npj Computational Materials

https://doi.org/10.1038/s41524-026-02067-0

Article in Press

Collecting diverse near-optimal samples via nested Thompson sampling

Received: 7 December 2025

Accepted: 24 March 2026

Cite this article as: Shibukawa, R., Matsuda, S., Nakamura, K. *et al.* Collecting diverse near-optimal samples via nested Thompson sampling. *npj Comput Mater* (2026). https://doi.org/10.1038/s41524-026-02067-0

Ryosuke Shibukawa, Shoichi Matsuda, Kazuha Nakamura, Ryo Tamura & Koji Tsuda

Collecting diverse near-optimal samples via nested Thompson sampling

Ryosuke Shibukawa[1], Shoichi Matsuda*[2], Kazuha Nakamura[2], Ryo Tamura*[3,1,4], and Koji Tsuda*[1,3,4]

[1]Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan
[2]Center for Green Research on Energy and Environmental Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
[3]Center for Basic Research on Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
[4]RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan
*matsuda.shoichi@nims.go.jp, tamura.ryo@nims.go.jp, tsuda@k.u-tokyo.ac.jp

ABSTRACT

Self-driving laboratories (SDLs) that combine automated experiments with machine learning have accelerated data-driven discovery. Although Bayesian optimization (BO) is widely used in SDLs to autonomously propose experimental conditions, many real systems require sampling diverse near-optimal candidates rather than identifying a single optimum. We propose nested Thompson sampling (NTS), a batch BO method that enhances diversity by incorporating the concept of nested sampling. In NTS, regions where the posterior exceeds a likelihood threshold are uniformly sampled, enabling exploration of multiple promising regions while requiring only one hyperparameter, that is, the threshold schedule. Benchmark studies using materials datasets demonstrated that NTS achieves higher sample diversity than a conventional batch BO method. Furthermore, application of NTS to automated electrolyte exploration in an SDL successfully produced diverse experimental samples. The NTS algorithm is implemented in the NIMO package, providing a practical framework for autonomous and diverse materials exploration.

Introduction

Research innovation is being driven by self-driving laboratories (SDLs), which realize closed-loop experimentation by combining automated experiments with machine learning-based experimental condition proposals[1,2]. As an algorithm for proposing next experimental conditions without human intervention, Bayesian optimization (BO) is most commonly used and has been adopted in many SDLs[3–15]. Furthermore, because many automated experimental systems are capable of parallel experiments, there is an increasing demand for batch BO[16–19], which simultaneously proposes multiple experimental conditions. For example, as a conventional batch BO method, Snoek et al.[20] proposed a sequential process, where the scores near already selected points are reduced, theoretically encouraging the selection of diverse points across the batch. This approach can be viewed as a sampling-based extension of the kriging believer strategy proposed by Ginsbourger et al.[21]

When performing self-driving optimization in a completely unexplored experimental system, the true objective is not always to find only the single optimal solution[22–25]. For instance, in materials science, materials with slightly inferior performance but simpler fabrication processes, or those with lower cost, are often more valuable candidates than the optimal one. In such cases, the goal is to sample many near-optimal solutions with diverse experimental conditions. To address this, batch BO using a determinantal point process (DPP) is a promising candidate[26–28]. DPP ensures diversity among simultaneously proposed experimental conditions through Markov chain Monte Carlo (MCMC)-based computations. Nevertheless, DPP-based algorithms require the predefinition of many hyperparameters before starting the closed-loop experiments, making their application to new problems nontrivial. Therefore, a method capable of proposing diverse samples with as few hyperparameters as possible is required in SDLs.

We propose nested Thompson sampling (NTS), a method that achieves batch BO while maintaining sample diversity by incorporating the concept of nested sampling. Nested sampling is a technique that gradually "shrinks" the sample space by raising the likelihood threshold step by step, and the only hyperparameter is the schedule for increasing this threshold[29–33]. In NTS, first, models are sampled from the Gaussian process used as the surrogate model in BO (Thompson sampling). Next, when the likelihood exceeds the threshold, the posterior is treated as a constant value; otherwise, it is set to zero (see Fig. 1).

This enables sampling from multiple peaks in a multimodal landscape, resulting in higher sample diversity. Unlike conventional Thompson sampling, which selects the single point maximizing the posterior function as the next experimental condition, NTS uniformly samples the next experimental conditions from a broader region where the posterior exceeds the threshold. Thus, batch BO with high diversity can be performed by NTS.

In this study, we investigated three strategies for increasing the threshold L_t, which is the only hyperparameter in NTS: the top 10%, 5%, and 1% of previously obtained values were used as thresholds, named "conservative", "moderate", and "aggressive", respectively. These thresholds are commonly used in statistical hypothesis testing (e.g., 1%, 5%, and 10% significance levels) and provide a practical range for balancing exploration and exploitation. To quantify the diversity of sampled conditions, we used the metric #Circles[34,35]. The definition of #Circles is as follows:

$$\#\mathrm{Circles}(S;d,r) = \max_{C\subseteq S} |C| \text{ s.t. } d(\boldsymbol{x}_i,\boldsymbol{x}_{i'}) > r, \forall \boldsymbol{x}_i,\boldsymbol{x}_{i'} \in C, i \neq i', \tag{1}$$

where $S = \{\boldsymbol{x}_1,...,\boldsymbol{x}_{|S|}\}$ is the dataset of experimental conditions for near-optimal samples, C is a subset of S, and r is radius. Here, $d(\boldsymbol{x}_i,\boldsymbol{x}_{i'})$ is a distance metric between $\boldsymbol{x}_i$ and $\boldsymbol{x}_{i'}$. As r increases, multiple points become included in the same subset, and thus the #Circles metric decreases monotonically with respect to r (see Fig. 2). As shown in Fig. 2, in the case of dense data points (left case), when the radius r is increased, a large number of points immediately fall into the same subset, causing the #Circles to decrease rapidly. In contrast, for diverse data points (right case), increasing r does not immediately cause data points to fall into the same subset, and #Circles metric bulges upward and lies in the upper-right part compared to the dense case. Therefore, a function that bulges upward in the r-dependency of the #Circles indicates that the dataset contains more diverse samples. To quantify #Circles metric, we defined the area under the #Circles curve (Circles-AUC) as follows:

$$I = \int_0^\infty \#\mathrm{Circles}(S;r)dr. \tag{2}$$

A larger Circles-AUC indicates that more diverse near-optimal samples are obtained.

We conducted a benchmark study using existing materials datasets: melting temperatures of an alloy system[36] and grain-boundary energies calculated by first-principles calculations[37]. As a result, NTS was found to produce experimental conditions with higher diversity compared to a conventional batch BO method (Snoek et al.) implemented in PHYSBO package[38]. Furthermore, NTS demonstrated better optimization performance than random sampling, confirming that it fulfills the motivation of efficiently sampling many near-optimal and diverse samples. In addition, we applied NTS to electrolyte discovery experiments using an SDL, and confirmed that diverse samples were successfully obtained in real experiments as well. The NTS algorithm has been implemented in the NIMO package, which is a software for SDLs and is available in GitHub (https://github.com/NIMS-DA/nimo)[39,40].

