.

Finally, we also studied the effect of batch size on convergence in **Figure 9**. Results present both similarities and differences with what we observe for synthetic benchmarks as in **Figure 6**. A lower batch size in qNEHVI

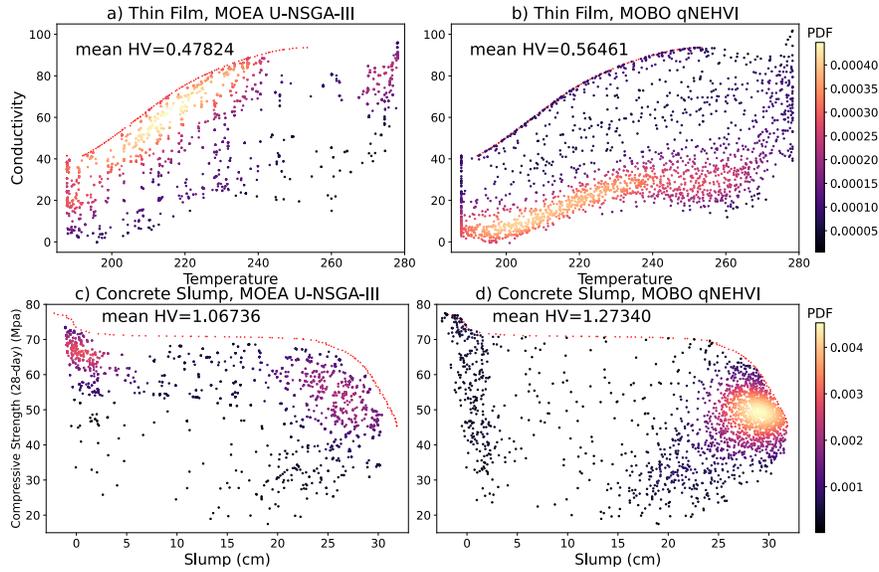


Figure 8: Probability density maps in objective space for 10 optimisation runs of 24 iterations x 8 points per batch. a-b) Thin Film. c-d) Concrete Slump. The evaluated data points are plotted with a Gaussian kernel density estimate using SciPy to illustrate the distribution of points across objective space, with a colour bar to represent the numerical value of probability density. Results are averaged over 10 runs, taking a smaller evaluation budget of 24 iterations x 8 points = 192. The results here reinforce the finding that qNEHVI has a more random distribution of points, but still outperforms U-NSGA-III for a low evaluation budget.

was better for both problems, as seen by the purple line, which is consistent with our findings for Figure 6.

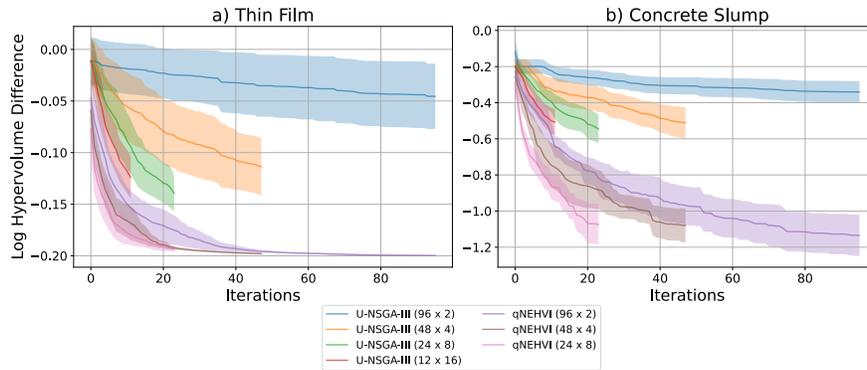


Figure 9: Convergence at different batch sizes with the same total evaluation budget of 24 x 8. a) Thin Film. b) Concrete Slump. We omitted qNEHVI for batch of 16 due to prohibitively high computation cost when scaling up. Plots are taken with mean and 95% confidence interval of $\log_{10}(HV_{\max} - HV_{\text{current}})$, with HV_{\max} being computed from known PF in pymoo. The results shown here support our conclusions for qNEHVI in Fig. 6 but have marked differences for U-NSGA-III.

Further discussion

We have compared qNEHVI and U-NSGA-III using both synthetic and real-world benchmarks, considering different experimental parameters such as dimensionality and batch size which materials scientists may face when implementing closed loop optimisation in HTE. Our results suggest that qNEHVI is extremely sample efficient in arriving at the PF to maximise HV gain but fails to exploit it. In contrast, we report that U-NSGA-III has a consistent optimisation trajectory, and better exploits the PF while maintaining more near-optimal solutions.

We thus make the case for MOEAs for materials experimentation besides computational design. We also argue that such implementations would be best when the objective space is mildly discontinuous (which can be the case for structural problems such as alloys) since small changes in inputs can cause the outputs to vary wildly in objective space, and an evolutionary-based strategy can navigate with better resolution. This is consistent with work by Liang Q. et al (Liang et al., 2021) on single-objective optimisation, which noted that having “multiple well-performing candidates allows one to not only observe regions in design space that frequently yield high-performing samples but also have backup options for further evaluation should the most optimal candidate fail in subsequent evaluations”.

