Adaptive Experimental Design for Optimizing Combinatorial Structures

Janardhan Rao Doppa

School of Electrical Engineering and Computer Science
Washington State University, Pullman, WA 99163
jana.doppa@wsu.edu

Abstract

Scientists and engineers in diverse domains need to perform expensive experiments to optimize combinatorial spaces, where each candidate input is a discrete structure (e.g., sequence, tree, graph) or a hybrid structure (mixture of discrete and continuous design variables). For example, in hardware design optimization over locations of processing cores and communication links for data transfer, design evaluation involves performing a computationally-expensive simulation. These experiments are often performed in a heuristic manner by humans and without any formal reasoning. In this paper, we first describe the key challenges in solving these problems in the framework of Bayesian optimization (BO) and our progress over the last five years in addressing these challenges. We also discuss exciting sustainability applications in domains such as electronic design automation, nanoporous materials science, biological sequence design, and electric transportation systems.

1 Introduction

Many design optimization problems in engineering and scientific domains are instantiations of the following general problem: *optimize the design of discrete (e.g., sequence and graph) and hybrid (mixture of discrete and continuous variables) structures guided by expensive experiments*, where the expense is measured by the resources consumed by the experiments. Consider the following examples. **1)** In designing application-specific hardware, we need to search over candidate placements of processing cores and communication links to optimize performance as measured by expensive computational simulations. **2)** To design drugs and vaccines, we need to search the space of molecules guided by physical lab experiments. **3)** In designing microbiomes for human health and agriculture applications, we need to search over subsets of microbial species (discrete variables) and environmental conditions (continuous variables) guided by lab experiments.

These experiments are often performed in a heuristic manner by the humans without formally reasoning about the available (computational or physical) resource budget and the usefulness of potential information that they may provide. There

is a great need for computational tools which will allow scientists and engineers to combine valuable domain knowledge and experimental data for automating *adaptive* experiment design. These tools will improve their productivity to achieve scientific discoveries in a resource-efficient manner. The overarching goals of this research program are to: **1)** develop general-purpose learning and reasoning algorithms to support engineers and scientists to select expensive experiments under a budget, and **2)** apply the algorithms to create high-impact applications via collaboration with domain experts from these application areas.

Bayesian optimization (BO) [Shahriari *et al.*, 2016] is an efficient framework for optimizing expensive functions. The key idea behind BO is to learn a cheap surrogate model, e.g., a Gaussian Process [Williams and Rasmussen, 2006] from past experimental data, which can be used to guide the selection of future experiments. Despite the huge successes of BO [Snoek *et al.*, 2012; Thornton *et al.*, 2013], current methods primarily focus on optimizing continuous spaces and there is little principled work on discrete and hybrid spaces. The key challenges behind BO over combinatorial spaces are analogous to those when we move from learning classifiers over simple outputs to structured outputs [Lafferty *et al.*, 2001; Taskar *et al.*, 2003; Tsochantaridis *et al.*, 2004; Roth and Yih, 2005; Daumé III *et al.*, 2009; Doppa *et al.*, 2014a; Doppa *et al.*, 2014b; Doppa *et al.*, 2014c]. The first challenge in moving from continuous spaces to discrete/hybrid spaces is to define an effective surrogate model over combinatorial structures. The second challenge is, given such a surrogate model, search the combinatorial space to select the most promising next structure for evaluation. Prior methods either employ simple surrogate models that admit tractable search or use complex models with heuristic search methods [Baptista and Poloczek, 2018; Hutter *et al.*, 2010; Hutter *et al.*, 2011]. To address these challenges, we developed novel approaches that can work with more complex models while following a more principled and effective search approach. We also describe our work on BO for optimizing multiple objectives, a relatively under-studied problem, and algorithms to leverage multi-fidelity experiments that trade-off accuracy and resource cost.

Finally, we discuss our work on exciting sustainability applications with high-societal impact in domains including electronic design automation (design of high-performance

and energy-efficient hardware), electric transportation systems (design of safe, low-cost, and efficient systems), nanoporous materials (design of materials to store, capture, and sense many different gases), and biological sequence design (design of DNA and proteins for medical applications).