Results

Benchmark using materials datasets

We used the melting-temperature data of Al-Mg-Zn alloys[36] and the grain boundary energy dataset[37] as benchmark materials datasets. The numbers of data points were 1,326 and 17,232 for the former and latter cases, respectively. The dimensionality of the search space was two (compositions) for the melting-temperature dataset and three (grain boundary structural descriptors) for the grain boundary energy dataset. In the batch setting, the batch size was fixed at $B = 3$ and 24, respectively. We evaluated the average simple regret and the #Circles metrics over five independent trials with the same initial samples but different random seeds, as shown in Fig. 3.

NTS was computed under three operating modes. The simple regret results (see Figs. 3(a) and (d)) showed that the conventional batch BO (PHYSBO) outperformed random sampling and NTS for the melting-temperature dataset, whereas for the grain boundary dataset, the optimization results were not significantly different from random sampling. To compute the #Circles metric, it is necessary to define the criterion for "near-optimal" points. Here, we assumed data points with target values greater than the 90th percentile of the entire dataset as near-optimal. The radius dependence of #Circles is shown in Figs. 3(b) and (e), where NTS expands toward the upper right, indicating that the obtained samples exhibit higher diversity. Here, to evaluate #Circles, we used the Euclidean distance as a distance metric. On the other hand, the results of the three NTS modes fall within the error bars for both regret and #Circles over almost the entire range of radii, suggesting that the overall differences among the modes are small for these datasets.

We compared the Circles-AUC values of the "conservative" mode of NTS, PHYSBO, and random sampling, as shown in Figs. 3(c) and (f). NTS yielded the highest values, confirming that the conditions sampled by NTS were more diverse. Note that the results discussed in the Supplementary Note A indicate that diverse sampling is maintained even when the batch size is varied in the NTS method. Furthermore, from Fig. 3(b), it can be inferred that the aggressive mode is preferable when faster optimization is desired, although the appropriate mode should be selected depending on the objective of the exploration.

Electrolyte exploration in self-driving laboratory using NTS

In this section, we present results for electrolyte explorations using NIMS automated robotic electrochemical experiments (NAREE)[7]. NAREE is a system that enables high-throughput screening of electrolytes by utilizing a microplate-type electrochemical cell. Electrochemical cells were fabricated using $LiFePO_4$ as positive electrode and Cu foil as negative electrode. Here, the objective function value was set to discharge time (sec). To maximize the objective function, a combination of electrolyte additives was optimized. Five different additives were selected from a list of 32 compounds (Table 1) and injected into an electrochemical cell containing 1 M Lithium bis(trifluoromethanesulfonyl)imide (LiTFSI) in propylene carbonate (PC). The total number of experimental conditions is ${}_{32}C_5 = 201,376$. To express the experimental condition as a vector, 32 binary variables are used: A value of 1 is set if the additive is used, and 0 if it is not. Consequently, for each vector, the number of variables with a value of 1 is always 5. Since NAREE is connected to some exploration algorithms through NIMO, the self-driving electrolyte explorations without human intervention can be performed. Thus, using NIMO, the closed-loop experiments of NAREE and NTS were performed (see Fig. 4(a)). Note that for binary variables, the squared Euclidean distance in the RBF kernel directly corresponds to the Hamming distance, allowing the model to effectively capture similarities in discrete spaces.

Batch size was set to $B = 48$ and 10 independent cycles were conducted using conventional batch BO ("PHYSBO" in NIMO) and NTS. As an initial dataset, we selected 48 experimental conditions randomly. The NTS exploration was performed in the "conservative" mode. The dependence of the maximum value of discharge time on cycles is shown in Fig. 4(b). From these results, electrolytes whose discharge time was larger than 0.14 sec were regarded as near-optimal. To examine the diversity of the near-optimal samples, the #Circles metrics for the samples whose discharge time was larger than 0.14 sec were plotted in Fig. 4(c). Here, to evaluate #Circles, we used the Hamming distance as a distance metric. The results of #Circles as a function of radius obtained by NTS extend further toward the upper-right region compared with those by conventional batch BO, indicating that the samples generated by NTS possess higher diversity than conventional batch BO. In Fig. 4(c), the Circles-AUC values are also shown, indicating that the diversity of samples by NTS is higher than that by conventional batch BO. Figure 4(d) also shows a two-dimensional mapping of the near-optimal samples using t-SNE, where the NTS samples exhibit a more widely spread distribution. These results demonstrate that, while conventional batch BO methods often prove insufficient, near-optimal sampling can be achieved with NTS in self-driving laboratories, even though the only hyperparameter that needs to be specified externally is the exploration mode.

Discussions

To efficiently collect many near-optimal samples, the nested Thompson sampling (NTS) method was proposed. This method supports a batch approach, making it suited for high-throughput parallel experimentation. Using materials datasets, we confirmed that the NTS method can obtain samples with higher diversity than conventional batch Bayesian optimization (BO). We then conducted automated electrolyte exploration by combining NTS with robotic electrochemical experiments, which successfully produced diverse samples in real-world experiments. Rather than focusing solely on the rapid identification of a single global optimum via conventional BO, a strategy that explores a diverse set of near-optimal candidates using NTS offers significant advantages in materials discovery. In this field, it is rarely the case that a single target property determines the overall excellence of a material. In practice, multiple competing factors, such as synthesis cost, stability, and environmental impact, must be balanced. Therefore, identifying a diverse set of near-optimal candidates is a more robust strategy. This approach enables the selection of the most suitable candidates from a broad pool of high-performing options, accounting for secondary criteria that may not be explicitly included in the automated optimization cycle.

The NTS method has been implemented in the NIMO package, enabling seamless integration into robotic experimental systems. By combining self-driving laboratories (SDLs) with NTS, it becomes possible to acquire numerous near-optimal samples, from which users can select the most suitable conditions according to their objectives, such as prioritizing simpler processes or lower costs. This new approach is expected to be highly compatible with SDLs, which can perform extensive experiments without human intervention, and we believe it will drive significant innovation in the research process.

Methods

Nested Thompson Sampling

Nested sampling was originally proposed for high-dimensional integration in Bayesian inference[29,30]. It has demonstrated success in a wide range of applications, including cosmology and particle physics[41,42]. Starting from randomly initialized points, the nested sampling algorithm progressively restricts the search space to more promising regions by replacing low-performing members of the population with higher-performing ones. In the presence of multimodality, the algorithm can simultaneously track multiple peaks, thereby promoting the generation of solutions with greater diversity.