Furthermore, MOEAs also scale better in terms of computational cost for a high dimensional and high throughput context, where they have the means to converge while maintaining both diversity and feasibility. HV-based MOBOs such as qNEHVI scale poorly to high dimensionality and many-objective problems due to the cost of computing HV. Depending on the HTE set-up, the ML component may not be able to leverage on powerful cluster computing for computationally intensive problems/models. MOEAs with lower computation overhead such as U-NSGA-III would be a better choice in such scenarios. With advancements in HTE set ups allowing for automation and parallel sampling, we expect research groups to leverage on higher throughput systems with short turnarounds. This makes the implementation of MOEAs much more practical to explore complex search spaces when paired with larger evaluation budgets of 10^3 to 10^4 data points.

The choice of batch size to balance optimisation performance while minimising experimental cycles is also important. Empirically, our results obtained suggest that a smaller batch size of around 4 is ideal for the limited evaluation budget of 192 points, although larger batch sizes are preferred for more complex problems (with added difficulty from disconnected regions in objective space, or perhaps presence of local optima).

A caveat of our work here is that the synthetic problems we chose are a generalisation of bi-objective spaces with specific Pareto geometry that may not translate well for real-life experimentation especially for many-objective ($M > 3$) problems. Newer benchmarks with higher difficulties and complex geometries/PFs (Fan et al., 2020) are tailored towards challenging MOEAs with massive evaluation budgets of up to 10^7 total observations. An example would be MW5 from the MW test suite, which has a narrow tunnel-like feasible regions that are practically impossible for GPs to model, resulting in MOBOs failing to converge. Indeed, R. W. Epps et al noted that it is “difficult to impose complex structure on the GPs, which often simply encode continuity, smoothness, or periodicity”. (Epps et al., 2020) We refer to other publications which study the differences between surrogate models in BO, (Liang et al., 2021; Lim et al., 2021; Yan et al., 2021) as well as AI techniques that scale MOBOs to higher dimensional spaces. (Moriconi et al., 2020; Eriksson and Jankowiak, 2021)

Furthermore, materials experimentation is usually afflicted with real-world imperfections and deviations during synthesis, or uncertainty due to characterization equipment resolution. For example, MacLeod et al noted that “the tendency of drop-casted samples to exhibit a wide range of downwards deviations in the apparent conductivity due to the poor sample morphology”. (MacLeod et al., 2022) The effect of noise causes deviations in objective values from the ‘true’ ground truth, and although unclear, is an unavoidable aspect of optimisation which should be tackled. (Koch et al., 2015; Horn et al., 2017) In SI 6, we perform a comparison of qNEHVI and U-NSGA-III on varying amounts of white noise on outputs.

Conclusion

In conclusion, our results illustrate that existing performance metrics such as HV do not really reflect the goal of fleshing out the PF region, where HV-based methods like qNEHVI may not achieve satisfactorily. This reflects an aspect of optimisation which might be neglected in the purview of multi-objective materials discovery: which is to find a diverse set of optimal solutions that can adequately convey the trade-offs between conflicting objectives. We thus present alternative illustrative means such as probability density maps to better benchmark the performance of optimisation strategies for such purposes. Moving ahead, we hope that this can spur further improvement for MOBOs as well as a stronger consideration for the use of MOEAs for materials problems due to its heuristic nature in exploiting the PF.

Acknowledgements

K.H. acknowledges funding from the Accelerated Materials Development for Manufacturing Program at A*STAR via the AME Programmatic Fund by the Agency for Science, Technology and Research under Grant No. A1898b0043. K.H. also acknowledges funding from the NRF Fellowship NRF-NRFF13-2021-0011.

K.Y.A.L. and K.H. conceived of the research. K.Y.A.L. working with E.V-G. and Y-F.L. developed and tested the algorithms and datasets, with key intellectual contributions from all authors. K.Y.A.L. wrote the manuscript, with input from all co-authors.

Conflict of interest

K.H. owns equity in a startup focused on applying Machine Learning for Materials.

Supporting Information

Source code for our work can be found in <https://github.com/Kedar-Materials-by-Design-Lab>. A supplementary document is also available.

References

- Kameel Abdel-Latif, Robert W. Epps, Fazel Bateni, Suyong Han, Kristofer G. Reyes, and Milad Abolhasani. Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow. *Advanced Intelligent Systems*, 3(2):2170022, feb 2021. doi: 10.1002/aisy.202170022. URL <https://doi.org/10.1002%2Faisy.202170022>.
- Nourin Alsharif, Joshua R. Uzarski, Timothy J. Lawton, and Keith A. Brown. High-Throughput Multiobjective Optimization of Patterned Multifunctional Surfaces. *ACS Applied Materials & Interfaces*, 12(28):32069–32077, jun 2020. doi: 10.1021/acsami.0c04202. URL <https://doi.org/10.1021%2Facsami.0c04202>.
- Vahé Asvatourian, Philippe Leray, Stefan Michiels, and Emilie Lanoy. Integrating expert’s knowledge constraint of time dependent exposures in structure learning for Bayesian networks. *Artificial Intelligence in*

