Cost-Sensitive Experimental Design for Atomistic Modeling

Chong Liu[1] CHONGLIU@CS.UCSB.EDU
Anirudh Raju Natarajan[2] ANIRUDH.NATARAJAN@EPFL.CH
Derick Evan Ober[3] DOBER@UCSB.EDU
Anton Van der Ven[3] AVDV@ENGINEERING.UCSB.EDU
Yu-Xiang Wang[1] YUXIANGW@CS.UCSB.EDU

1 *Department of Computer Science, University of California, Santa Barbara, CA 93106, USA*
2 *Laboratory of Materials Design and Simulation, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*
3 *Department of Materials, University of California, Santa Barbara, CA 93106, USA*

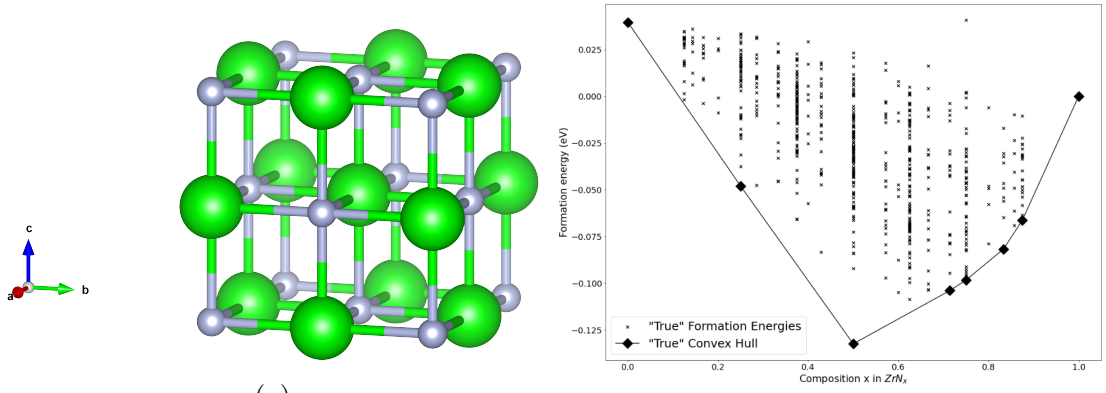
Abstract

Parameterizing surrogate models that capture chemical interactions at the atomistic scale is a long-standing challenge in materials science. It is desirable to obtain highly predictive surrogate models from the smallest possible pool of training data points. Conventional methods parameterize these models by systematically counting over symmetrically-distinct arrangements of chemical species with increasing periodicities. However, it is unclear if such an approach provides an optimal route to attaining robust models with a minimal computational cost. In this paper, we formulate the problem as a cost-sensitive experimental design problem and propose a new algorithm to minimize the computational cost. The key motivation is that structures with larger periodicities may lead to better modeling in earlier stages than conventional algorithms. By a new definition of improvement called Maximum Improvement Per Unit Cost (MI-PUC), our algorithm adaptively and efficiently deals with the performance-cost trade-off and works well in our experiments. The methodology described in this study provides material scientists with a rigorous benchmark of the performance-cost trade-off and a systematic way of attaining accurate atomistic models.

1. Introduction

Atomistic models are invaluable in predicting the high-temperature properties of several technologically important materials. These models serve as surrogates that replace computationally expensive quantum mechanical calculations. Typically, a small number of expensive quantum mechanical simulations are used to train surrogate atomistic models. Modern regression algorithms have been very helpful in parameterizing surrogate models that perform well on datasets that the model was not trained on. However, an outstanding problem in the field is the identification of the smallest possible pool of training points that can still be used to train accurate and predictive surrogate atomistic models.

Cluster expansion hamiltonians are a versatile and popular framework to construct surrogate atomistic models. They provide a rigorous framework that can be used to describe the properties of any arbitrary arrangement of chemical species on a crystal structure like



(a) (b)

Figure 1: (a) Conventional unit cell for the Rocksalt ZrN crystal. Zr atoms are represented by green spheres, and N atoms are represented by gray spheres. Removing some of the N atoms in larger unit cells will give different crystal configurations. Each configuration has a different normalized free energy, plotted in (b). The thermodynamically favored configurations have formation energies that lie on the lower convex hull in space defined by formation energy and composition x in ZrN_x.

Figure 1. In exact formulation, this model is an expansion (similar in spirit to a Taylor expansion) with infinite terms. Practically however, the expansions are truncated to contain a finite number of features and associated regression coefficients. Despite truncation, cluster expansion Hamiltonians still contain more features than the number of training data points. Feature selection algorithms such as genetic algorithms, lasso and ridge regression have been employed in previous studies to alleviate the problem of training underdetermined linear models. Alternately, cluster expansion models can also be parameterized with non-linear models such as neural networks. While all of these techniques are useful in achieving excellent surrogate models given a training dataset they do not provide any guidance on choosing the optimal set of points that can help parameterize a predictive model.