By incorporating the idea of nested sampling into black-box optimization, it becomes possible to collect diverse near-optimal samples. In our algorithm, Thompson sampling is employed to define a constrained posterior distribution that determines the next proposals, which we refer to as nested Thompson sampling (NTS). Gaussian process (GP) is a probability distribution where a function $f(\boldsymbol{x})$ corresponds to the realization of a random variable[43]. The fundamental idea of GP is that a function gives similar outputs for similar inputs: the similarity between $f(\boldsymbol{x}_1)$ and $f(\boldsymbol{x}_2)$ is close to that of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. To measure the similarity between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, a kernel function is needed. In this work, a popular radial basis function (RBF) kernel was adopted. In Gaussian process (GP) regression, $\mathrm{GP}(f|D_t)$ models the function $f(\boldsymbol{x})$ as a probability distribution, allowing one to infer the likely shapes of the underlying function from the observed dataset $D_t = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_{|D_t|}, y_{|D_t|})\}$. In BO using Thompson sampling, a function is sampled from the posterior GP distribution $\mathrm{GP}(f|D_t)$, and the next proposal is determined by the maximum of the sampled function.

Instead of maximum distribution, we consider a constrained posterior distribution $\pi_t(\boldsymbol{x})$, which is defined as the probability that $\boldsymbol{x}$ yields a function value exceeding a threshold L_t. Then, $\pi_t(\boldsymbol{x})$ is given as

$$\pi_t(\boldsymbol{x}) \propto \mathbb{E}_{\tilde{f}\sim \mathrm{GP}(f|D_t)}\left[\mathbb{1}_{\tilde{f}(\boldsymbol{x})>L_t}(\boldsymbol{x})\right], \tag{3}$$

where L_t is a threshold and $\tilde{f}(\boldsymbol{x})$ is sampled function from GP regression. $\mathbb{1}_A(\boldsymbol{x})$ is indicator function defined as

$$\mathbb{1}_A(\boldsymbol{x}) = \begin{cases} 1 & \boldsymbol{x} \in A \\ 0 & \boldsymbol{x} \notin A \end{cases}. \tag{4}$$

In addition, $\mathbb{E}_{\tilde{f}\sim \mathrm{GP}(f|D_t)}[\cdot]$ is approximated by the sample average from the posterior GP. Thus, $\pi_t(\boldsymbol{x})$ is given by the sample average of the indicator function. In our implementation, number of generated samples is fixed as 100. In NTS, multiple candidates are generated by sampling from the constrained posterior distribution $\pi_t(\boldsymbol{x})$, enabling simultaneous exploration of diverse solution space. Algorithm 1 describes the whole procedure.

Algorithm 1 Nested Thompson sampling

procedure NTS
  $D_0 \leftarrow \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_{|D_0|}, y_{|D_0|})\}$
  $t \leftarrow 0$
  repeat
    compute Gaussian process posterior $\mathrm{GP}(f \mid D_t)$
    select a threshold L_t
    perform Thompson sampling multiple times and construct $\pi_t(\boldsymbol{x})$
    sample the next batch $B = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{|\mathrm{batch\_size}|}\}$ by drawing i.i.d. samples from π_t
    $D_B \leftarrow \{(\boldsymbol{x}, f(\boldsymbol{x})) | \boldsymbol{x} \in B\}$
    $D_{t+1} \leftarrow D_t \cup D_B$
    $t \leftarrow t + 1$
  until until the predetermined number of cycles is reached

To determine the value of L_t, we proposed three strategies termed "conservative", "moderate", and "aggressive" which represent the only hyperparameter in NTS. In these strategies, the top 10%, 5%, and 1% of previously obtained values were used as L_t, respectively. Consequently, L_t is monotonically increased with each cycle.

Calculation of #Circles Metrics

As a measure of the diversity of near-optimal samples, we adopted the #Circles metric that is currently proposed for evaluating the diversity of chemical space[34,35], which is defined in Eq. (1). The #Circles metric corresponds to the packing number in graph theory[34]. In practice, one can construct an undirected graph $G = (V, E)$ by adding an edge between $\boldsymbol{x}_i, \boldsymbol{x}_{i'} \in S$ if $d(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) \leq r$. The #Circles metric can then be obtained by computing a maximal independent set of G using any graph software. In this work, we employed the `maximal_independent_set` method of NetworkX 3.4.2[44] to compute the #Circles metric.

Implementation in NIMO

We implemented the NTS algorithm in NIMO[39,40]. NTS can be applied through the `nimo.selection` function as follows

```
nimo.selection(method = "NTS",
               input_file = "candidates.csv",
```

```
output_file = "proposals.csv",
num_objectives = 1,
num_proposals = 48,
sample_mode = "conservative")
```

The candidates are stored beforehand in `candidates.csv`, and the proposals are saved to `proposals.csv`. Subsequently, experiments will be conducted according to the experimental conditions listed in `proposals.csv`. The parameter `sample_mode` specifies the strategy to control a threshold L_t, where "conservative", "moderate", and "aggressive" can be selected. If the data contains significant noise, using the "aggressive" mode with a strict threshold from the early stages may lead to the omission of promising regions. Conversely, since the threshold increases monotonically even in "conservative" mode, it is essential to prepare a sufficient amount of initial data to ensure that potential candidate regions are not overlooked.

Data availability

The datasets generated and analyzed during the current study are available in the GitHub repository at https://github.com/shibukawar/nimo.

Code availability

The code supporting the findings of this study is available from the GitHub https://github.com/shibukawar/nimo.

Acknowledgements

This work was supported by Ministry of Education, Culture, Sports, Science, and Technology (MEXT) Program: Data Creation and Utilization Type Materials Research and Development Project [Grant Number JPMXP1121467561], JST CREST [Grant Number JPMJCR21O2], and JST PRESTO [Grant Number JPMJPR24T8].

Author Contributions

The authors confirm contributions to the paper as follows: study conception and design: R.T. and K. T.; methodology: R.S. and K.T.; code development and execution: R.S. and R.T.; experiments: S.M. and K.N. writing—initial draft: R.S. and R.T.; writing—review and editing: all authors.

Competing interests

The authors declare no competing interests.

References

1. Tom, G. *et al.* Self-driving laboratories for chemistry and materials science. *Chem. Rev.* **124**, 9633–9732 (2024).
2. Yoshikawa, N. *et al.* Self-driving laboratories in Japan. *Digit. Discov.* **4**, 1384–1403 (2025).
3. Wakabayashi, Y. K. *et al.* Machine-learning-assisted thin-film growth: Bayesian optimization in molecular beam epitaxy of $SrRuO_3$ thin films. *APL Mater.* **7** (2019).
4. Burger, B. *et al.* A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
5. MacLeod, B. P. *et al.* Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **6** (2020).
6. Shimizu, R., Kobayashi, S., Watanabe, Y., Ando, Y. & Hitosugi, T. Autonomous materials synthesis by machine learning and robotics. *APL Mater.* **8** (2020).
7. Matsuda, S., Lambard, G. & Sodeyama, K. Data-driven automated robotic experiments accelerate discovery of multi-component electrolyte for rechargeable Li–O_2 batteries. *Cell Reports Phys. Sci.* **3**, 100832 (2022).
8. Wakabayashi, Y. K. *et al.* Bayesian optimization with experimental failure for high-throughput materials growth. *npj Comput. Mater.* **8**, 180 (2022).
9. Rooney, M. B. *et al.* A self-driving laboratory designed to accelerate the discovery of adhesive materials. *Digit. Discov.* **1**, 382–389 (2022).
10. Fébba, D. M. *et al.* Autonomous sputter synthesis of thin film nitrides with composition controlled by Bayesian optimization of optical plasma emission. *APL Mater.* **11** (2023).