- Medicine*, 107:101874, jul 2020. doi: 10.1016/j.artmed.2020.101874. URL <https://doi.org/10.1016%2Fj.artmed.2020.101874>.
- Anne Auger, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications. *Theoretical Computer Science*, 425:75–103, mar 2012. doi: 10.1016/j.tcs.2011.03.012. URL <https://doi.org/10.1016%2Fj.tcs.2011.03.012>.
- Patrick Avery, Zackary Falls, and Eva Zurek. XtalOpt Version r10: An open-source evolutionary algorithm for crystal structure prediction. *Computer Physics Communications*, 217:210–211, aug 2017. doi: 10.1016/j.cpc.2017.04.001. URL <https://doi.org/10.1016%2Fj.cpc.2017.04.001>.
- Daniil Bash, Yongqiang Cai, Chellappan Vijila, Swee Liang Wong, Yang Xu, Pawan Kumar, Jin Da Tan, Anas Abutaha, Jayce Jian Wei Cheng, Yee Fun Lim, Isaac Parker Siyu Tian, Zekun Ren, Wai Kuan Wong, Flore Mekki-Berrada, Jatin Nitin Kumar, Saif Khan, Qianxiao Li, Tonio Buonassisi, and Kedar Hippalgaonkar. Machine Learning and High-Throughput Robust Design of P3HT-CNT Composite Thin Films for High Electrical Conductivity. nov 2020. doi: 10.26434/chemrxiv.13265288. URL <https://doi.org/10.26434%2Fchemrxiv.13265288>.
- Daniil Bash, Frederick Hubert Chenardy, Zekun Ren, Jayce J Cheng, Tonio Buonassisi, Ricardo Oliveira, Jatin N Kumar, and Kedar Hippalgaonkar. Accelerated automated screening of viscous graphene suspensions with various surfactants for optimal electrical conductivity. *Digital Discovery*, 1(2):139–146, 2022. doi: 10.1039/d1dd00008j. URL <https://doi.org/10.1039%2Fd1dd00008j>.
- Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Max-value entropy search for multi-objective Bayesian optimization with constraints. *arXiv preprint arXiv:2009.01721*, 2020.
- Enrico Berardo, Lukas Turcani, Marcin Miklitz, and Kim E. Jelfs. An evolutionary algorithm for the discovery of porous organic cages. *Chemical Science*, 9(45):8513–8527, 2018. doi: 10.1039/c8sc03560a. URL <https://doi.org/10.1039%2Fc8sc03560a>.
- Eric Bradford, Artur M. Schweidtmann, and Alexei Lapkin. Efficient multiobjective optimization employing Gaussian processes spectral sampling and a genetic algorithm. *Journal of Global Optimization*, 71(2):407–438, feb 2018. doi: 10.1007/s10898-018-0609-2. URL <https://doi.org/10.1007%2Fs10898-018-0609-2>.
- Benjamin Burger, Phillip M. Maffettone, Vladimir V. Gusev, Catherine M. Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M. Alston, Buyi Li, Rob Clowes, Nicola Rankin, Brandon Harris, Reiner Sebastian Sprick, and Andrew I. Cooper. A mobile robotic chemist. *Nature*, 583(7815):237–241, jul 2020. doi: 10.1038/s41586-020-2442-2. URL <https://doi.org/10.1038%2Fs41586-020-2442-2>.
- Liwei Cao, Danilo Russo, Kobi Felton, Daniel Salley, Abhishek Sharma, Graham Keenan, Werner Mauer, Huanhuan Gao, Leroy Cronin, and Alexei A. Lapkin. Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE. *Cell Reports Physical Science*, 2(1):100295, jan 2021. doi: 10.1016/j.xcrp.2020.100295. URL <https://doi.org/10.1016%2Fj.xcrp.2020.100295>.
- Rodrigo P. Carvalho, Cleber F. N. Marchiori, Daniel Brandell, and C. Moyses Araujo. Tuning the Electrochemical Properties of Organic Battery Cathode Materials: Insights from Evolutionary Algorithm DFT Calculations. *ChemSusChem*, 13(9):2402–2409, mar 2020. doi: 10.1002/cssc.201903450. URL <https://doi.org/10.1002%2Fc8sc.201903450>.
- Melodie Christensen, Lars P. E. Yunker, Folarin Adedeji, Florian Häse, Loïc M. Roch, Tobias Gensch, Gabriel dos Passos Gomes, Tara Zepel, Matthew S. Sigman, Alán Aspuru-Guzik, and Jason E. Hein. Data-science driven autonomous process optimization. *Communications Chemistry*, 4(1), aug 2021. doi: 10.1038/s42004-021-00550-x. URL <https://doi.org/10.1038%2Fs42004-021-00550-x>.
- Carlos A. Coello Coello and Ricardo Landa Becerra. Evolutionary Multiobjective Optimization in Materials Science and Engineering. *Materials and Manufacturing Processes*, 24(2):119–129, jan 2009. doi: 10.1080/10426910802609110. URL <https://doi.org/10.1080%2F10426910802609110>.