Given a large design pool of data points, the experimental design problem studies how to select a subset of data points to maximize the statistical efficiency regressed on these selected points (Allen-Zhu et al., 2021). Here maximizing statistical efficiency can be measured by minimizing several optimality criteria, including A(verage), D(eterminant), T(race), E(igen), V(ariance), and G-optimality. To establish a robust regression model of the formation energy w.r.t. data points, traditional method usually uses SYStematic enumeration (SYS) to select data points with monotonically increasing periodicities. Here the periodcity of data points can be understood as the "size". Exactly evaluating formation energy of a large size data point involves expensive computational cost. Therefore, SYS uses the minimum computational cost after observing certain number of data points, it may result in sub-optimal modeling in total practically.

In this paper, we formulate the problem as a cost-sensitive experimental design problem which aims at minimizing the optimality criteria and computational cost at the same time. The key motivation is that large periodcity data points may lead to good modeling in earlier

stages of the algorithm than SYS method. We design the Maximum Improvement Per Unit Cost (MI-PUC) algorithm, which runs sequentially to select data points to build the model. The key idea is to select the data point that leads to the maximum optimality criterion reduction per unit cost at each round.

Contributions. (1) We reformulate the formation energy modeling as a cost-sensitive experimental design problem, which aims at building a robust regression model and minimizing computational cost at the same time. (2) By selecting data point that leads to the maximum optimality criterion reduction per unit cost at each round, our MI-PUC algorithm adaptively selects data points for formation energy modeling. (3) We do experiments on a rocksalt-structured Zirconium Nitride dataset. The crystal structure of Rocksalt Zirconium Nitride resembles a checkerboard pattern, extended to three dimensions. Different data points in the experiment correspond to formation energies of different atomic arrangements on the crystal lattice. The results show that MI-PUC performs better than classical SYS method in all three evaluation criteria. Our approach provides material scientists with a rigorous benchmark of the performance-cost trade-off and a systematic way of attaining accurate atomistic models.

2. Problem Statement

Recent studies have described counting algorithms to systematically generate all possible arrangements of chemical elements on a crystal structure with a particular periodicity. These algorithms provide a route to exhaustively generating all data points that have periodicities smaller than a user defined maximum value. Given the large pool of chemical orderings with periodicities smaller than a user-defined maximum, it is desirable to find the smallest set of training points for which to perform expensive electronic structure calculations such that we are still able to obtain an accurate atomistic model, which is the cost-sensitive experimental design problem defined as follows.

Definition 1 (Cost-sensitive experimental design problem) *Let $x_1, ..., x_n \in \mathbb{R}^d$ be n data points, $c(x)$ be computational cost of x, and $f : \mathbb{S}_d^+ \to \mathbb{R}^+$ be a non-negative function defined over $\mathbb{S}_p^+$, the class of all d-dimensional positive definite matrices. Given a constant λ, the cost-sensitive experimental design problem is to select k data points such that:*

$$\min_{s \in \mathcal{S}_k} \; f\Big(\sum_{i=1}^{n} s_i \cdot x_i x_i^\top\Big) + \lambda \sum_{i=1}^{n} c(x_i), \qquad \text{s.t. } \mathcal{S}_k = \Big\{ s \in \{0,1\}^n, \sum_{i=1}^{n} s_i \le k \Big\}. \tag{1}$$

Let Σ denote the summation over data point outer products, then Eq. (1) becomes A-optimal and V-optimal design problem when $f(\Sigma) = \mathrm{tr}(\Sigma^{-1})$ and $f(\Sigma, X_{\text{test}}) = \mathrm{tr}(X_{\text{test}} \Sigma^{-1} X_{\text{test}}^\top)$, respectively. From material science literature (Kresse and Furthmüller, 1996), we know that given a data point x and its periodicity s_x, its computational cost $c(x) \propto s_x^2 \log(s_x)$.

3. Method

In this section, we first propose the Maximum Improvement Per Unit Cost (MI-PUC) algorithm. To run the algorithm in practice, we simplify the algorithm and introduce dimension reduction methods to address the singular matrix issue.

Algorithm 1 MI-PUC

Input: Initial selected points $\{X_0, Y_0\} = \{x_i, y_i\}_{i=1}^{N_0}$, unselected points $X_U = \{x_i\}_{i=1}^{N_U}$, testing set $X_{\text{test}} = \{x_i\}_{i=1}^{N_{\text{test}}}$, time horizon T.