11. Low, A. K. *et al.* Evolution-guided Bayesian optimization for constrained multi-objective optimization in self-driving labs. *npj Comput. Mater.* **10**, 104 (2024).

12. Schilter, O. *et al.* Combining Bayesian optimization and automation to simultaneously optimize reaction conditions and routes. *Chem. Sci.* **15**, 7732–7741 (2024).

13. Fukui, Y. *et al.* Automated odor-blending with one-pot Bayesian optimization. *Digit. Discov.* **3**, 969–976 (2024).

14. Zheng, Y. B. *et al.* A self-driving physical vapor deposition system making sample-specific decisions on the fly. *npj Comput. Mater.* **11**, 327 (2025).

15. Sin, J. W. *et al.* Highly parallel optimisation of chemical reactions through automation and machine intelligence. *Nat. Commun.* **16**, 6464 (2025).

16. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **104**, 148–175 (2015).

17. Roch, L. M. *et al.* ChemOS: orchestrating autonomous experimentation. *Sci. Robotics* **3**, eaat5559 (2018).

18. Häse, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. Phoenics: a Bayesian optimizer for chemistry. *ACS central science* **4**, 1134–1145 (2018).

19. Adachi, M. *et al.* Looping in the human: Collaborative and explainable Bayesian optimization. In Dasgupta, S., Mandt, S. & Li, Y. (eds.) *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, vol. 238 of *Proceedings of Machine Learning Research*, 505–513 (PMLR, 2024).

20. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. In Pereira, F., Burges, C., Bottou, L. & Weinberger, K. (eds.) *Advances in Neural Information Processing Systems*, vol. 25 (Curran Associates, Inc., 2012).

21. Ginsbourger, D., Le Riche, R. & Carraro, L. *Kriging Is Well-Suited to Parallelize Optimization*, 131–162 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010).

22. Konakovic Lukovic, M., Tian, Y. & Matusik, W. Diversity-guided multi-objective Bayesian optimization with batch evaluations. *Adv. Neural Inf. Process. Syst.* **33**, 17708–17720 (2020).

23. Ahmadianshalchi, A., Belakaria, S. & Doppa, J. R. Pareto front-diverse batch multi-objective Bayesian optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 10784–10794 (2024).

24. Rolnick, D. *et al.* Position: Application-driven innovation in machine learning. *Proc. Mach. Learn. Res.* **235**, 42707–42718 (2024).

25. Maguire, C., Hardcastle, C., Hastings, T., Arróyave, R. & Vela, B. Good enough is better: Feasibility vs. pareto-optimality in alloy design. *arXiv preprint arXiv:2510.20125* (2025).

26. Kathuria, T., Deshpande, A. & Kohli, P. Batched Gaussian process bandit optimization via determinantal point processes. *Adv. neural information processing systems* **29** (2016).

27. Oh, C., Bondesan, R., Gavves, E. & Welling, M. Batch Bayesian optimization on permutations using the acquisition weighted kernel. *Adv. Neural Inf. Process. Syst.* **35**, 6843–6858 (2022).

28. Nava, E., Mutny, M. & Krause, A. Diversified sampling for batched Bayesian optimization with determinantal point processes. In *International Conference on Artificial Intelligence and Statistics*, 7031–7054 (PMLR, 2022).

29. Skilling, J. Nested sampling. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 735, 395–405 (2004).

30. Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Analysis* **1**, 833 (2006).

31. Sivia, D. & Skilling, J. *Data analysis: a Bayesian tutorial* (OUP Oxford, 2006).

32. Chopin, N. & Robert, C. P. Properties of nested sampling. *Biometrika* **97**, 741–755 (2010).

33. Buchner, J. Nested sampling methods. *Stat. Surv.* **17**, 169–215 (2023).

34. Xie, Y., Xu, Z., Ma, J. & Mei, Q. How much space has been explored? measuring the chemical space covered by databases and machine-generated molecules. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)* (2023).

35. Renz, P., Luukkonen, S. & Klambauer, G. Diverse hits in de novo molecule design: Diversity-based comparison of goal-directed generators. *J. Chem. Inf. Model.* **64**, 5756–5761 (2024).

36. Hayashi, N., Nakashima, K., Enoki, M. & Ohtani, H. Thermodynamic analysis of the Al-Mg-Zn ternary system. *Nippon. Kinzoku Gakkaishi/J. Jpn. Inst. Met.* **84**, 141–150 (2020).

37. Kiyohara, S., Oda, H., Tsuda, K. & Mizoguchi, T. Acceleration of stable interface structure searching using a kriging approach. *Jpn. J. Appl. Phys.* **55**, 045502 (2016).

38. Motoyama, Y. *et al.* Bayesian optimization package: PHYSBO. *Comput. Phys. Commun.* **278**, 108405 (2022).

39. Tamura, R., Tsuda, K. & Matsuda, S. NIMS-OS: an automation software to implement a closed loop between artificial intelligence and robotic experiments in materials science. *Sci. Technol. Adv. Materials: Methods* **3**, 2232297 (2023).

40. Tamura, R. *et al.* Seamless integration of legacy robotic systems into a self-driving laboratory via NIMO: a case study on liquid handler automation. *Sci. Technol. Adv. Materials: Methods* 2565144 (2025).

41. Speagle, J. S. dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Notices Royal Astron. Soc.* **493**, 3132–3158 (2020).

42. Ashton, G. *et al.* Nested sampling for physical scientists. *Nat. Rev. Methods Primers* **2**, 39 (2022).

43. Williams, C. K. & Rasmussen, C. E. *Gaussian processes for machine learning*, vol. 2 (MIT press Cambridge, MA, 2006).

44. Varoquaux, G., Vaught, T. & Millman, J. (eds.). *Exploring Network Structure, Dynamics, and Function using NetworkX* (Pasadena, CA USA, 2008).

Figure 1. Brief description of nested Thompson sampling (NTS). From the Gaussian process (left panel), one model is sampled (Thompson sampling). In conventional BO, the proposed point is determined by the maximum posterior (middle panel). In NTS (right panel), the posterior takes a constant value when the sampled function value is higher than L_t, and becomes zero when it is lower than the threshold value L_t. Multiple conditions can then be uniformly selected from the regions with constant and finite posterior values.

Figure 2. Illustration of the #Circles metric. A higher curve indicates a more diverse sample set. The area under the curve is defined as Circles-AUC, I, and datasets with higher diversity exhibit larger areas, that is, $I_2 > I_1$.