- Juan-Pablo Correa-Baena, Kedar Hippalgaonkar, Jeroen van Duren, Shaffiq Jaffer, Vijay R. Chandrasekhar, Vladan Stevanovic, Cyrus Wadia, Supratik Guha, and Tonio Buonassisi. Accelerating Materials Development via Automation Machine Learning, and High-Performance Computing. *Joule*, 2(8):1410–1420, aug 2018. doi: 10.1016/j.joule.2018.05.009. URL <https://doi.org/10.1016%2Fj.joule.2018.05.009>.
- Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. *Advances in Neural Information Processing Systems*, 33:9851–9864, 2020.
- Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement. *Advances in Neural Information Processing Systems*, 34:2187–2200, 2021.
- Adarsh Dave, Jared Mitchell, Kirthevasan Kandasamy, Han Wang, Sven Burke, Biswajit Paria, Barnabás Póczos, Jay Whitacre, and Venkatasubramanian Viswanathan. Autonomous Discovery of Battery Electrolytes with Robotic Experimentation and Machine Learning. *Cell Reports Physical Science*, 1(12):100264, dec 2020. doi: 10.1016/j.xcrp.2020.100264. URL <https://doi.org/10.1016%2Fj.xcrp.2020.100264>.
- Kalyanmoy Deb. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction. In *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, pages 3–34. Springer London, 2011. doi: 10.1007/978-0-85729-652-8_1. URL https://doi.org/10.1007%2F978-0-85729-652-8_1.
- James R. Deneault, Jorge Chang, Jay Myung, Daylond Hooper, Andrew Armstrong, Mark Pitt, and Benji Maruyama. Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer. *MRS Bulletin*, 46(7):566–575, apr 2021. doi: 10.1557/s43577-021-00051-1. URL <https://doi.org/10.1557%2Fs43577-021-00051-1>.
- Robert W. Epps, Michael S. Bowen, Amanda A. Volk, Kameel Abdel-Latif, Suyong Han, Kristofer G. Reyes, Aram Amassian, and Milad Abolhasani. Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot. *Advanced Materials*, 32(30):2001626, jun 2020. doi: 10.1002/adma.202001626. URL <https://doi.org/10.1002%2Fadma.202001626>.
- David Eriksson and Martin Jankowiak. High-dimensional Bayesian optimization with sparse axis-aligned subspaces. In *Uncertainty in Artificial Intelligence*, pages 493–503. PMLR, 2021.
- Timothy Erps, Michael Foshey, Mina Konaković Luković, Wan Shou, Hanns Hagen Goetzke, Herve Dietsch, Klaus Stoll, Bernhard von Vacano, and Wojciech Matusik. Accelerated discovery of 3D printing materials using data-driven multiobjective optimization. *Science Advances*, 7(42), oct 2021. doi: 10.1126/sciadv.abf7435. URL <https://doi.org/10.1126%2Fsciadv.abf7435>.
- Zhun Fan, Wenji Li, Xinye Cai, Han Huang, Yi Fang, Yugen You, Jiajie Mo, Caimin Wei, and Erik Goodman. An improved epsilon constraint-handling method in MOEA/D for CMOPs with large infeasible regions. *Soft Computing*, 23(23):12491–12510, feb 2019. doi: 10.1007/s00500-019-03794-x. URL <https://doi.org/10.1007%2Fs00500-019-03794-x>.
- Zhun Fan, Wenji Li, Xinye Cai, Hui Li, Caimin Wei, Qingfu Zhang, Kalyanmoy Deb, and Erik Goodman. Difficulty Adjustable and Scalable Constrained Multiobjective Test Problem Toolkit. *Evolutionary Computation*, 28(3):339–378, sep 2020. doi: 10.1162/evco.a.00259. URL https://doi.org/10.1162%2Fevco_a_00259.
- S. Ganguly, S. Datta, and N. Chakraborti. Genetic Algorithms in Optimization of Strength and Ductility of Low-Carbon Steels. *Materials and Manufacturing Processes*, 22(5):650–658, jun 2007. doi: 10.1080/10426910701323607. URL <https://doi.org/10.1080%2F10426910701323607>.
- Eduardo C. Garrido-Merchán and Daniel Hernández-Lobato. Predictive Entropy Search for Multi-objective Bayesian Optimization with Constraints. *Neurocomputing*, 361:50–68, oct 2019. doi: 10.1016/j.neucom.2019.06.025. URL <https://doi.org/10.1016%2Fj.neucom.2019.06.025>.