2 Combinatorial Bayesian Optimization

Problem setup. Let $\mathcal{X}$ be a space of combinatorial structures. We assume the availability of a black-box objective function $f : \mathcal{X} \mapsto R$ defined over the combinatorial space $\mathcal{X}$. Evaluating each candidate structure $x \in \mathcal{X}$ using function f (also called an experiment) is *expensive* in terms of the resources consumed and produces an output $y = f(x)$. For example, in the drug design application, each $x \in \mathcal{X}$ is a molecule, and $f(x)$ corresponds to running a physical lab experiment. Our overall goal is to find a structure $x \in \mathcal{X}$ that approximately optimizes f by minimizing the number of experiments and observing their outcomes. To develop an effective BO algorithm, we need to instantiate the following three key elements: **1)** Statistical model of the black-box function $f(x)$, e.g., Gaussian process; **2)** Acquisition function to score the utility of evaluating candidate inputs using the statistical model, e.g., expected improvement; and **3)** Acquisition function optimizer to select the input with maximum utility.

BO over discrete spaces. In this setting, each input $x \in \mathcal{X}$ is a discrete structure (e.g., set, sequence, tree, graph). SMAC [Hutter *et al.*, 2010; Hutter *et al.*, 2011] is one canonical baseline which employs random forest as surrogate model and a *hand-designed* local search procedure for optimizing the expected improvement (EI) acquisition function. BOCS employs a linear Bayesian model over binary variables as the surrogate model, Thompson sampling (TS) as acquisition function, and semi-definite programming solution for acquisition function optimization. COMBO [Oh *et al.*, 2019] employs Gaussian process (GP) model with discrete diffusion kernels [Kondor and Lafferty, 2002] and performs local search to optimize EI. BO over discrete spaces has been reduced to BO over continuous spaces [Gómez-Bombarelli *et al.*, 2018]. The key idea is to employ an encoder-decoder architecture to learn a latent representation from data and perform BO in this latent space. Some of the main drawbacks of prior work include **1)** Can work with only limited acquisition functions; **2)** Ineffective search for acquisition function optimization with complex statistical models; and **3)** Cannot handle complex constraints, e.g., communication links should be placed such that all processing cores are connected to handle data transfer, to select only valid structures for evaluation.

We developed a learning-to-search framework referred to as L2S-DISCO [Deshwal *et al.*, 2020b] for selecting the sequence of combinatorial structures for evaluation. The key idea behind L2S-DISCO is to employ a combinatorial search procedure (e.g., local search with multiple restarts) guided by search control knowledge (e.g., heuristic function to select good starting states), and *continuously* improve the control knowledge using machine learning. L2S-DISCO can work with any choice of statistical model and acquisition function, and uses advances in machine learning to tune search-based optimizers on-the-fly to improve their accuracy in selecting

valid structures for evaluation. L2S-DISCO provides a new family of BO-style approaches with many instantiations to explore in future work. We studied a concrete instantiation of L2S-DISCO for local search procedure by specifying the form of training data, and a rank learning formulation to update the search heuristic to guide the selection of high-utility starting states.

We also designed a novel approach referred as MerCBO [Deshwal *et al.*, 2021c], which improves over COMBO to be able to perform tractable acquisition function optimization. COMBO employs a combinatorial graph representation of the input space $\mathcal{X}$ that allows using diffusion kernels to define smooth functions over $\mathcal{X}$ [Oh *et al.*, 2019]. MercBO computes a closed-form expression of this class of smooth functions in terms of explicit feature maps of diffusion kernels, which are referred as *Mercer features*. The key insight is to exploit the structure of the graph representation in COMBO to extract powerful low-dimensional features. *Mercer features can be seen as a bridge between BO over discrete spaces and BO over continuous spaces*. They allow us to leverage a large number of acquisition functions and algorithms from prior work on BO over continuous spaces to improve the BO performance for combinatorial spaces. In MerCBO, we employ Thompson sampling (TS) as the acquisition function by sampling parametric functions from GP posterior via Mercer features. We showed that the acquisition function optimization problem with TS is a Binary Quadratic Program (BQP). Inspired by the success of submodular relaxation in the structured prediction literature [Gorelick *et al.*, 2014], we studied a fast and scalable submdular relaxation method [Deshwal *et al.*, 2020a] to solve BQPs for selecting structures. One interesting future direction is to extend the construction of Mercer features to non-binary variables.