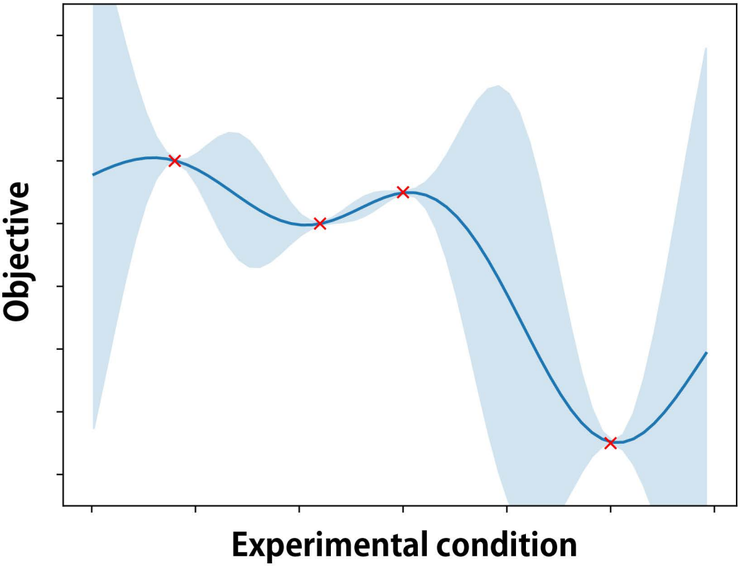
Figure 3. Benchmark results on materials datasets. The melting temperature and grain boundary datasets are used for panels (a)–(c) and (d)–(f), respectively. Simple regret as a function of cycle ((a) and (d)), number of circles (#Circles) as a function of radius ((b) and (e)), and Circles-AUC for near-optimal samples ((c) and (f)). Near-optimal samples are defined as data points whose target values are greater than the 90th percentile of the entire dataset. The lines represent the average value, and the shaded area and error bars indicate the deviation from five independent trials.

Figure 4. Exploration of electrolytes using the NAREE system and the NTS method. (a) Workflow of the exploration process without human intervention. (b) Objective function values (discharge time (sec)) as a function of cycle obtained by PHYSBO and the NTS method. (c) Number of circles (#Circles) as a function of radius. (d) Dimensionality reduction results for near-optimal samples obtained using t-SNE. Near-optimal samples are defined as those having target values greater than 0.14.

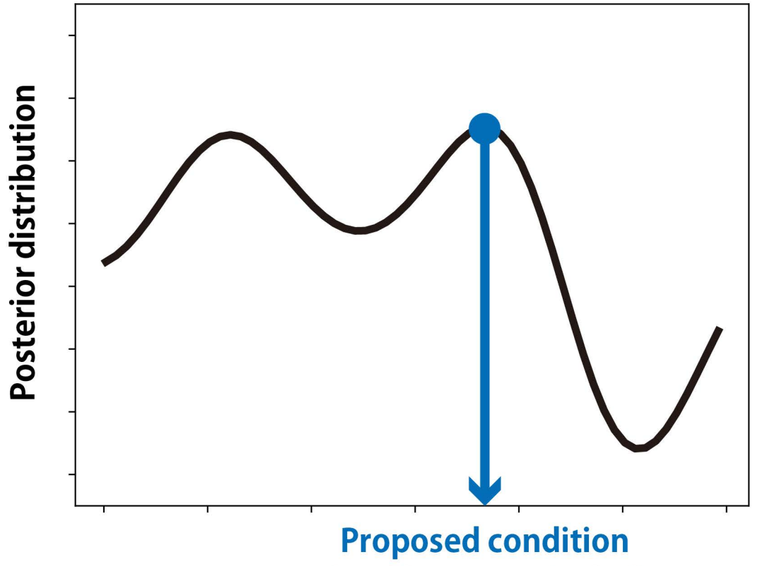
Table 1. 32 additives used in electrolyte exploration in a self-driving laboratory.

No.	Solution
1	1M Lithium bis(pentafluoroethanesulfonyl)imide in PC
2	1M Lithium bis(fluorosulfonyl)imide in PC
3	1M Lithium trifluoromethanesulfonate in PC
4	1M Lithium tetrafluoroborate in PC
5	1M Lithium hexafluorophosphate in PC
6	1M Lithium difluoro(oxalato)borate in PC
7	1M Lithium perchlorate in PC
8	50vol% Dimethylsulfamoyl fluoride in PC
9	0.1M Lithium bis(pentafluoroethanesulfonyl)imide in PC
10	0.1M Lithium bis(fluorosulfonyl)imide in PC
11	0.1M Lithium trifluoromethanesulfonate in PC
12	0.1M Lithium tetrafluoroborate in PC
13	0.1M Lithium hexafluorophosphate in PC
14	0.1M Lithium difluoro(oxalato)borate in PC
15	0.1M Lithium perchlorate in PC
16	5vol% Dimethylsulfamoyl fluoride in PC
17	50vol% Trimethyl phosphate in PC
18	50vol% Triethyl phosphate in PC
19	50vol% N-methyl-2-pyrrolidone in PC
20	50vol% Tetraethylene glycol dimethyl ether in PC
21	50vol% Dimethyl sulfoxide in PC
22	50vol% 1,3-Propane sultone in PC
23	50vol% 2-Propynyl methanesulfonate in PC
24	50vol% Sulfolane in (1M LiTFSI in PC)
25	5vol% Trimethyl phosphate in PC
26	5vol% Triethyl phosphate in PC
27	5vol% N-methyl-2-pyrrolidone in PC
28	5vol% Tetraethylene glycol dimethyl ether in PC
29	5vol% Dimethyl sulfoxide in PC
30	5vol% 1,3-Propane sultone in PC
31	5vol% 2-Propynyl methanesulfonate in PC
32	5vol% Sulfolane in (1M LiTFSI in PC)