- Aldair E. Gongora, Bowen Xu, Wyatt Perry, Chika Okoye, Patrick Riley, Kristofer G. Reyes, Elise F. Morgan, and Keith A. Brown. A Bayesian experimental autonomous researcher for mechanical design. *Science Advances*, 6(15), apr 2020. doi: 10.1126/sciadv.aaz1708. URL <https://doi.org/10.1126%2Fsciadv.aaz1708>.
- Abhijith M. Gopakumar, Prasanna V. Balachandran, Dezhen Xue, James E. Gubernatis, and Turab Lookman. Multi-objective Optimization for Materials Discovery via Adaptive Design. *Scientific Reports*, 8(1), feb 2018. doi: 10.1038/s41598-018-21936-3. URL <https://doi.org/10.1038%2Fs41598-018-21936-3>.
- Jonathan Grizou, Laurie J. Points, Abhishek Sharma, and Leroy Cronin. A curious formulation robot enables the discovery of a novel protocell behavior. *Science Advances*, 6(5), jan 2020. doi: 10.1126/sciadv.aay4237. URL <https://doi.org/10.1126%2Fsciadv.aay4237>.
- Andreia P Guerreiro, Carlos M Fonseca, and Luís Paquete. The hypervolume indicator: Problems and algorithms. *arXiv preprint arXiv:2005.00515*, 2020.
- Kyohei Hanaoka. Bayesian optimization for goal-oriented multi-objective inverse material design. *iScience*, 24(7):102781, jul 2021. doi: 10.1016/j.isci.2021.102781. URL <https://doi.org/10.1016%2Fj.isci.2021.102781>.
- Kyohei Hanaoka. Comparison of conceptually different multi-objective Bayesian optimization methods for material design problems. *Materials Today Communications*, 31:103440, jun 2022. doi: 10.1016/j.mtcomm.2022.103440. URL <https://doi.org/10.1016%2Fj.mtcomm.2022.103440>.
- Daniel Horn, Melanie Dagge, Xudong Sun, and Bernd Bischl. First Investigations on Noisy Model-Based Multi-objective Optimization. In *Lecture Notes in Computer Science*, pages 298–313. Springer International Publishing, 2017. doi: 10.1007/978-3-319-54157-0_21. URL https://doi.org/10.1007%2F978-3-319-54157-0_21.
- Max Hort and Federica Sarro. The effect of offspring population size on NSGA-II. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, jul 2021. doi: 10.1145/3449726.3459479. URL <https://doi.org/10.1145%2F3449726.3459479>.
- Florian Häse, Loïc M. Roch, and Alán Aspuru-Guzik. Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chemical Science*, 9(39):7642–7655, 2018. doi: 10.1039/c8sc02239a. URL <https://doi.org/10.1039%2Fc8sc02239a>.
- Jon Paul Janet, Sahasrajit Ramesh, Chenru Duan, and Heather J. Kulik. Accurate Multiobjective Design in a Space of Millions of Transition Metal Complexes with Neural-Network-Driven Efficient Global Optimization. *ACS Central Science*, 6(4):513–524, mar 2020. doi: 10.1021/acscentsci.0c00026. URL <https://doi.org/10.1021%2Facscentsci.0c00026>.
- Paul C. Jennings, Steen Lysgaard, Jens Strabo Hummelshøj, Tejs Vegge, and Thomas Bligaard. Genetic algorithms for computational materials discovery accelerated by machine learning. *npj Computational Materials*, 5(1), apr 2019. doi: 10.1038/s41524-019-0181-4. URL <https://doi.org/10.1038%2Fs41524-019-0181-4>.
- Masayuki Karasuyama, Hiroki Kasugai, Tomoyuki Tamura, and Kazuki Shitara. Computational design of stable and highly ion-conductive materials using multi-objective bayesian optimization: Case studies on diffusion of oxygen and lithium. *Computational Materials Science*, 184:109927, nov 2020. doi: 10.1016/j.commatsci.2020.109927. URL <https://doi.org/10.1016%2Fj.commatsci.2020.109927>.
- Aaron Klein, Simon Bartels, Stefan Falkner, Philipp Hennig, and Frank Hutter. Towards efficient Bayesian optimization for big data. In *NIPS 2015 Bayesian Optimization Workshop*, 2015.
- Patrick Koch, Tobias Wagner, Michael T.M. Emmerich, Thomas Bäck, and Wolfgang Konen. Efficient multi-criteria optimization on noisy machine learning problems. *Applied Soft Computing*, 29:357–370, apr 2015. doi: 10.1016/j.asoc.2015.01.005. URL <https://doi.org/10.1016%2Fj.asoc.2015.01.005>.

- Stefan Langner, Florian Häse, José Darío Perea, Tobias Stubhan, Jens Hauch, Loïc M. Roch, Thomas Heumueller, Alán Aspuru-Guzik, and Christoph J. Brabec. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Advanced Materials*, 32(14):1907801, apr 2020. doi: 10.1002/adma.201907801. URL <https://doi.org/10.1002%2Fadma.201907801>.
- Bingdong Li, Jinlong Li, Ke Tang, and Xin Yao. Many-objective evolutionary algorithms: A survey. *ACM Computing Surveys (CSUR)*, 48(1):1–35, 2015.
- Jiagen Li, Junzi Li, Rulin Liu, Yuxiao Tu, Yiwen Li, Jiaji Cheng, Tingchao He, and Xi Zhu. Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nature Communications*, 11(1), apr 2020. doi: 10.1038/s41467-020-15728-5. URL <https://doi.org/10.1038%2Fs41467-020-15728-5>.
- Zhiming Li, Konda Gokuldoss Pradeep, Yun Deng, Dierk Raabe, and Cemal Cem Tasan. Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off. *Nature*, 534(7606):227–230, may 2016. doi: 10.1038/nature17981. URL <https://doi.org/10.1038%2Fnature17981>.
- Qiaohao Liang, Aldair E. Gongora, Zekun Ren, Armi Tiihonen, Zhe Liu, Shijing Sun, James R. Deneault, Daniil Bash, Flore Mekki-Berrada, Saif A. Khan, Kedar Hippalgaonkar, Benji Maruyama, Keith A. Brown, John Fisher III, and Tonio Buonassisi. Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. *npj Computational Materials*, 7(1), nov 2021. doi: 10.1038/s41524-021-00656-9. URL <https://doi.org/10.1038%2Fs41524-021-00656-9>.
- Yee-Fun Lim, Chee Koon Ng, U.S. Vaiteswar, and Kedar Hippalgaonkar. Extrapolative Bayesian Optimization with Gaussian Process and Neural Network Ensemble Surrogate Models. *Advanced Intelligent Systems*, 3(11):2100101, sep 2021. doi: 10.1002/aisy.202100101. URL <https://doi.org/10.1002%2Faisy.202100101>.
- Yue Liu, Tianlu Zhao, Wangwei Ju, and Siqi Shi. Materials discovery and design using machine learning. *Journal of Materiomics*, 3(3):159–177, sep 2017. doi: 10.1016/j.jmat.2017.08.002. URL <https://doi.org/10.1016%2Fj.jmat.2017.08.002>.
- Zhe Liu, Nicholas Rolston, Austin C. Flick, Thomas W. Colburn, Zekun Ren, Reinhold H. Dauskardt, and Tonio Buonassisi. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. *Joule*, 6(4):834–849, apr 2022. doi: 10.1016/j.joule.2022.03.003. URL <https://doi.org/10.1016%2Fj.joule.2022.03.003>.
- T. Lookman, P. V. Balachandran, D. Xue, G. Pilania, T. Shearman, J. Theiler, J. E. Gubernatis, J. Hogden, K. Barros, E. BenNaim, and F. J. Alexander. A Perspective on Materials Informatics: State-of-the-Art and Challenges. In *Information Science for Materials Discovery and Design*, pages 3–12. Springer International Publishing, dec 2015. doi: 10.1007/978-3-319-23871-5_1. URL https://doi.org/10.1007%2F978-3-319-23871-5_1.
- Zhongwei Ma and Yong Wang. Evolutionary Constrained Multiobjective Optimization: Test Suite Construction and Performance Comparisons. *IEEE Transactions on Evolutionary Computation*, 23(6):972–986, dec 2019. doi: 10.1109/tevc.2019.2896967. URL <https://doi.org/10.1109%2Ftevc.2019.2896967>.
- B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette. Self-driving laboratory for accelerated discovery of thin-film materials. *Science Advances*, 6(20), may 2020. doi: 10.1126/sciadv.aaz8867. URL <https://doi.org/10.1126%2Fsciadv.aaz8867>.
- Benjamin P. MacLeod, Fraser G. L. Parlane, Connor C. Rupnow, Kevan E. Dettelbach, Michael S. Elliott, Thomas D. Morrissey, Ted H. Haley, Oleksii Proskurin, Michael B. Rooney, Nina Taherimakhosousi, David J. Dvorak, Hsi N. Chiu, Christopher E. B. Waizenegger, Karry Ocean, Mehrdad