BO over hybrid spaces. In this setting, each input $x \in \mathcal{X}$ is a hybrid structure (mixture of discrete and continuous variables). There is very limited work on BO methods to optimize hybrid spaces [Hutter *et al.*, 2010; Hutter *et al.*, 2011; Daxberger *et al.*, 2020]. Most of them employ non-GP based surrogate models as it is challenging to define a generic kernel over hybrid spaces that can account for complex interactions between variables. To address this open challenge, our recent HyBO approach [Deshwal *et al.*, 2021b] builds GP based surrogate models using diffusion kernels, which are naturally defined over continuous and discrete spaces [Kondor and Lafferty, 2002]. We provide a principled approach that employs the general formulation of additive Gaussian process kernels [Duvenaud *et al.*, 2011] to define *additive hybrid diffusion* kernels. The key idea is to assign a base kernel for each discrete/continuous variable and construct an overall kernel by summing over all possible orders of interaction between these kernels. This construction procedure has two advantages: 1) Allows to leverage existing kernels for continuous and discrete spaces; and 2) Can automatically identify the strength of different orders of interaction in a data-driven manner for a given application. We also proved that the proposed hybrid diffusion kernel has universal approximation property, i.e., given sufficient data, it allows us to approximate any black-box function defined over hybrid spaces. One inter-

esting future direction is to use random Fourier features for continuous diffusion kernels and Mercer features for discrete diffusion kernels to devise a variant of HyBO with tractable acquisition function optimization.

3 Multi-Objective Bayesian Optimization

Problem setup. In many real-world applications, we need to optimize multiple objectives. For example, hardware design to optimize performance, power, and area. The key challenge is that objectives are conflicting in nature and they cannot be optimized simultaneously. Thus, we need to find the *Pareto optimal* set of solutions $\mathcal{X}^* \subset X$ and the corresponding set of function values referred as the optimal *Pareto front*. A solution is called Pareto optimal if it cannot be improved in any of the objectives without compromising some other objective. The goal of multi-objective BO is to approximate $\mathcal{X}^*$ by minimizing the number of expensive function evaluations. The most common setting is that each selected input $x \in \mathcal{X}$ is evaluated to measure *all* objective values.

Single-fidelity algorithms. In this setting, we assume that all objective function evaluations are accurate and expensive. There is relatively limited prior work on multi-objective BO. Many algorithms reduce the problem to single-objective optimization by designing appropriate acquisition functions, e.g., expected improvement in Pareto hypervolume [Knowles, 2006; Emmerich and Klinkenberg, 2008]. This can potentially lead to aggressive exploitation behavior. Additionally, algorithms to optimize Pareto hypervolume based acquisition functions scale poorly as the number of objectives and the dimensionality of input space grows. PESMO relies on *input space entropy* based acquisition function [Hernández-Lobato *et al.*, 2016] to select the candidate inputs for evaluation. However, it is computationally expensive to approximate and optimize this acquisition function.

We developed a simple and highly-effective approach referred to as USeMO [Belakaria *et al.*, 2020e]. The key insight behind USeMO is a two-stage search procedure to improve the accuracy and computational-efficiency of the reasoning process for selecting inputs for evaluation. First, it solves a cheap MOO problem defined in terms of the acquisition functions (one for each black-box objective) to identify a list of promising candidate inputs. Second, it selects the candidate from this list with maximum uncertainty. USeMO has several advantages: a) Does not reduce to single objective optimization problem; b) Allows to leverage acquisition functions designed for single objective BO and can be seen as a wrapper approach; and c) Computationally-efficient to solve problems with many objectives. We also proved asymptotic regret bounds for USeMO.