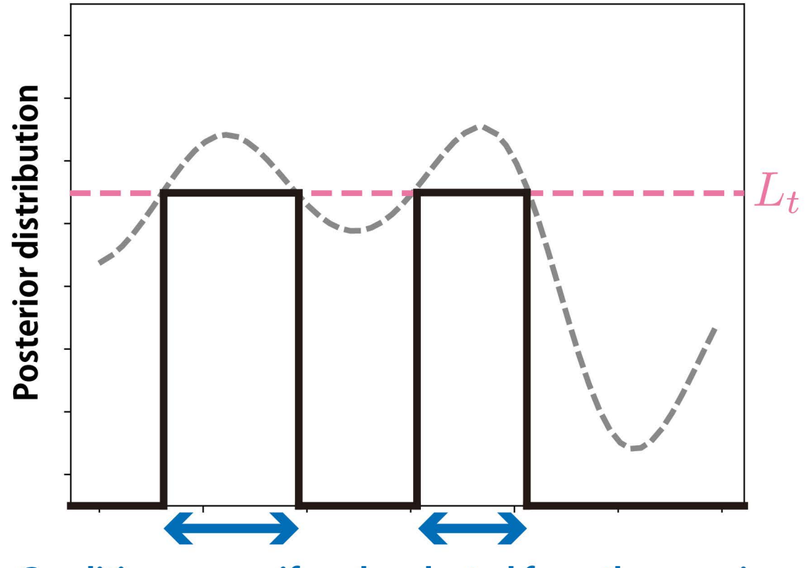
Gaussian process regression



Thompson sampling



Nested Thompson sampling



Dense data points

small radius

#Circles = 4

large radius

#Circles = 1

#Circles

n

I_2

I_1

1

radius

$I_1 < I_2$

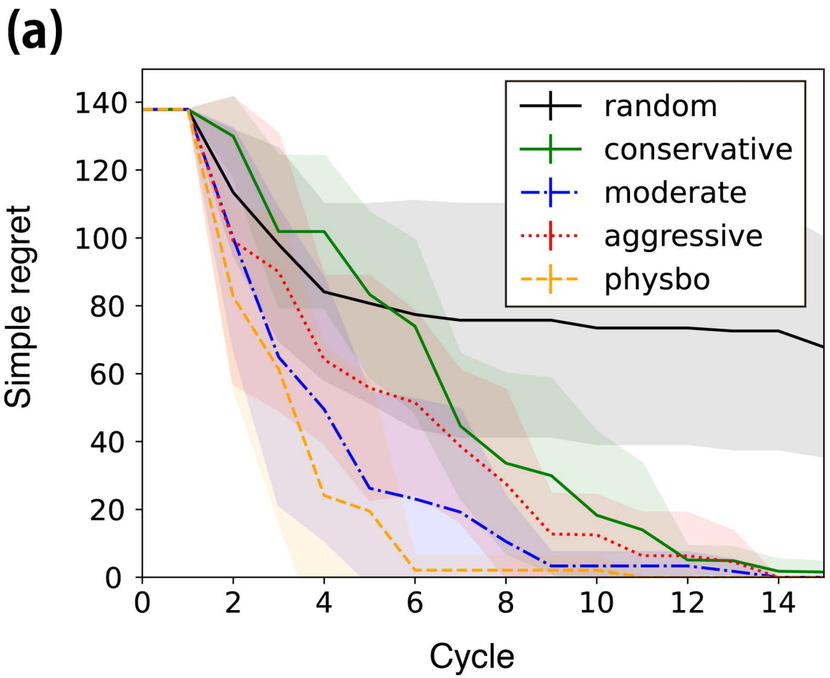
Diverse data points

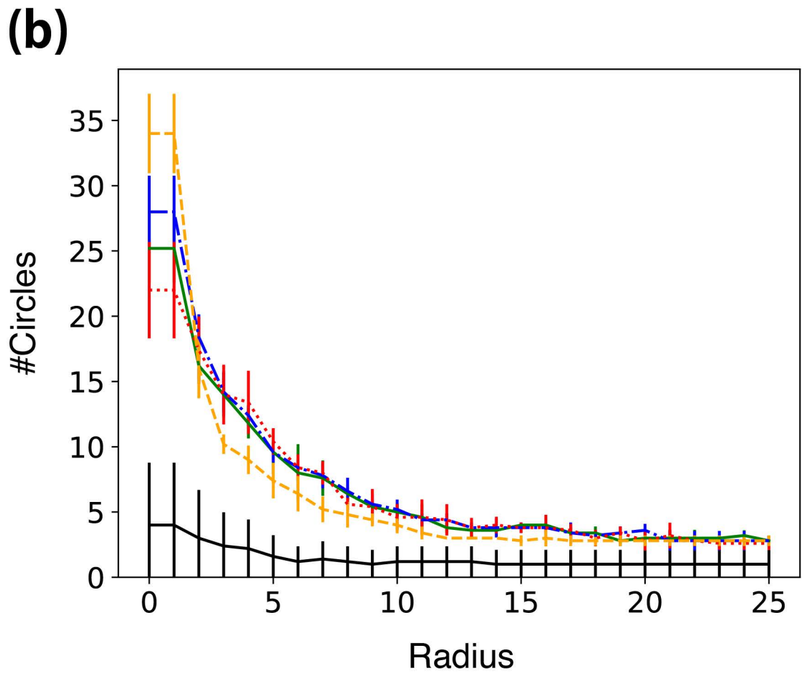
small radius

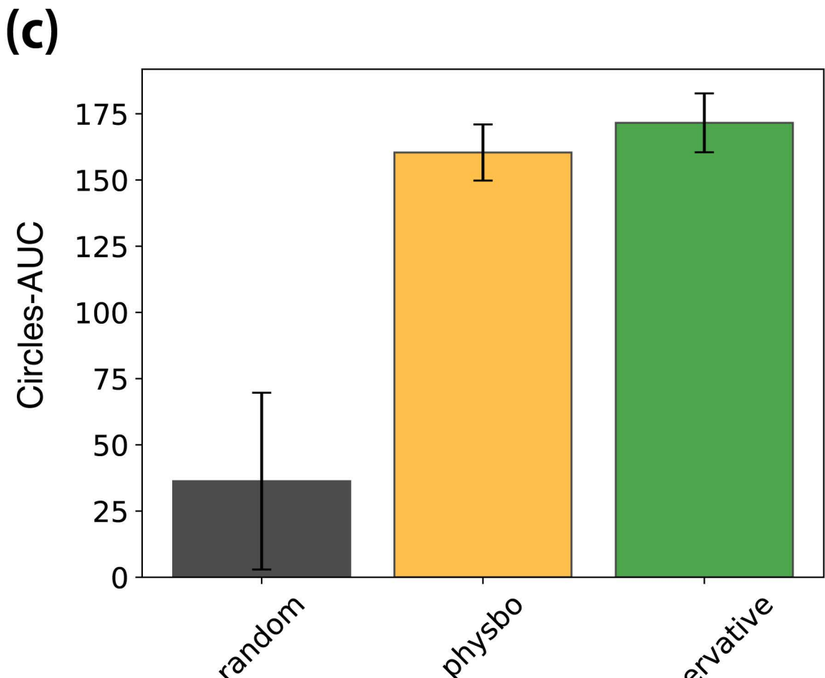
#Circles = 5