- Mokhtari, and Curtis P. Berlinguette. A self-driving laboratory advances the Pareto front for material properties. *Nature Communications*, 13(1), feb 2022. doi: 10.1038/s41467-022-28580-6. URL <https://doi.org/10.1038/s41467-022-28580-6>.
- M. Mahfouf, M. Jamei, and D. A. Linkens. Optimal Design of Alloy Steels Using Multiobjective Genetic Algorithms. *Materials and Manufacturing Processes*, 20(3):553–567, may 2005. doi: 10.1081/amp-200053580. URL <https://doi.org/10.1081/amp-200053580>.
- Arun Mannodi-Kanakithodi, Ghanshyam Pilania, Rampi Ramprasad, Turab Lookman, and James E. Gubernatis. Multi-objective optimization techniques to design the Pareto front of organic dielectric polymers. *Computational Materials Science*, 125:92–99, dec 2016. doi: 10.1016/j.commatsci.2016.08.018. URL <https://doi.org/10.1016/j.commatsci.2016.08.018>.
- Flore Mekki-Berrada, Zekun Ren, Tan Huang, Wai Kuan Wong, Fang Zheng, Jiaxun Xie, Isaac Parker Siyu Tian, Senthilnath Jayavelu, Zackaria Mahfoud, Daniil Bash, Kedar Hippalgaonkar, Saif Khan, Tonio Buonassisi, Qianxiao Li, and Xiaonan Wang. Two-Step Machine Learning Enables Optimized Nanoparticle Synthesis. jul 2020. doi: 10.26434/chemrxiv.12673742. URL <https://doi.org/10.26434/chemrxiv.12673742>.
- Steven M. Mennen, Carolina Alhambra, C. Liana Allen, Mario Barberis, Simon Berritt, Thomas A. Brandt, Andrew D. Campbell, Jesús Castañón, Alan H. Cherney, Melodie Christensen, David B. Damon, J. Eugenio de Diego, Susana García-Cerrada, Pablo García-Losada, Rubén Haro, Jacob Janey, David C. Leitch, Ling Li, Fangfang Liu, Paul C. Lobben, David W. C. MacMillan, Javier Magano, Emma McInturff, Sebastien Monfette, Ronald J. Post, Danielle Schultz, Barbara J. Sitter, Jason M. Stevens, Iulia I. Strambeanu, Jack Twilton, Ke Wang, and Matthew A. Zajac. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Organic Process Research & Development*, 23(6):1213–1242, may 2019. doi: 10.1021/acs.oprd.9b00140. URL <https://doi.org/10.1021/acs.oprd.9b00140>.
- Ederm Menou, Gérard Ramstein, Emmanuel Bertrand, and Franck Tancret. Multi-objective constrained design of nickel-base superalloys using data mining- and thermodynamics-driven genetic algorithms. *Modelling and Simulation in Materials Science and Engineering*, 24(5):055001, apr 2016. doi: 10.1088/0965-0393/24/5/055001. URL <https://doi.org/10.1088/0965-0393/24/5/055001>.
- Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.
- Riccardo Moriconi, Marc Peter Deisenroth, and K. S. Sesh Kumar. High-dimensional Bayesian optimization using low-dimensional feature spaces. *Machine Learning*, 109(9-10):1925–1943, sep 2020. doi: 10.1007/s10994-020-05899-z. URL <https://doi.org/10.1007/s10994-020-05899-z>.
- Radu Stefan Niculescu, Tom M Mitchell, R Bharat Rao, Kristin P Bennett, and Emilio Parrado-Hernández. Bayesian network learning with parameter constraints. *Journal of machine learning research*, 7(7), 2006.
- Maria Pakhnova, Ivan Kruglov, Alexey Yanilkin, and Artem R. Oganov. Search for stable cocrystals of energetic materials using the evolutionary algorithm USPEX. *Physical Chemistry Chemical Physics*, 22(29):16822–16830, 2020. doi: 10.1039/d0cp03042b. URL <https://doi.org/10.1039/d0cp03042b>.
- Tarak K. Patra, Venkatesh Meenakshisundaram, Jui-Hsiang Hung, and David S. Simmons. Neural-Network-Biased Genetic Algorithms for Materials Design: Evolutionary Algorithms That Learn. *ACS Combinatorial Science*, 19(2):96–107, jan 2017. doi: 10.1021/acscmbosci.6b00136. URL <https://doi.org/10.1021/acscmbosci.6b00136>.
- Ivan Ramirez, Martina Causa', Yufei Zhong, Natalie Banerji, and Moritz Riede. Key Tradeoffs Limiting the Performance of Organic Photovoltaics. *Advanced Energy Materials*, 8(28):1703551, may 2018. doi: 10.1002/aenm.201703551. URL <https://doi.org/10.1002/aenm.201703551>.

- Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.
- Shaoxuan Ren, Dorian Joulié, Danielle Salvatore, Kristian Torbensen, Min Wang, Marc Robert, and Curtis P. Berlinguette. Molecular electrocatalysts can mediate fast selective CO. 365(6451):367–369, jul 2019. doi: 10.1126/science.aax4608. URL <https://doi.org/10.1126%2Fscience.aax4608>.
- Ashish Sabharwal, Horst Samulowitz, and Gerald Tesauro. Selecting Near-Optimal Learners via Incremental Data Allocation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), mar 2016. doi: 10.1609/aaai.v30i1.10316. URL <https://doi.org/10.1609%2Faaai.v30i1.10316>.
- Daniel Salley, Graham Keenan, Jonathan Grizou, Abhishek Sharma, Sergio Martín, and Leroy Cronin. A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles. *Nature Communications*, 11(1), jun 2020. doi: 10.1038/s41467-020-16501-4. URL <https://doi.org/10.1038%2Fs41467-020-16501-4>.
- Artur M. Schweidtmann, Adam D. Clayton, Nicholas Holmes, Eric Bradford, Richard A. Bourne, and Alexei A. Lapkin. Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chemical Engineering Journal*, 352:277–282, nov 2018. doi: 10.1016/j.cej.2018.07.031. URL <https://doi.org/10.1016%2Fj.cej.2018.07.031>.
- Haitham Seada and Kalyanmoy Deb. A Unified Evolutionary Optimization Procedure for Single Multiple, and Many Objectives. *IEEE Transactions on Evolutionary Computation*, 20(3):358–369, jun 2016. doi: 10.1109/tevc.2015.2459718. URL <https://doi.org/10.1109%2Ftevc.2015.2459718>.
- Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1):148–175, jan 2016. doi: 10.1109/jproc.2015.2494218. URL <https://doi.org/10.1109%2Fjproc.2015.2494218>.
- Ryota Shimizu, Shigeru Kobayashi, Yuki Watanabe, Yasunobu Ando, and Taro Hitosugi. Autonomous materials synthesis by machine learning and robotics. *APL Materials*, 8(11):111110, nov 2020. doi: 10.1063/5.0020370. URL <https://doi.org/10.1063%2F5.0020370>.
- Alexandros Solomou, Guang Zhao, Shahin Boluki, Jobin K. Joy, Xiaoning Qian, Ibrahim Karaman, Raymundo Arróyave, and Dimitris C. Lagoudas. Multi-objective Bayesian materials discovery: Application on the discovery of precipitation strengthened NiTi shape memory alloys through micromechanical modeling. *Materials & Design*, 160:810–827, dec 2018. doi: 10.1016/j.matdes.2018.10.014. URL <https://doi.org/10.1016%2Fj.matdes.2018.10.014>.
- Shijing Sun, Noor T. P. Hartono, Zekun D. Ren, Felipe Oviedo, Antonio M. Buscemi, Mariya Layurova, De Xin Chen, Tofunmi Ogunfunmi, Janak Thapa, Savitha Ramasamy, Charles Settens, Brian L. DeCost, Aaron Gilad Kusne, Zhe Liu, Siyu I. P. Tian, I. Marius Peters, Juan-Pablo Correa-Baena, and Tonio Buonassisi. Accelerating Photovoltaic Materials Development via High-Throughput Experiments and Machine-Learning-Assisted Diagnosis. 2018. URL <https://arxiv.org/abs/1812.01025v1>. Accessed on Thu, December 08, 2022.
- Shijing Sun, Noor T.P. Hartono, Zekun D. Ren, Felipe Oviedo, Antonio M. Buscemi, Mariya Layurova, De Xin Chen, Tofunmi Ogunfunmi, Janak Thapa, Savitha Ramasamy, Charles Settens, Brian L. DeCost, Aaron G. Kusne, Zhe Liu, Siyu I.P. Tian, Ian Marius Peters, Juan-Pablo Correa-Baena, and Tonio Buonassisi. Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis. *Joule*, 3(6):1437–1451, jun 2019. doi: 10.1016/j.joule.2019.05.014. URL <https://doi.org/10.1016%2Fj.joule.2019.05.014>.
- Shinya Suzuki, Shion Takeno, Tomoyuki Tamura, Kazuki Shitara, and Masayuki Karasuyama. Multi-objective Bayesian optimization using pareto-frontier entropy. In *International Conference on Machine Learning*, pages 9279–9288. PMLR, 2020.

