Preference-Aware Constrained Multi-Objective Bayesian Optimization

Alaleh Ahmadianshalchi
Washington State University
Pullman, WA, USA
a.ahmadianshalchi@wsu.edu

Syrine Belakaria
Stanford University
Stanford, CA, USA
syrineb@cs.stanford.edu

Janrdhan Rao Doppa
Washington State University
Pullman, WA, USA
jana.doppa@wsu.edu

ABSTRACT

This paper addresses the problem of constrained multi-objective optimization over black-box objective functions with practitioner-specified preferences over the objectives when a large fraction of the input space is infeasible (i.e., violates constraints). This problem arises in many engineering design problems, including analog circuits and electric power system design. We aim to approximate the optimal Pareto set over the small fraction of feasible input designs. The key challenges include the massive size of the design space, multiple objectives, a large number of constraints, and the small fraction of feasible input designs, which can be identified only after performing expensive experiments/simulations. We propose a novel and efficient preference-aware constrained multi-objective Bayesian optimization approach referred to as PAC-MOO to address these challenges. The key idea is to learn surrogate models for both output objectives and constraints, and select the candidate input for evaluation in each iteration that maximizes the information gained about the optimal constrained Pareto front while factoring in the preferences over objectives. Our experiments on synthetic and challenging real-world analog circuit design optimization problems demonstrate the efficacy of PAC-MOO over baseline methods.

CCS CONCEPTS

• **Mathematics of computing** → **Stochastic processes**; **Bayesian computation**; • **Computing methodologies** → *Supervised learning by regression*; • **Applied computing** → *Computer-aided design.*

KEYWORDS

Blackbox function optimization, Bayesian optimization, probabilistic models, information theory, engineering design

ACM Reference Format:
Alaleh Ahmadianshalchi, Syrine Belakaria, and Janrdhan Rao Doppa. 2024. Preference-Aware Constrained Multi-Objective Bayesian Optimization. In *7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD) (CODS-COMAD 2024), January 04–07, 2024, Bangalore, India.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3632410.3632427

CODS-COMAD 2024, January 04–07, 2024, Bangalore, India

ACM ISBN 979-8-4007-1634-8/24/01.
https://doi.org/10.1145/3632410.3632427

1 INTRODUCTION

A large number of engineering design problems involve making design choices to optimize multiple objectives. Some examples include electric power systems design [10, 74], design of aircrafts [76], design of analog circuits and hardware accelerators [13, 16, 23, 43, 44, 50, 78, 79, 82], dynamic resource management in computing systems [22, 49, 56, 57, 60], and nanoporous materials discovery [24].

The common challenges in constrained multi-objective optimization (MOO) problems include the following. **1)** The objective functions are unknown and we need to perform expensive experiments to evaluate each candidate design choice. **2)** The objectives are conflicting in nature and all of them cannot be optimized simultaneously. **3)** The constraints need to be satisfied, but we cannot evaluate them for a given input design without performing expensive experiments. **4)** Only a small fraction of the input design space is feasible (i.e., satisfies all constraints). Therefore, we need to find the Pareto optimal set of solutions from this small subset of (unknown) feasible inputs, which is akin to finding needles in a haystack. Additionally, in several real-world applications, the practitioners have specific preferences over the objectives. For example, the designer prefers efficiency over settling time when optimizing analog circuits.

Bayesian optimization (BO) is an efficient framework to solve black-box optimization problems with expensive objective function evaluations [45, 65]. The key idea behind BO is to learn surrogate models (e.g., Gaussian processes [77]) of the expensive objective function and intelligently select the sequence of inputs for evaluation using an acquisition function (e.g., expected improvement [59]).

As we discuss in the related work, there are some BO algorithms for handling multiple expensive objective functions. However, there are no BO algorithms designed specifically for simultaneously handling the challenges of black-box constraints, a large fraction of invalid input space (where a considerable fraction of the input designs do not satisfy all constraints), and preferences over objectives.

To fill this important gap, we propose a novel and efficient information-theoretic approach referred to as **P***reference*-**A***ware* **C***onstrained* **M***ulti*-**O***bjective Bayesian* **O***ptimization (PAC-MOO)*[1]. PAC-MOO builds surrogate models for both output objectives and constraints based on the training data from past function evaluations. PAC-MOO employs an acquisition function in each iteration to select a candidate input design for performing expensive function evaluations. The selected input design maximizes the information gain about the constrained optimal Pareto front while factoring in the designer preferences over objectives. The experimental results on two challenging real-world analog circuit design benchmarks demonstrate that PAC-MOO was able to find circuit configurations

[1]A preliminary version of this paper [2] was presented at NeurIPS-2022 workshop on Machine Learning for Systems

with higher preferred objective values (efficiency) as intended by sacrificing the overall Pareto hypervolume indicator.