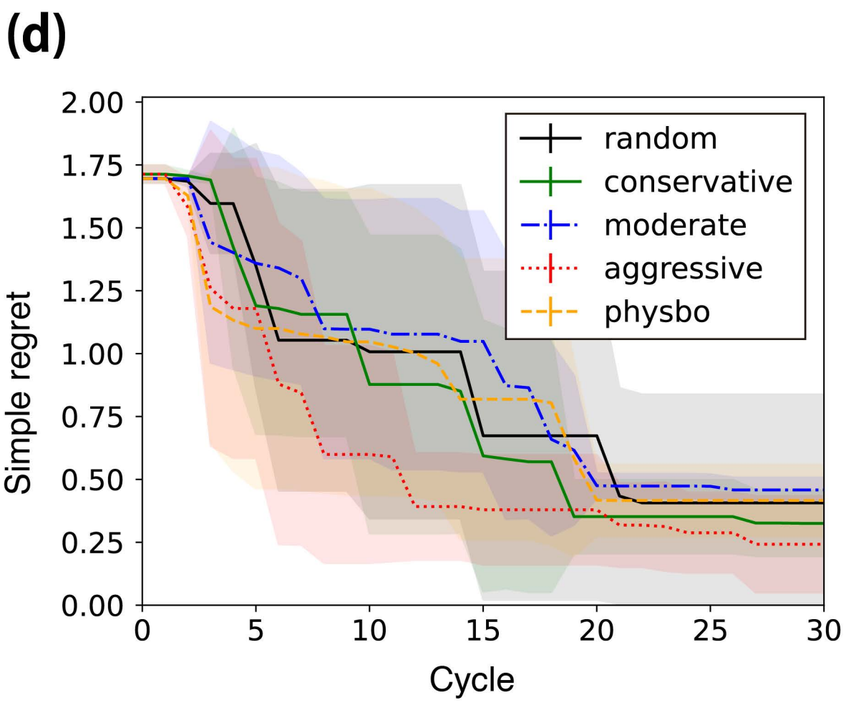
large radius

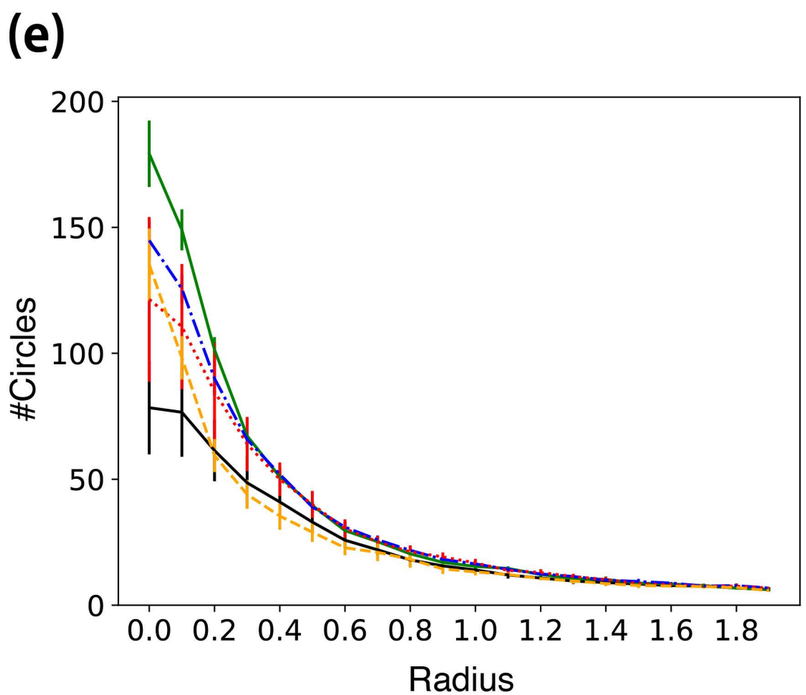
#Circles = 5

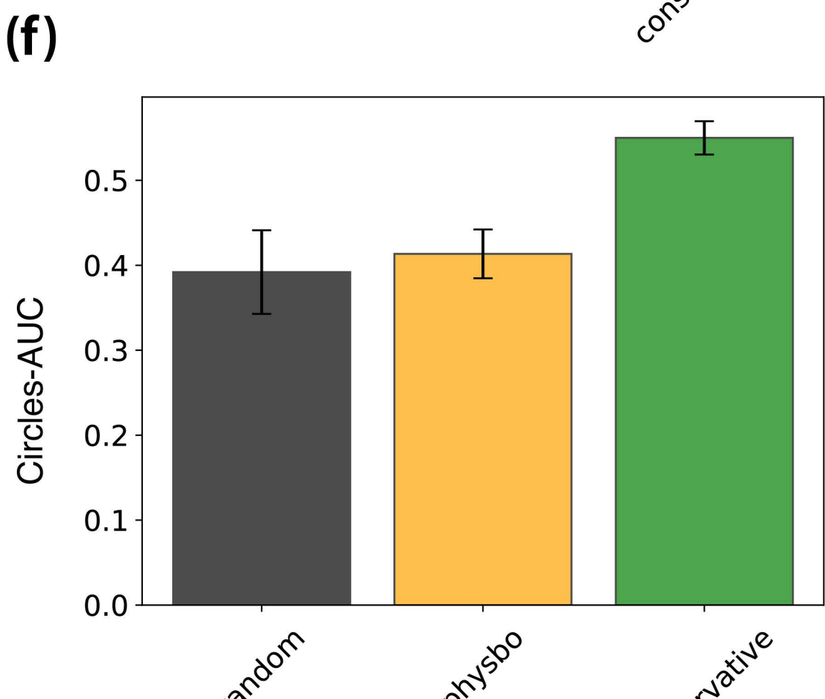

(a)
Simple regret
Cycle
random
conservative
moderate
aggressive
physbo

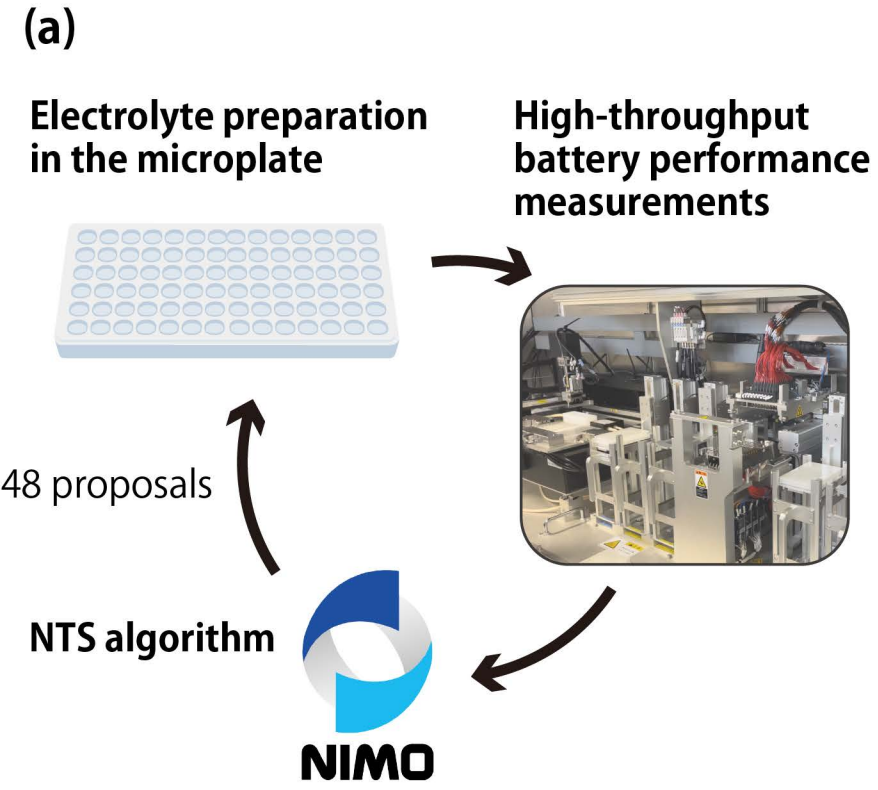

(b)
#Circles
Radius

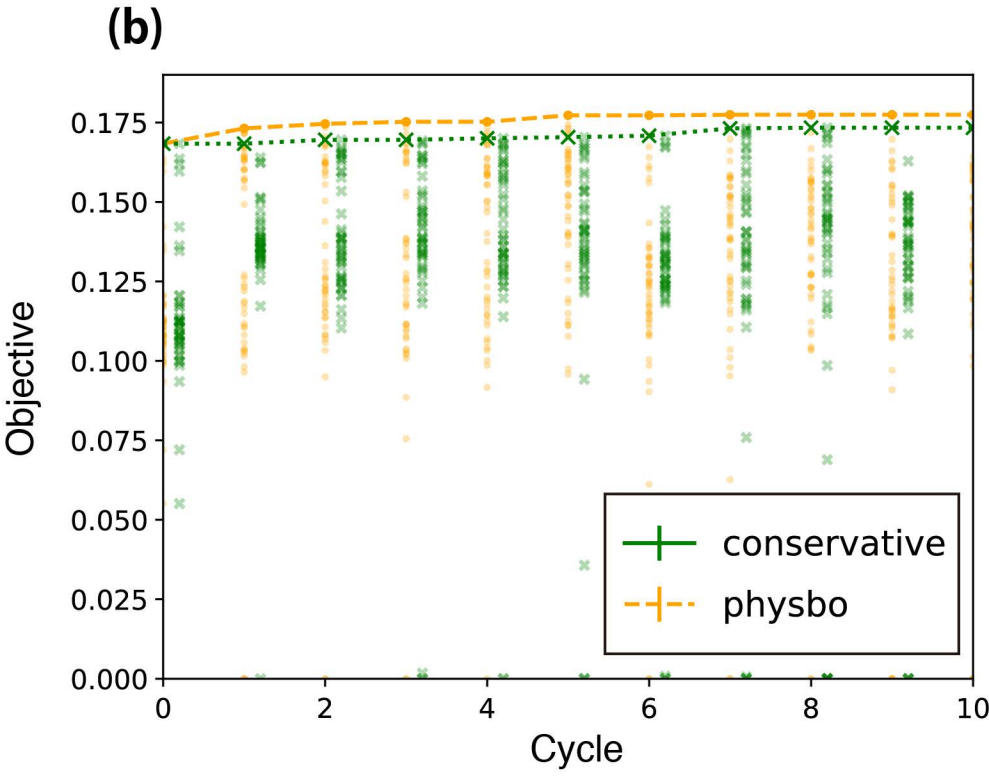

(c)
Circles-AUC
random
physbo
conservative

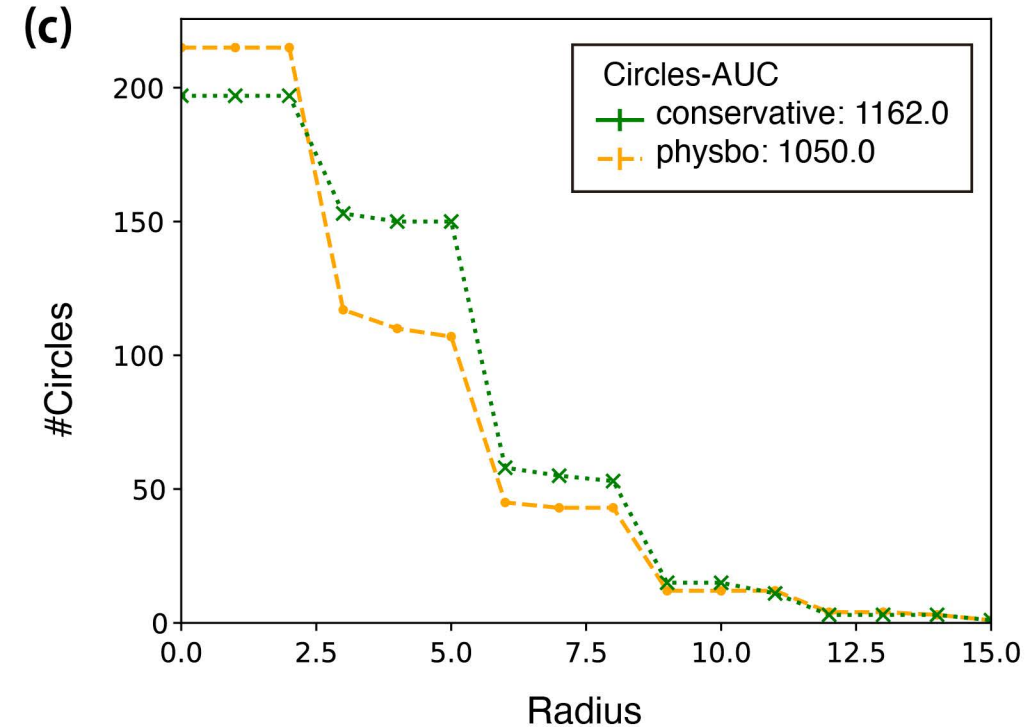

(d)
Simple regret