1: **for** $t = [T]$ **do**
2: x_t is selected according to eq. (2) or eq. (3).
3: Query y_t for x_t and update $X_t \leftarrow \{X_{t-1}, x_t\}, Y_t \leftarrow \{Y_{t-1}, y_t\}$.
4: Update $X_U \leftarrow X_U \setminus x_t$.
5: **end for**

Output: A regression model $\hat{w}$ trained on $\{X_T, Y_T\}$.

3.1 Our Algorithm

MI-PUC is a greedy algorithm that runs in the online fashion. At each time step t, the algorithm select one data point x_t. The key idea of MI-PUC is to select data point that leads to the maximum optimality criterion reduction per unit cost. It works with all kinds of optimality criteria. For example, At round t, A-optimality of previous data points is defined as $\text{tr}((X_{t-1}^\top X_{t-1})^{-1})$. Given a new point x, the new A-optimality would become $\text{tr}(([X_{t-1}; x^\top]^\top [X_{t-1}; x^\top])^{-1})$. Therefore, the acquisition function of MI-PUC-A is

$$x_t = \underset{x \in X_U}{\text{argmax}} \frac{\text{tr}((X_{t-1}^\top X_{t-1})^{-1}) - \text{tr}(([X_{t-1}; x^\top]^\top [X_{t-1}; x^\top])^{-1})}{c(x)}. \tag{2}$$

Note V-optimality is able to access the testing set X_{test}. Similarly, the acquisition function of MI-PUC-V is

$$x_t = \underset{x \in X_U}{\text{argmax}} \frac{\text{tr}(X_{\text{test}}(X_{t-1}^\top X_{t-1})^{-1} X_{\text{test}}^\top) - \text{tr}(X_{\text{test}}([X_{t-1}; x^\top]^\top [X_{t-1}; x^\top])^{-1} X_{\text{test}}^\top)}{c(x)}. \tag{3}$$

3.2 Practical Considerations

Lemma 2 *Eq. (2) is equivalent to* $x_t = \text{argmax}_{x \in X_U} \frac{x^\top (X_{t-1}^\top X_{t-1})^{-2} x}{c(x)(1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x)}$, *and eq. (3) is equivalent to* $x_t = \text{argmax}_{x \in X_U} \frac{x^\top (X_{t-1}^\top X_{t-1})^{-1} X_{\text{test}}^\top X_{\text{test}} (X_{t-1}^\top X_{t-1})^{-1} x}{c(x)(1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x)}$.

The Sherman-Morrsion formula is used in the proof which is shown in Appendix B. The simplification allows us to save cost in acquisition function calculation. In equations above, $(X_{t-1}^\top X_{t-1})^{-1}$ is needed to be calculated only once, and then matrix computations in acquisition functions are performed only in dimension d rather than number of data points.

Also, in early stages of our algorithm, e.g., $t < d$, the dimension d maybe too high for very few data points, which leads to singular matrices. We use the Principal Component Analysis (PCA) method to do dimension reduction to avoid this problem.

Table 1: Statistics of datasets.

Periodicity s	1	2	3	4	5	6	7	8
No. of points N_s	2	2	6	19	28	80	102	388
Cumulative No. of points $N_{1:s}$	2	4	10	29	57	137	239	627
No. of initial selected points N_0					25	47	111	193
No. of unselected points N_U					478	456	392	310

4. Experiments

4.1 Experimental Settings

Dataset. Our dataset represents a collection of atomic configurations for Zirconium Nitride in a Rocksalt crystal structure. The crystal structure is composed of two sublattices; there is one sublattice occupied by Zirconium, and another occupied by Nitrogen. The Zirconium sublattice is left unchanged across all data points, but Nitrogen atoms can be removed. An absence of a Nitrogen atom is called a vacancy. Different arrangements of vacancies and Nitrogen atoms on the Nitrogen sublattice can produce different formation energies. The formation energy of an arrangement is related to the arrangement's thermodynamic stability. Each arrangement is represented by a data point in the "cluster" or "correlation" basis. The total number of data points is 627, each of which have 253 dimensions. To fairly compare our MI-PUC method with SYS, we randomly sample 20% of data points with periodicity $= 4, 5, 6, 7, 8$ reserved as testing set where $N_{\text{test}} = 124$. We repeat the whole process for 10 rounds and report mean performance. See Table 1 for more details.

Model training. We use *LASSO* in Scikit-Learn (Pedregosa et al., 2011) with fixed penalty parameter $\alpha = 0.000025$ to train the regression model, i.e., $\hat{w} = \text{argmin}_{w \in \mathbb{R}^d} \|Xw - y\|_2^2 + \alpha \|w\|_1^2$. Also, in our experiments, we set $T = N_U$ meaning all unselected points will be selected in the end of experiments.