Contributions. Our key contribution is the development and evaluation of the PAC-MOO algorithm to solve a constrained multi-objective optimization problem. Specific contributions include:

- A tractable acquisition function based on information gain to select candidate inputs for expensive function evaluations.
- Approaches to increase the chances of finding feasible candidate designs and to incorporate preferences over objectives.
- Evaluation of PAC-MOO on real-world problems and comparison with prior methods.
- Our implementation for PAC-MOO is publicly available at https://github.com/Alaleh/PAC-MOO.

2 RELATED WORK

There are three families of approaches for solving constrained multi-objective optimization problems with expensive black-box functions. First, we can employ heuristic search algorithms such as multi-objective variants of simulated annealing [35, 38, 72], genetic algorithms [36, 37], and particle swarm optimization [29, 46, 68, 73] to solve them. The main drawback of this family of methods is that they require a large number of expensive function evaluations. Second, methods that build analytical models using a given functional form to model the black-box function behavior. Typical examples include approaches based on geometric programming [11, 21, 26, 47] and general polynomials [62, 75]. The key limitation of these methods is that the accuracy of solutions critically depends on the accuracy of analytical models over the entire input space, and the construction of accurate models requires a considerable number of expensive function evaluations. Third, Bayesian optimization (BO) methods employ surrogate statistical models to overcome the drawbacks of the previous families of approaches. The surrogate models are initialized using a small set of randomly sampled training data, i.e., input-output pairs of design parameters and objective evaluations. They are iteratively refined during the optimization process to actively collect a new training example in each iteration through an acquisition function (e.g., expected improvement). There is a large body of work on BO for single-objective optimization [25, 28]. Standard BO methods have been applied to a variety of problems including solving simple analog circuit design optimization and synthesis problems [41, 42, 52–55, 70, 71, 81].

Multi-objective BO (MOBO) is a relatively less-studied problem setting compared to the single-objective problem. Some of the recent work on MOBO include Predictive Entropy Search for Multi-objective Bayesian Optimization (PESMO) [39], Max-value Entropy Search for Multi-Objective Bayesian optimization (MESMO) [4, 6], Uncertainty-aware Search framework for Multi-Objective Bayesian Optimization (USEMO)[9], Pareto-Frontier Entropy Search (PFES) [66], Expected Hypervolume Improvement [17, 27], Multi-Objective Bayesian Optimization over High-Dimensional Search Spaces [19], and MVAR Approximation via Random Scalarizations [18]. Each of these methods has been shown to perform well on a variety of MOO problems. MESMO [4] is one of the state-of-the-art algorithms that is based on the principle of entropy search in the output space, which is low-dimensional compared to the input space.

Recent work extended existing approaches to the multi-objective constrained setting to account for black box constraints, notably Predictive Entropy Search for Multi-objective Bayesian Optimization with Constraints (PESMOC) [33], Parallel Predictive Entropy Search for Multi-objective Bayesian Optimization with Constraints (PPESMOC) [32], Max-value Entropy Search for Multi-Objective Bayesian Optimization with Constraints (MESMOC) [5, 8], and Uncertainty aware Search Framework for Multi-Objective Bayesian Optimization with Constraints (USEMOC) [7]. Existing algorithms can handle constraints that are evaluated using expensive function evaluations. However, they might not perform well when the fraction of feasible designs in the input space is small because they are hard to locate. Additionally, none of them supports preference specifications over the output objectives.

There has been a large body of work on incorporating preferences in multi-objective optimization using evolutionary techniques [12, 48, 69]. A parallel line of work includes several proposed optimization approaches to incorporate preferences between different objectives [1, 14, 34, 51, 63]. However, simultaneously handling constraints and preferences is not well-studied. The goal of this paper is to fill this gap motivated by real-world problems in analog circuit design and electric power systems design.

3 BACKGROUND AND PROBLEM SETUP

In this section, we provide background on the general BO framework and formally define the constrained MOO problem with preferences we are trying to solve.

3.1 Background on Bayesian Optimization

Bayesian Optimization (BO) is a general framework for solving expensive black-box optimization problems in a sample-efficient manner, i.e., minimizing the number of calls for expensive function evaluations. BO iterates within a feedback loop between (i) conducting expensive experiment with a candidate design parameters; (ii) updating our beliefs in the form of surrogate models about the relationship between design parameters and output objective(s); and (iii) selecting candidate design parameters for the next experiment.

The two key components of BO are surrogate model(s) and an acquisition function. The term "surrogate" in the surrogate model refers to "substitute for the expensive experiment". It is a probabilistic model (e.g., Gaussian process) to capture the relationship between the design parameters and output objectives, and is trained on all the training data from past expensive experiments. The surrogate model cheaply predicts the output objective(s) of the unevaluated design parameters and critically, quantifies the uncertainty in its predictions. The acquisition function (e.g., expected improvement [59]) is used to make the decision of which design parameters to select for the next simulation. It uses the surrogate model to score the utility of selecting design parameters for the next experiment by trading-off exploitation (select design parameters with high predictions from the model) and exploration (select design parameters for which the model is highly uncertain). The overall goal is to quickly direct the search towards high-quality design configurations.