We studied another efficient approach referred as MESMO [Belakaria et al., 2019] based on the principle of *output space entropy (OSE) search* [Wang and Jegelka, 2017; Hoffman and Ghahramani, 2015]. MESMO is inspired by the prior success of the OSE principle for solving single-objective BO problems and is an extension of [Wang and Jegelka, 2017] for multi-objective optimization setting. The key idea is to select the input that maximizes the information gain about the optimal Pareto front in each iteration. Output space entropy search has many advantages over algorithms based on input space entropy search [Belakaria et al., 2019]: a) it allows much tighter approximation; b) it is cheaper to compute; and c) it naturally lends itself to robust optimization with respect to the number of Monte Carlo (MC) samples used for acquisition function computation. In fact, MESMO obtained excellent results even with a single MC sample!

Multi-fidelity algorithms. In the multi-fidelity setting, we assume the availability of function evaluations that vary in the amount of resources consumed and their accuracy. The overall goal is to approximate the optimal Pareto set of solutions by minimizing the resources consumed for experiments. For example, in power system design optimization, we need to find designs that trade-off cost, size, efficiency, and thermal tolerance using multi-fidelity simulators for design evaluations. This setting is challenging and hasn't been studied for multi-objective optimization problems. We extended our MESMO algorithm to both discrete-fidelity (small number of function approximations) [Belakaria *et al.*, 2020c] and continuous-fidelity (continuous function approximations which can even be infinite) [Belakaria *et al.*, 2020a] settings. The key idea is to select the candidate input and fidelity-vector (one for each objective) pair that maximizes the information gained about the true Pareto front per unit resource cost. We also provide multiple approximations to efficiently compute the entropy, which is a key step for these algorithms.

Constrained BO algorithms. In this setting, the solutions may need to satisfy black-box constraints, which cannot be evaluated without performing experiments. For example, in aviation power system design applications, we need to find the designs that trade-off total energy and mass while satisfying specific thresholds for motor temperature and voltage of cells (safety constraints). This setting is challenging and hasn't been studied for multi-objective optimization problems. We extended our USeMO and MESMO algorithms to handle constraints. The key idea is to build statistical models for each black-box constraint and use it as additional source of information to guide the selection of inputs for evaluation. For example, in USeMOC [Belakaria *et al.*, 2020d; Zhou *et al.*, 2020], we solve a cheap constrained MOO problem using these learned constraints to select promising candidates. Similarly, in MESMOC [Belakaria *et al.*, 2020b], we maximize the information gain between the next candidate input for evaluation and the constrained optimal Pareto front.

4 Sustainability Applications

In this section, we list some exciting and high-impact applications from diverse domains based on our BO algorithms.

Electronic design automation. Advanced computing systems have long been enablers for breakthroughs in science, engineering, and new technologies. However, with the slowing down of Moore's law and the relentless needs of Big-Data applications, e.g., deep learning and graph analytics, current solutions are not adequate. There is a need for innovative computer architectures and computationally efficient methods to design application-specific hardware systems ranging from IoT and mobile to manycores and datacenters to optimize performance, power consumption, and reliability [Ceze

et al., 2016; Doppa *et al.*, 2019].

Our research has addressed three out of five grand challenges posed by leading computer architects in the ARCH2030 vision paper [Ceze *et al.*, 2016; Hennessy and Patterson, 2018]. First, hardware design should be democratized so that developing specialized hardware for emerging applications is easy and fast. Second, 3D integration provides a new dimension of scalability in chip design to overcome the end of Moore's law. Third, machine learning for diverse applications has emerged as a key workload. We explored novel manycore architecture spaces enabled by heterogeneous integration of cores (e.g., CPUs, GPUs, TPUs, and ASICs) [Choi *et al.*, 2018]; 3D integration (Through-Silicon-Via and Monolithic 3D) for improved communication backbone [Musavvir *et al.*, 2020; Joardar *et al.*, 2021a; Joardar *et al.*, 2018a; Lee *et al.*, 2018; Lee *et al.*, 2019]; and Processing-in-memory to reduce the data movement by breaking the memory wall [Joardar *et al.*, 2021c; Arka *et al.*, 2021a]. We have applied our BO algorithms and domain-specific methods [Das *et al.*, 2017a; Joardar *et al.*, 2018b; Deshwal *et al.*, 2019] by leveraging prior knowledge for efficient design optimization [Das *et al.*, 2017b; Arka *et al.*, 2021b; Joardar *et al.*, 2021b]. We also developed machine learning algorithms for on-chip resource management to improve performance and energy-efficiency [Deshwal *et al.*, 2021a; Kim *et al.*, 2017; Mandal *et al.*, 2019; Mandal *et al.*, 2020].