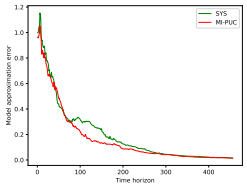
Evaluation criteria. Totally we use three criteria, all of which are the lower the better. The first criterion is Model Approximation Error (`MAE`). Ideally, we could generate infinite number of data points and their formation energies, which could lead to a perfect model w^*. However, given a fixed dataset, the best empirical model $\hat{w}^*$ can only be obtained through all data points in the set. Therefore, $\hat{w}^*$ is defined as the coefficient trained with all 627 data points in our problem, using the same *LASSO* model as above. Formally, we define `MAE` as the difference between $\hat{w}$ and $\hat{w}^*$ normalized by the magnitude of $\hat{w}^*$, which is between 0 and 1: $\texttt{MAE}(\hat{w}) = \|\hat{w} - \hat{w}^*\|_2^2 / \|\hat{w}^*\|_2^2$.

The second criterion is testing error (`TestErr`), which is measured as the Root Mean Square Error (RMSE) of $\hat{w}$ on the testing set: $\texttt{TestErr}(\hat{w}) = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (X_{\text{test},i} \hat{w} - y_{\text{test},i})^2}$.

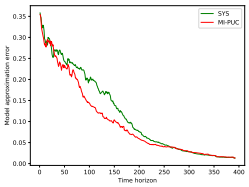
Our last criterion is ground state prediction error (`GSE`), which measures the fraction of incorrectly predicted ground states of $\hat{w}$. A ground state is defined as an atomic configuration on the lower convex hull of all data points in the space defined by composition and formation energy. If the simplices of the convex hull are defined by the outward-facing normal vectors, the lower hull is constructed by all simplices that have surface normals facing in the negative

formation energy direction. In the case of Zirconium Nitride, composition x is a scalar between 0 and 1 corresponding to the fraction of Nitrogen in the chemical formula ZrN_x. Let $S(\hat{w})$ denote the set of predicted ground states by $\hat{w}$ and S^* denote the set of true ground states. Formally, $\texttt{GSE}(\hat{w}) = |S(\hat{w}) - S^*|/|S^*|$.

4.2 Experimental Results

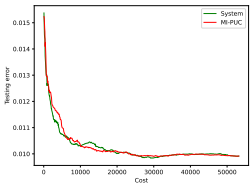


(a) 6

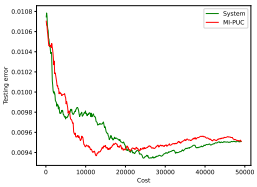


(b) 7

Figure 2: `MAE` of MI-PUC-V and SYS.



(a) 6



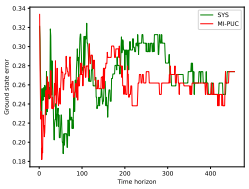
(b) 7

Figure 3: `TestErr` of MI-PUC-V and SYS

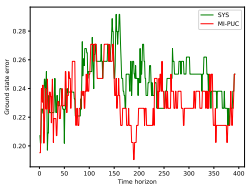
We mainly report the performance comparison between MI-PUC-V (working with V-optimality) and SYS with unselected data point periodicities starting from 6 and 7. In Appendix C, we show detailed results of the experiments, e.g., cost, periodicities, training error, and testing error w.r.t. time horizon.

From Figure 2, we learn that MI-PUC approximates the ideal model $\hat{w}^*$ significantly better than SYS, especially when round is between 80 and 200. In Figure 3, note the testing error is measured w.r.t. evaluation cost. MI-PUC performs similarly to SYS when unselected data point periodicity starts from 6, but it's different when starting from 7. In Figure 3(b), MI-PUC achieves the minimum testing error at cost of only $10,000$, where testing error of SYS is much higher than MI-PUC.

In Figure 4(a), the red line is lower than the green line, meaning MI-PUC predicts the ground states better than SYS. The phenomena is more obvious in Figure 4(b). All of these six figures show that MI-PUC is able to build a nearly optimal model at the early stages of the experiments. Moreover, in Appendix C, we report comparison between MI-PUC-A (working with A-optimality) and SYS, which shows similar results.



(a) 6



(b) 7

Figure 4: `GSE` of MI-PUC-V and SYS.

5. Impacts in Material Science

This study provides a first step in experimental design scheme that is sensitive to not only the quality of regression model but also computational cost of creating training dataset. Our results show that accounting for computational cost when choosing the next data point can result in more accurate fits at smaller costs than conventional methods while also being able to capture the essential physics of the underlying material system. Our work also serves as a benchmark of existing state-of-the art methods to explore the configurational space of alloys.

References

Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, 186(1):439–478, 2021.