- Ryoji Tanabe and Akira Oyama. The Impact of Population Size Number of Children, and Number of Reference Points on the Performance of NSGA-III. In *Lecture Notes in Computer Science*, pages 606–621. Springer International Publishing, 2017. doi: 10.1007/978-3-319-54157-0_41. URL https://doi.org/10.1007%2F978-3-319-54157-0_41.
- Ye Tian, Yajie Zhang, Yansen Su, Xingyi Zhang, Kay Chen Tan, and Yaochu Jin. Balancing Objective Optimization and Constraint Satisfaction in Constrained Evolutionary Multiobjective Optimization. *IEEE Transactions on Cybernetics*, 52(9):9559–9572, sep 2022. doi: 10.1109/teyb.2020.3021138. URL <https://doi.org/10.1109%2Fteyb.2020.3021138>.
- Lu Wang, Lanja R. Karadaghi, Richard L. Brutchey, and Noah Malmstadt. Self-optimizing parallel millifluidic reactor for scaling nanoparticle synthesis. *Chemical Communications*, 56(26):3745–3748, 2020. doi: 10.1039/d0cc00064g. URL <https://doi.org/10.1039%2Fd0cc00064g>.
- Qi Wang, Libing Wang, Wen Huang, Zhihong Wang, Shuming Liu, and Dragan A. Savić. Parameterization of NSGA-II for the Optimal Design of Water Distribution Systems. *Water*, 11(5):971, may 2019. doi: 10.3390/w11050971. URL <https://doi.org/10.3390%2Fw11050971>.
- Shuai Wu, Craig M. Hamel, Qiji Ze, Fengyuan Yang, H. Jerry Qi, and Ruike Zhao. Evolutionary Algorithm-Guided Voxel-Encoding Printing of Functional Hard-Magnetic Soft Active Materials. *Advanced Intelligent Systems*, 2(8):2000060, may 2020. doi: 10.1002/aisy.202000060. URL <https://doi.org/10.1002%2Faisy.202000060>.
- Bin Xu and Zonghao Zhang. Constrained Optimization Based on Ensemble Differential Evolution and Two-Level-Based Epsilon Method. *IEEE Access*, 8:213981–213997, 2020. doi: 10.1109/access.2020.3040647. URL <https://doi.org/10.1109%2Faccess.2020.3040647>.
- Yonggang Yan, Dan Lu, and Kun Wang. Accelerated discovery of single-phase refractory high entropy alloys assisted by machine learning. *Computational Materials Science*, 199:110723, nov 2021. doi: 10.1016/j.commatsci.2021.110723. URL <https://doi.org/10.1016%2Fj.commatsci.2021.110723>.
- I-Cheng Yeh. Modeling slump of concrete with fly ash and superplasticizer. *Computers and Concrete*, 5(6): 559–572, dec 2008. doi: 10.12989/cac.2008.5.6.559. URL <https://doi.org/10.12989%2Fcac.2008.5.6.559>.
- Wei Yong, Hongtao Zhang, Huadong Fu, Yaliang Zhu, Jie He, and Jianxin Xie. Improving prediction accuracy of high-performance materials via modified machine learning strategy. *Computational Materials Science*, 204:111181, mar 2022. doi: 10.1016/j.commatsci.2021.111181. URL <https://doi.org/10.1016%2Fj.commatsci.2021.111181>.
- Ruihao Yuan, Zhen Liu, Prasanna V. Balachandran, Deqing Xue, Yumei Zhou, Xiangdong Ding, Jun Sun, Dezhen Xue, and Turab Lookman. Accelerated Discovery of Large Electrostrains in BaTiO₃. 30(7):1702884, jan 2018. doi: 10.1002/adma.201702884. URL <https://doi.org/10.1002%2Fadma.201702884>.
- Guiying Zhang and David E. Block. Using highly efficient nonlinear experimental design methods for optimization of *Lactococcus lactis* fermentation in chemically defined media. *Biotechnology Progress*, pages NA–NA, 2009. doi: 10.1002/btpr.277. URL <https://doi.org/10.1002%2Fbtpr.277>.
- Hongtao Zhang, Huadong Fu, Shuaicheng Zhu, Wei Yong, and Jianxin Xie. Machine learning assisted composition effective design for precipitation strengthened copper alloys. *Acta Materialia*, 215:117118, aug 2021a. doi: 10.1016/j.actamat.2021.117118. URL <https://doi.org/10.1016%2Fj.actamat.2021.117118>.
- Peng Zhang, Yiyu Qian, and Quan Qian. Multi-objective optimization for materials design with improved NSGA-II. *Materials Today Communications*, 28:102709, sep 2021b. doi: 10.1016/j.mtcomm.2021.102709. URL <https://doi.org/10.1016%2Fj.mtcomm.2021.102709>.

Qingfu Zhang, Wudong Liu, Edward Tsang, and Botond Virginas. Expensive Multiobjective Optimization by MOEA/D With Gaussian Process Model. *IEEE Transactions on Evolutionary Computation*, 14(3):456–474, jun 2010. doi: 10.1109/tevc.2009.2033671. URL <https://doi.org/10.1109%2Ftevc.2009.2033671>.

Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, jun 2000. doi: 10.1162/106365600568202. URL <https://doi.org/10.1162%2F106365600568202>.