Nanoporous materials design. Metal-organic frameworks (MOFs) are a class of crystalline materials that consist of coordination bonds between transition-metal cations and multidentate organic linkers. The structure of MOFs is characterized by an open framework that can be porous (porous materials) [Deshwal *et al.*, 2021d]. Currently, we are exploring the design of MOF materials for three applications: (i) sensing gases to construct an electronic nose to diagnose disease, and monitor pollution; (ii) separating gases cheaply and more energy-efficiently to reduce carbon dioxide emissions; (iii) storing gases; for example, storing hydrogen gas onboard vehicles for a clean, sustainable fuel source. We applied our BO algorithms to accelerate the search of optimized MOF materials using computationally-expensive molecular simulations [Deshwal *et al.*, 2021d]. Future challenges include constructing accurate MOF representations; exploring multi-fidelity algorithms using molecular simulations and real physical experiments; and AI algorithms for synthesis of MOF materials.

Biological sequence design. Design of optimized biological sequences such as DNA and proteins is a fundamental problem with many medical applications [Yang *et al.*, 2019]. The most popular design method in bio-molecular engineering is directed evolution, which explores sequence space by making small mutations to existing sequences. These design problems have some unique requirements: uncover a diverse set of sequences (*diversity*); select a large batch of sequences in each round to perform parallel evaluations (*large-scale parallel experiments*); and use parallel experimental resources to accelerate optimization (*real-time accelerated design*). Our MerCBO approach produced good results to address these challenges [Deshwal *et al.*, 2021c]. There is room for innovative batch selection methods to further improve diversity of high-performing sequences which scale to long sequences.

Electric transportation power systems. Electrification of vehicles introduces significant complexity to electrical system design. Electric transportation power systems design involves multiple inter-dependent subsystems and components. Hence, a large amount of engineering time and effort is needed to find optimized designs from a large search space. Multiphysics-based domain models and simulations are used to evaluate designs, which are computationally expensive (taking hours to days depending on the system complexity). We have applied our BO algorithms for multi-objective optimization to reduce the time and cost [Belakaria *et al.*, 2020f]. These advances have the potential to accelerate research and development throughout the transportation sector.

5 Conclusions and Future Work

Selecting expensive experiments to efficiently optimize combinatorial structures is an important and challenging problem with many real-world applications in science and engineering. This paper described the key challenges and our progress over the last five years focusing on Bayesian optimization based algorithms. We also listed several sustainability applications from diverse application domains where these BO algorithms have been applied and found to be effective.

There are many important directions for future work. First, creating effective representations by combining domain knowledge and data is the key to solve applications in domains such as catalysis and material science. Second, there are many open problems in the development of algorithms for batch BO, scalability to high-dimensional combinatorial spaces, and incorporating resource constraints [Dolatnia *et al.*, 2016]. Third, there are little to no theoretical studies to characterize the efficacy of BO algorithms for discrete and hybrid spaces. Fourth, availability of challenging benchmarks that are easy to use will allow the AI community to understand the strengths and weaknesses of existing methods to drive future research. Fifth, development of declarative languages to specify combinatorial spaces and domain constraints is both useful and necessary for adoption of this technology by engineers and scientists from application domains.

References

[Arka *et al.*, 2021a] Aqeeb Iqbal Arka, Biresh Kumar Joardar, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. ReGraphX: NoC-enabled 3D heterogeneous ReRAM architecture for training graph neural networks. In *DATE*, 2021.

[Arka *et al.*, 2021b] Aqeeb Iqbal Arka, Biresh Kumar Joardar, Ryan Gary Kim, Dae Hyun Kim, Janardhan Rao Doppa, and Partha Pratim Pande. HeM3D: Heterogeneous manycore architecture based on monolithic 3D vertical integration. *ACM Trans. Design Autom. Electr. Syst.*, 26(2):16:1–16:21, 2021.