Michal Černỳ and Milan Hladík. Two complexity results on c-optimality in experimental design. *Computational Optimization and Applications*, 51(3):1397–1408, 2012.

Alex Yong Sheng Eng, Chhail Bihari Soni, Yanwei Lum, Edwin Khoo, Zhenpeng Yao, SK Vineeth, Vipin Kumar, Jun Lu, Christopher S Johnson, Christopher Wolverton, et al. Theory-guided experimental design in battery materials research. *Science Advances*, 8 (19):1–17, 2022.

Peter I Frazier. Bayesian optimization. In *Recent advances in optimization and modeling of contemporary problems*, pages 255–278. Informs, 2018.

Konstantin Gubaev, Evgeny V Podryabinkin, Gus LW Hart, and Alexander V Shapeev. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials. *Computational Materials Science*, 156:148–156, 2019.

Ryosuke Jinnouchi, Jonathan Lahnsteiner, Ferenc Karsai, Georg Kresse, and Menno Bokdam. Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with bayesian inference. *Physical Review Letters*, 122:225701, 2019.

Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996.

Saehong Park, Dylan Kato, Zach Gima, Reinhardt Klein, and Scott Moura. Optimal experimental design for parameterization of an electrochemical lithium-ion battery model. *Journal of The Electrochemical Society*, 165(7):1309–1323, 2018.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

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, 2015.

Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics*, 21(1): 124–127, 1950.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *International Conference on Machine Learning*, 2010.

Appendix A. Related Work

Given a collection of data points, experimental design studies how to select a subset of data points to build a statistical efficient regression model, which is widely used in applied science (Park et al., 2018; Eng et al., 2022). The statistical efficiency is measured by some optimality criteria (Pukelsheim, 2006). Exact optimizing T-optimality is trivial however exact optimizing D/E-optimality is NP-hard (Černỳ and Hladík, 2012). Recently, Allen-Zhu et al. (2021) proposes a polynomial-time regret algorithm approximately solves the A/D/T/E/V/G-optimality problem.

Unlike experimental design, Bayesian optimization (Frazier, 2018; Shahriari et al., 2015) tries to optimize a black-box function defined on a continuous domain. It models the objective function with Gaussian process. Then at each round, it applies acquisition function to select data point to query. Popular choices of acquisition functions include expected improvement (Jones et al., 1998) and upper confidence bound (Srinivas et al., 2010). Our work is closely related to the expected improvement per unit cost (EI-PUC) algorithm (Snoek et al., 2012) which selects the maximum expected improvement per unit cost and efficiently solves the Bayesian optimization with constrained cost.

The challenge of accounting for computational cost while parameterizing atomistic models has largely remained uninvestigated in the field of materials science. Recent work by Gubaev et al. (2019) employs the D-optimality criteria with an active learning scheme to learn off-lattice interatomic models. With Gaussian approximation, Jinnouchi et al. (2019) uses machine learning-driven isothermal-isobaric simulations to give direct insight into the underlying microscopic mechanisms.

Appendix B. Missing Details in the Main Paper

Definition 3 (Optimality criteria) *The objective function f can be chosen from:*

$$f_A(\Sigma) = \frac{\operatorname{tr}(\Sigma^{-1})}{d}, \tag{4}$$

$$f_D(\Sigma) = (\det \Sigma)^{-\frac{1}{d}}, \tag{5}$$

$$f_T(\Sigma) = \frac{d}{\operatorname{tr}(\Sigma)}, \tag{6}$$

$$f_E(\Sigma) = \|\Sigma^{-1}\|_2, \tag{7}$$

$$f_V(\Sigma) = \frac{\operatorname{tr}(X\Sigma^{-1}X^\top)}{n}, \tag{8}$$

$$f_G(\Sigma) = \max \operatorname{diag}(X\Sigma^{-1}X^\top), \tag{9}$$

where $X = [x_1^\top, ..., x_n^\top]^\top \in \mathbb{R}^{n\times d}$ is the data matrix.

Lemma 4 (Restatement of Lemma 2) *The acquisition function of MI-PUC-A (eq. (2)) is equivalent to*

$$x_t = \operatorname*{argmax}_{x\in X_U} \frac{x^\top (X_{t-1}^\top X_{t-1})^{-2} x}{c(x)(1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x)}. \tag{10}$$

And the Acquisition function of MI-PUC-V (eq. (3)) is equivalent to

$$x_t = \underset{x \in X_U}{\operatorname{argmax}} \frac{x^\top (X_{t-1}^\top X_{t-1})^{-1} X_{\text{test}}^\top X_{\text{test}} (X_{t-1}^\top X_{t-1})^{-1} x}{c(x)(1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x)}. \tag{11}$$

Proof First we rewrite $([X_{t-1}; x^\top]^\top [X_{t-1}; x^\top])^{-1}$, which is

$$\left(\begin{bmatrix} X_{t-1} \\ x^\top \end{bmatrix}^\top \begin{bmatrix} X_{t-1} \\ x^\top \end{bmatrix} \right)^{-1} = \left(\begin{bmatrix} X_{t-1}^\top & x \end{bmatrix} \begin{bmatrix} X_{t-1} \\ x^\top \end{bmatrix} \right)^{-1} = (X_{t-1}^\top X_{t-1} + x x^\top)^{-1}. \tag{12}$$

The following proof relies on the Sherman-Morrison formula, shown below.