Gaussian process for surrogate modeling. Gaussian processes (GPs) [77] are the most commonly used surrogate models in BO owing to their flexibility to approximate complex objectives, principled

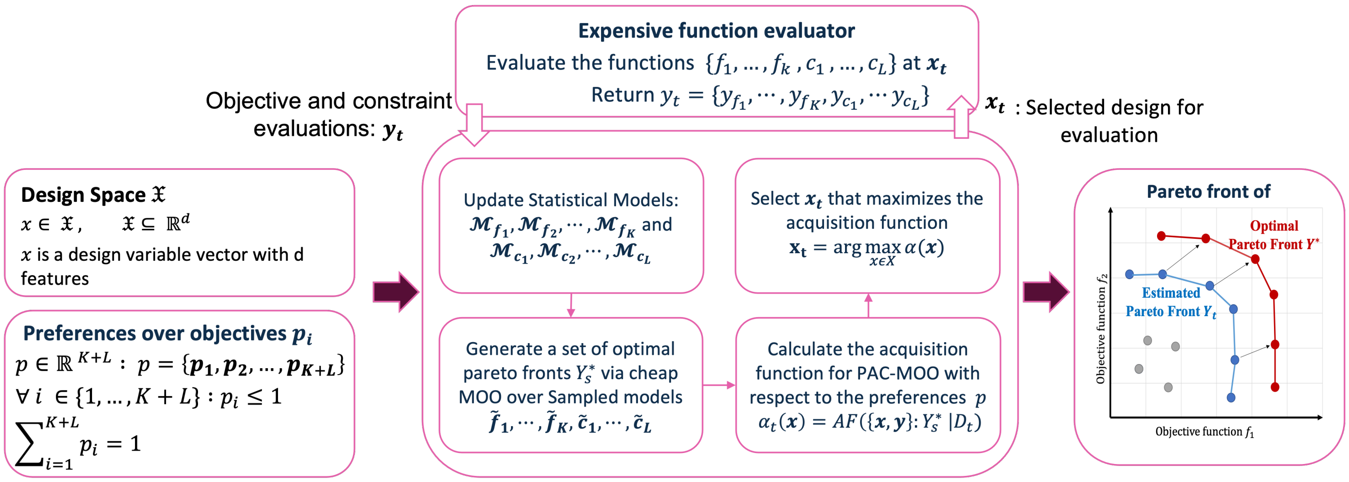


Figure 1: A high-level overview of the PAC-MOO algorithm. It takes as input the input space $\mathfrak{X}$ and preferences over objectives p, and produces a Pareto set of candidate points as per the preferences after T iterations of PAC-MOO. In each iteration t, PAC-MOO selects a candidate point $x_t \in \mathfrak{X}$ to perform expensive function evaluations and the surrogate models for both objective functions and constraints are updated based on training data from the evaluated point.

uncertainty quantification, and incorporating domain knowledge in the form of kernels to measure similarity between any two given inputs. A GP is a random process over domain space $\mathfrak{X} \rightarrow \mathcal{R}$ such that every finite collection of those random variables has a multivariate normal distribution; It can be characterized using its mean $\mu : \mathfrak{X} \rightarrow \mathcal{R}$ and its covariance or kernel function $\kappa : \mathfrak{X} \times \mathfrak{X} \rightarrow \mathcal{R}$. The posterior mean and standard deviation of a GP provide the prediction and uncertainty, respectively. Intuitively, uncertainty will be low for input design parameters $\mathbf{x} \in \mathfrak{X}$ that are close to the ones in our training data and will increase as the distance grows.

GPs provide a probabilistic framework for modeling the objective function. This probabilistic nature allows BO to measure uncertainty about the true function, which is crucial in making informed decisions about where to evaluate the function next. GPs are highly flexible models; They can adapt to a wide range of objective functions, including non-linear, non-convex, and multimodal functions. While they can be computationally demanding for large datasets, their expressiveness makes them valuable for tackling complex optimization problems. The point estimate and uncertainty provided by GPs, are crucial in real-world applications where making decisions under uncertainty is essential.