Cycle
random
conservative
moderate
aggressive
physbo

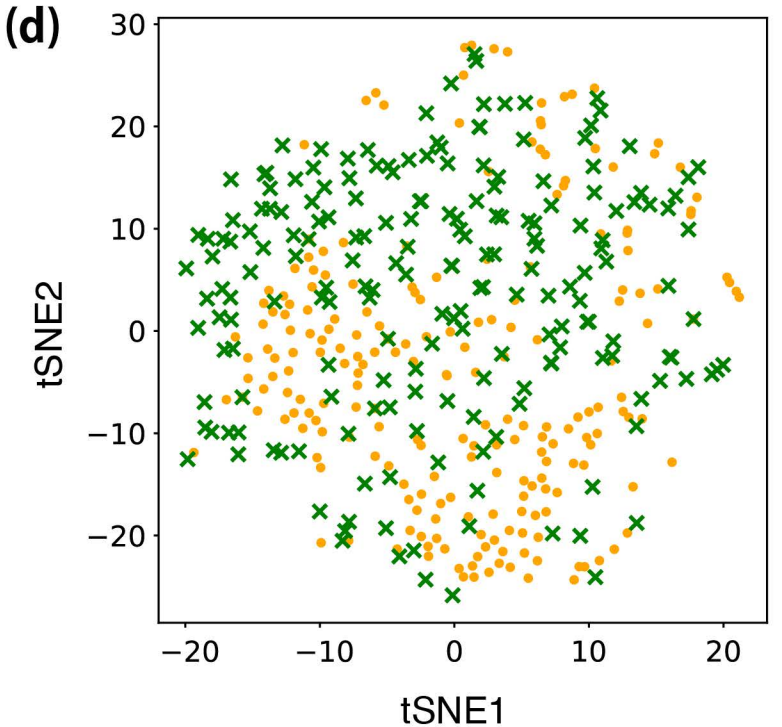

(e)
#Circles
Radius


(f)
Circles-AUC
random
physbo
conservative


(a)
Electrolyte preparation in the microplate
High-throughput battery performance measurements
48 proposals
NTS algorithm
NIMO


(b)
Objective
Cycle
conservative
physbo


(c)
Circles-AUC
conservative: 1162.0
physbo: 1050.0
#Circles
Radius


(d)
tSNE2
tSNE1