Lemma 5 (Sherman-Morrison Formula (Sherman and Morrison, 1950)) *Let A denote a matrix and b, c denote two vectors. Then*

$$(A + bc^\top)^{-1} = A^{-1} - \frac{A^{-1} b c^\top A^{-1}}{1 + c^\top A^{-1} b}. \tag{13}$$

Apply Sherman-Morrison Formula (Lemma 5),

$$([X_{t-1}; x^\top]^\top [X_{t-1}; x^\top])^{-1} = (X_{t-1}^\top X_{t-1})^{-1} - \frac{(X_{t-1}^\top X_{t-1})^{-1} x x^\top (X_{t-1}^\top X_{t-1})^{-1}}{1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x}. \tag{14}$$

Therefore, eq. (2) can be written as

$$\frac{\operatorname{tr}((X_{t-1}^\top X_{t-1})^{-1}) - \operatorname{tr}(([X_{t-1}; x^\top]^\top [X_{t-1}; x^\top])^{-1})}{\operatorname{cost}(x)} = \frac{\operatorname{tr}((X_{t-1}^\top X_{t-1})^{-1} x x^\top (X_{t-1}^\top X_{t-1})^{-1})}{\operatorname{cost}(x)(1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x)} \tag{15}$$

$$= \frac{\operatorname{tr}(x^\top (X_{t-1}^\top X_{t-1})^{-2} x)}{\operatorname{cost}(x)(1 + x^\top (X_{t-1}^\top X_{t-1})^{-1} x)}, \tag{16}$$

where the second equation is due to property of trace. The proof for MI-PUC-V works similarly. ■

Appendix C. More Experimental Results

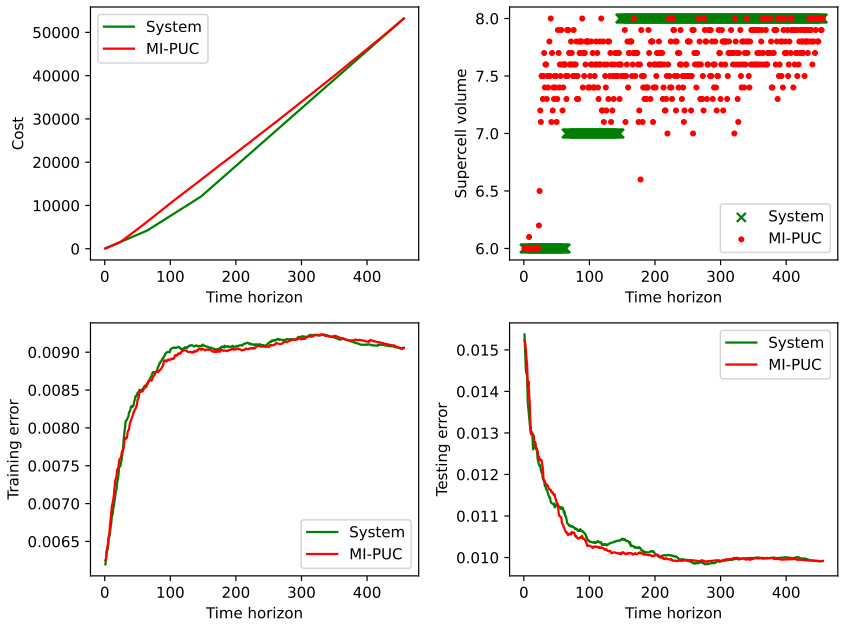


Figure 5: Details of MI-PUC-V with unselected periodicity (supercell volume) starting from 6.