[Baptista and Poloczek, 2018] Ricardo Baptista and Matthias Poloczek. Bayesian optimization of combinatorial structures. In *ICML*, pages 462–471, 2018.

[Belakaria *et al.*, 2020a] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Information-theoretic multi-objective Bayesian optimization with continuous approximations. *CoRR*, abs/2009.05700, 2020.

[Belakaria *et al.*, 2020b] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Max-value entropy search for multi-objective Bayesian optimization with constraints. *CoRR*, abs/2009.01721, 2020.

[Belakaria *et al.*, 2020c] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Multi-fidelity multi-objective Bayesian optimization: An output space entropy search approach. In *AAAI conference on Artificial Intelligence (AAAI)*, 2020.

[Belakaria *et al.*, 2020d] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Uncertainty aware search framework for multi-objective Bayesian optimization with constraints. *CoRR*, abs/2008.07029, 2020.

[Belakaria *et al.*, 2020e] Syrine Belakaria, Aryan Deshwal, Nitthilan Kannappan Jayakodi, and Janardhan Rao Doppa. Uncertainty-aware search framework for multi-objective Bayesian optimization. In *AAAI*, 2020.

[Belakaria *et al.*, 2020f] Syrine Belakaria, Derek Jackson, Yue Cao, Janardhan Rao Doppa, and Xiaonan Lu. Machine learning enabled fast multi-objective optimization for electrified aviation power system design. In *ECCE*, 2020.

[Belakaria et al., 2019] Belakaria et al. Max-value entropy search for multi-objective Bayesian optimization. In *NeurIPS*, 2019.

[Ceze *et al.*, 2016] Luis Ceze, Mark D. Hill, and Thomas F. Wenisch. Arch2030: A vision of computer architecture research over the next 15 years. *CoRR*, abs/1612.03182, 2016.

[Choi *et al.*, 2018] Wonje Choi, Karthi Duraisamy, Ryan Gary Kim, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. On-chip communication network for efficient training of deep convolutional networks on heterogeneous manycore systems. *IEEE Transactions on Computers (TC)*, 67(5):672–686, 2018.

[Das *et al.*, 2017a] Sourav Das, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. Design-space exploration and optimization of an energy-efficient and reliable 3D small-world network-on-chip. *IEEE TCAD*, 36(5), 2017.

[Das *et al.*, 2017b] Sourav Das, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. Monolithic 3d-enabled high performance and energy efficient network-on-chip. In *ICCD*, pages 233–240, 2017.

[Daumé III *et al.*, 2009] Hal Daumé III, John Langford, and Daniel Marcu. Search-based Structured Prediction. *MLJ*, 75(3), 2009.

[Daxberger *et al.*, 2020] Erik Daxberger, Anastasia Makarova, Matteo Turchetta, and Andreas Krause. Mixed-variable bayesian optimization. In *IJCAI*, pages 2633–2639, 2020.

[Deshwal *et al.*, 2019] Aryan Deshwal, Nitthilan Kannappan Jayakodi, Biresh Kumar Joardar, Janardhan Rao Doppa, and Partha Pratim Pande. MOOS: A multi-objective design space exploration and optimization framework for NoC enabled manycore systems. *ACM TECS*, 2019.

[Deshwal *et al.*, 2020a] Aryan Deshwal, Syrine Belakaria, and Janardhan Rao Doppa. Scalable combinatorial Bayesian optimization with tractable statistical models. *CoRR*, abs/2008.08177, 2020.

[Deshwal *et al.*, 2020b] Aryan Deshwal, Syrine Belakaria, Janardhan Rao Doppa, and Alan Fern. Optimizing discrete spaces via expensive evaluations: A learning to search framework. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[Deshwal *et al.*, 2021a] Aryan Deshwal, Syrine Belakaria, Ganapati Bhat, Janardhan Rao Doppa, and Partha Pratim Pande. Learning pareto-frontier resource management policies for heterogeneous socs: An information-theoretic approach. In *(DAC)*, 2021.

[Deshwal *et al.*, 2021b] Aryan Deshwal, Syrine Belakaria, and Janardhan Rao Doppa. Bayesian optimization over hybrid spaces. In *ICML*, 2021.