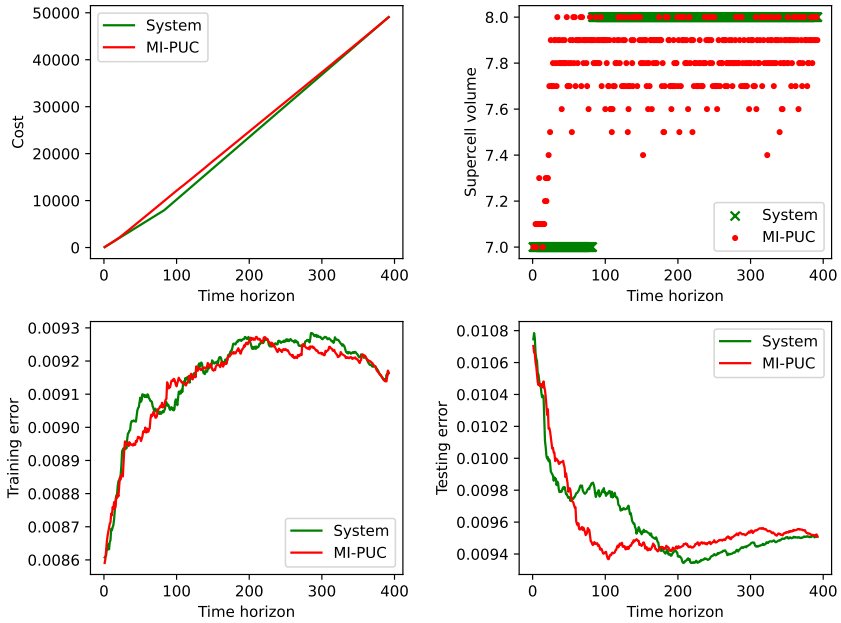


Figure 6: Details of MI-PUC-V with unselected periodicity (supercell volume) starting from 7.

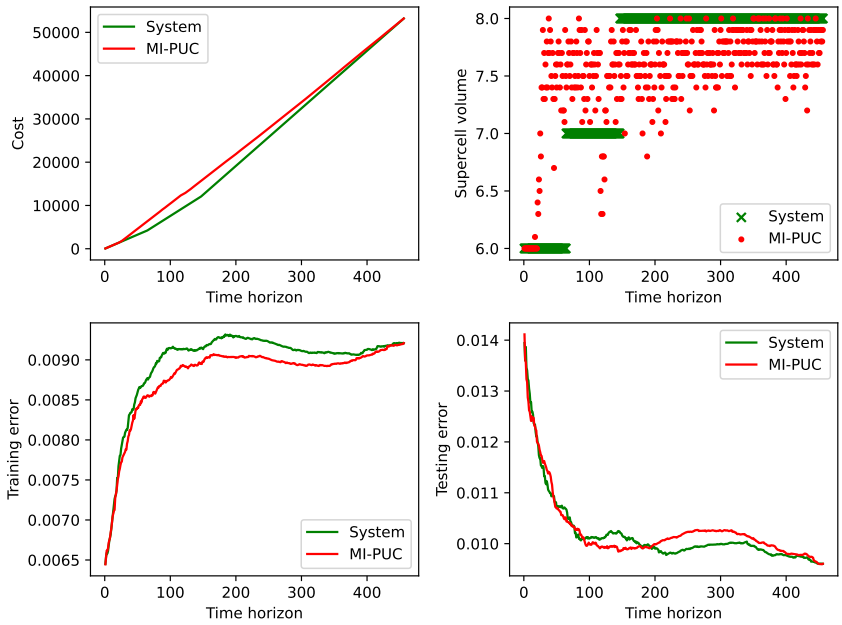


Figure 7: Details of MI-PUC-A with unselected periodicity (supercell volume) starting from 6.