[Deshwal *et al.*, 2021c] Aryan Deshwal, Syrine Belakaria, and Janardhan Rao Doppa. Mercer features for efficient combinatorial Bayesian optimization. In *AAAI*, 2021.

[Deshwal *et al.*, 2021d] Aryan Deshwal, Cory Simon, and Janardhan Rao Doppa. Bayesian optimization of nanoporous materials. *ChemRxiv*, 2021.

[Dolatnia *et al.*, 2016] Nima Dolatnia, Alan Fern, and Xiaoli Z. Fern. Bayesian optimization with resource constraints and production. In *ICAPS*, pages 115–123, 2016.

[Doppa *et al.*, 2014a] Janardhan Rao Doppa, Alan Fern, and Prasad Tadepalli. HC-Search: A Learning Framework for Search-based Structured Prediction. *JAIR*, 2014.

[Doppa *et al.*, 2014b] Janardhan Rao Doppa, Alan Fern, and Prasad Tadepalli. Structured prediction via output space search. *JMLR*, 2014.

[Doppa *et al.*, 2014c] Janardhan Rao Doppa, Jun Yu, Chao Ma, Alan Fern, and Prasad Tadepalli. HC-Search for Multi-Label Prediction: An Empirical Study. In *AAAI*, 2014.

[Doppa *et al.*, 2019] Janardhan Rao Doppa, Justinian Rosca, and Paul Bogdan. Autonomous design space exploration of computing systems for sustainability: Opportunities and challenges. *IEEE Design and Test*, 36(5):35–43, 2019.

[Duvenaud *et al.*, 2011] David K Duvenaud, Hannes Nickisch, and Carl E Rasmussen. Additive Gaussian Processes. In *NIPS*, 2011.

[Emmerich and Klinkenberg, 2008] Michael Emmerich and Jan-willem Klinkenberg. The computation of the expected improvement in dominated hypervolume of pareto front approximations. *Technical Report, Leiden University*, 34, 2008.

[Gómez-Bombarelli *et al.*, 2018] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

[Gorelick *et al.*, 2014] Lena Gorelick, Yuri Boykov, Olga Veksler, Ismail Ben Ayed, and Andrew Delong. Submodularization for binary pairwise energies. In *CVPR*, pages 1154–1161, 2014.

[Hennessy and Patterson, 2018] John L. Hennessy and David A. Patterson. Turing lecture: A new golden age for computer architecture: Domain-specific hardware/software co-design, enhanced security, open instruction sets, and agile chip development. http://iscaconf.org/isca2018/turing_lecture.html, 2018. Accessed: 2021-06-03.

[Hernández-Lobato *et al.*, 2016] Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah, and Ryan Adams. Predictive entropy search for multi-objective Bayesian optimization. In *ICML*, pages 1492–1501, 2016.

[Hoffman and Ghahramani, 2015] Matthew W Hoffman and Zoubin Ghahramani. Output-space predictive entropy search for flexible global optimization. In *NIPS workshop on BO*, 2015.

[Hutter *et al.*, 2010] F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. Technical Report TR-2010-10, University of British Columbia, Department of Computer Science, 2010.

[Hutter *et al.*, 2011] Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *LION*, pages 507–523, 2011.

[Joardar *et al.*, 2018a] Biresh Kumar Joardar, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. Hybrid on-chip communication architectures for heterogeneous manycore systems. In *ICCAD*, 2018.

[Joardar *et al.*, 2018b] Biresh Kumar Joardar, Ryan Gary Kim, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. Learning-based application-agnostic 3D NoC design for heterogeneous manycore systems. *IEEE Transactions on Computers*, 68(6):852–866, 2018.

[Joardar *et al.*, 2021a] Biresh Kumar Joardar, Aqeeb Iqbal Arka, Janardhan Rao Doppa, and Partha Pratim Pande. 3D++: Unlocking the next generation of high-performance and energy-efficient architectures using M3D integration. In *DATE*, 2021.

[Joardar *et al.*, 2021b] Biresh Kumar Joardar, Aryan Deshwal, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. High-throughput training of deep CNNs on ReRAM-based heterogeneous architectures via optimized normalization layers. *IEEE TCAD*, 2021.

[Joardar *et al.*, 2021c] Biresh Kumar Joardar, Janardhan Rao Doppa, Partha Pratim Pande, Hai Li, and Krishnendu Chakrabarty. AccuReD: High accuracy training of cnns on ReRAM/GPU heterogeneous 3D architecture. *IEEE TCAD*, 40(5):971–984, 2021.

[Kim *et al.*, 2017] Ryan Gary Kim, Wonje Choi, Zhuo Chen, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. Imitation learning for dynamic VFI control in large-scale manycore systems. *TVLSI*, 25(9), 2017.

[Knowles, 2006] Joshua Knowles. Parego: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE TEC*, 10(1):50–66, 2006.

[Kondor and Lafferty, 2002] Risi Kondor and John Lafferty. Diffusion kernels on graphs and other discrete structures. In *ICML*, pages 315–322, 2002.

[Lafferty *et al.*, 2001] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *ICML*, pages 282–289, 2001.

[Lee *et al.*, 2018] Dongjin Lee, Sourav Das, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. Performance and thermal tradeoffs for energy-efficient monolithic 3D network-on-chip. *ACM TODAES*, 23(5):60:1–60:25, 2018.

[Lee *et al.*, 2019] Dongjin Lee, Sourav Das, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. Impact of electrostatic coupling on monolithic 3D-enabled network on chip. *ACM TODAES*, 24(6):62:1–62:22, 2019.

[Mandal *et al.*, 2019] Sumit Mandal, Ganapati Bhatt, Chetan Arvid Patel, Janardhan Rao Doppa, Partha Pratim Pande, and Umit Ogras. Dynamic resource management of heterogeneous mobile platforms via imitation learning. *IEEE TVLSI*, 2019.

[Mandal *et al.*, 2020] Sumit K. Mandal, Ganapati Bhat, Janardhan Rao Doppa, Partha Pratim Pande, and Ümit Y. Ogras. An energy-aware online learning framework for resource management in heterogeneous platforms. *ACM TODAES*, 25(3), 2020.

[Musavvir *et al.*, 2020] Shouvik Musavvir, Anwesha Chatterjee, Ryan Gary Kim, Dae Hyun Kim, Janardhan Rao Doppa, and Partha Pratim Pande. Power, performance, and thermal trade-offs in M3D-enabled manycore chips. In *DATE*, 2020.

[Oh *et al.*, 2019] Changyong Oh, Jakub Tomczak, Efstratios Gavves, and Max Welling. Combinatorial Bayesian Optimization using the Graph Cartesian Product. In *NeurIPS*, 2019.

[Roth and Yih, 2005] Dan Roth and Wen-tau Yih. Integer linear programming inference for conditional random fields. In *ICML*, 2005.

[Shahriari *et al.*, 2016] 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, 2016.

[Snoek *et al.*, 2012] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical Bayesian optimization of machine learning algorithms. In *NIPS*, pages 2960–2968, 2012.

[Taskar *et al.*, 2003] Benjamin Taskar, Carlos Guestrin, and Daphne Koller. Max-Margin Markov Networks. In *NIPS*, 2003.

[Thornton *et al.*, 2013] Chris Thornton, Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *KDD*, 2013.

[Tsochantaridis *et al.*, 2004] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. Support Vector Machine Learning for Interdependent and Structured Output Spaces. In *ICML*, 2004.

[Wang and Jegelka, 2017] Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient Bayesian optimization. In *ICML*, 2017.

[Williams and Rasmussen, 2006] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.

[Yang *et al.*, 2019] Kevin K Yang, Zachary Wu, and Frances H Arnold. Machine-learning-guided directed evolution for protein engineering. *Nature methods*, 16(8):687–694, 2019.

[Zhou *et al.*, 2020] Zhiyuan Zhou, Syrine Belakaria, Aryan Deshwal, Wookpyo Hong, Janardhan Rao Doppa, Partha Pratim Pande, and Deukhyoun Heo. Design of multi-output switched-capacitor voltage regulator via machine learning. In *DATE*, 2020.