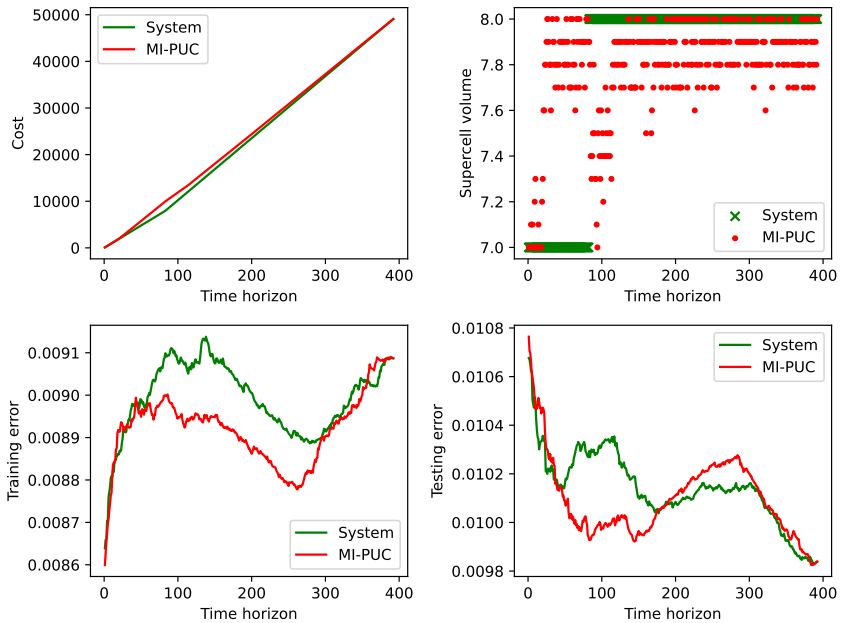
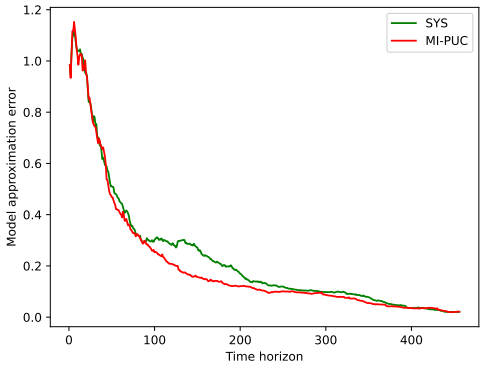
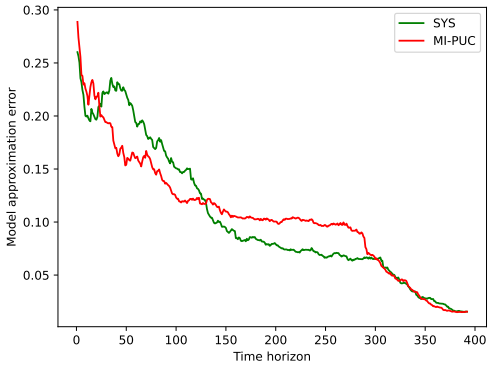


Figure 8: Details of MI-PUC-A with unselected periodicity (supercell volume) starting from 7.

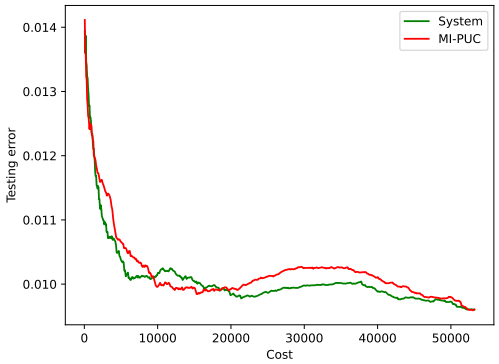


(a) Unselected periodicity starts from 6

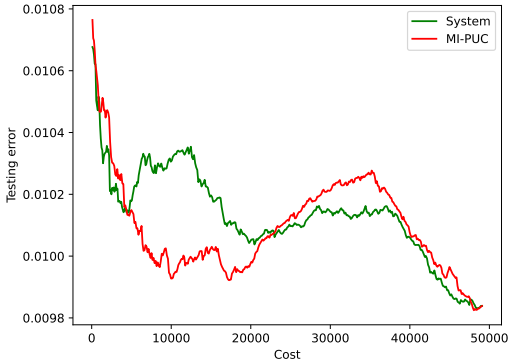


(b) Unselected periodicity starts from 7

Figure 9: `MAE` of MI-PUC-A and SYS.

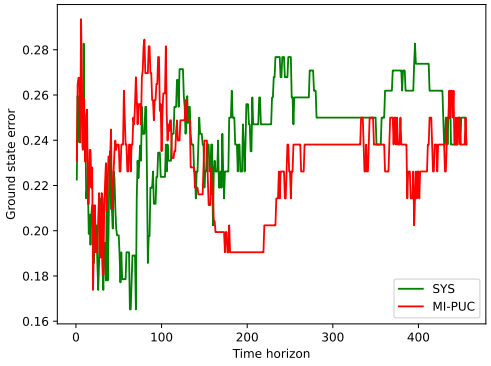


(a) Unselected periodicity starts from 6

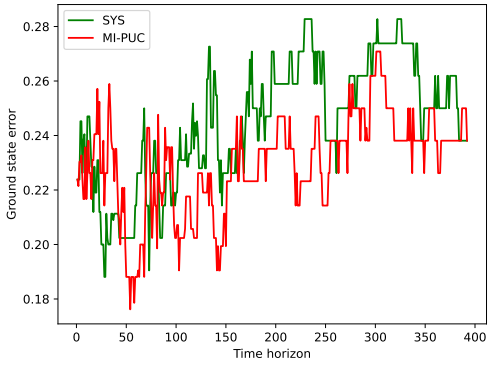


(b) Unselected periodicity starts from 7

Figure 10: `TestErr` of MI-PUC-A and SYS.



(a) Unselected periodicity starts from 6



(b) Unselected periodicity starts from 7

Figure 11: `GSE` of MI-PUC-A and SYS.