3.2 Problem Setup

Constrained multi-objective optimization w/ preferences. Constrained MOO is the problem of optimizing $\mathbf{K} \ge 2$ real-valued objective functions $\{f_1(x), \cdots, f_{\mathbf{K}}(x)\}$, while satisfying $\mathbf{L}$ black-box constraints of the form $c_1 \ge 0, \cdots, c_{\mathbf{L}}(x) \ge 0$ over the given design space $\mathfrak{X} \subset \mathbb{R}^d$. A function evaluation with the candidate parameters $\mathbf{x} \in \mathfrak{X}$ generates two vectors, one consisting of objective values and one consisting of constraint values $\mathbf{y} = (y_{f_1}, \cdots, y_{f_{\mathbf{K}}}, y_{c_1}, \cdots, y_{c_{\mathbf{L}}})$ where $y_{f_j} = f_j(x)$ for all $j \in \{1, \cdots, K\}$ and $y_{c_i} = C_i(x)$ for all $i \in \{1, \cdots, L\}$. We define an input vector $\mathbf{x}$ as feasible if and only if it satisfies all constraints. The input vector $\mathbf{x}$ *Pareto-dominates* another input vector $\mathbf{x}'$ if $f_j(\mathbf{x}) \le f_j(\mathbf{x}')$ $\forall j$ and there exists some $j \in \{1, \cdots, K\}$ such that $f_j(\mathbf{x}) < f_j(\mathbf{x}')$.

The optimal solution of the MOO problem with constraints is a set of input vectors $\mathcal{X}^* \subset \mathfrak{X}$ such that no configuration $\mathbf{x}' \in \mathfrak{X} \setminus \mathcal{X}^*$ Pareto-dominates another input $\mathbf{x} \in \mathcal{X}^*$ and all configurations in $\mathcal{X}^*$ are feasible. The solution set $\mathcal{X}^*$ is called the optimal constrained *Pareto set* and the corresponding set of function values $\mathcal{Y}^*$ is called the optimal constrained *Pareto front*. The most commonly used measure to evaluate the quality of a given Pareto set is by calculating the Pareto hypervolume (PHV) indicator [3] of the corresponding Pareto front of $(\mathbf{y}_{\mathbf{f}_1}, \mathbf{y}_{\mathbf{f}_2}, \cdots, \mathbf{y}_{\mathbf{f}_{\mathbf{K}}})$ with respect to a reference point $\mathbf{r}$. Our overall goal is to approximate the constrained Pareto set $\mathcal{X}^*$ by minimizing the total number of expensive function evaluations. When a preference specification p over the objectives is provided, the MOO algorithm should prioritize producing a Pareto set of inputs that optimize the preferred objective functions.

Preferences over black-box functions. The designer/practitioner can define input preferences over multiple black-box functions through the notion of preference specification, which is defined as a vector of scalars $\mathbf{p} = \{p_{f_1}, \cdots, p_{f_K}, p_{c_1}, \cdots, p_{c_L}\}$ with $0 \le p_i \le 1$ and $\sum_{i \in \mathcal{I}} p_i = 1$ such that $\mathcal{I} = \{f_1, \cdots, f_K, c_1, \cdots, c_L\}$. Higher values of p_i mean that the corresponding objective function f_i is highly preferred. In such cases, the solution to the MOO problem should prioritize producing design parameters that optimize the preferred objective functions.

4 PREFERENCE-AWARE CONSTRAINED MULTI-OBJECTIVE BO

The general strategy behind the BO process is to employ an acquisition function to iteratively select a candidate input (i.e., design parameters) to evaluate using the information provided by the surrogate models. The surrogate models are updated based on new training examples (design parameters as input, and evaluations of objectives and constraints from function evaluations as output).

Overview of PAC-MOO. PAC-MOO algorithm is an instance of the BO framework, which takes as input the input space $\mathfrak{X}$, preferences over objectives p, expensive objective functions and constraints evaluator, and produces a Pareto set of candidate inputs as per the preferences after T iterations of PAC-MOO as shown in Algorithm 1. In each iteration t, PAC-MOO selects a candidate input design $\mathbf{x}_t \in \mathfrak{X}$ to perform a function evaluation. Consequently, the surrogate models for both objective functions and constraints are updated based on training data from the function evaluations.

4.1 Surrogate Modeling

Gaussian Processes (GPs) [77] are suitable for solving black-box optimization problems with expensive function evaluations since they are rich and flexible models which can mimic any objective function. Intuitively, two candidate design parameters that are close to each other will potentially exhibit approximately similar performance in terms of output objectives. We model the objective functions and black-box constraints by independent GP models $\mathcal{GP}_{f_1}, \cdots, \mathcal{GP}_{f_K}, \mathcal{GP}_{c_1}, \cdots, \mathcal{GP}_{c_K}$ with zero mean and i.i.d. observation noise. Let $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{t-1}$ be the training data from past $t-1$ function evaluations, where $\mathbf{x}_i \in \mathfrak{X}$ is a candidate design and $\mathbf{y}_i = \{y^i_{f_1}, \cdots, y^i_{f_K}, y^i_{c_1}, \cdots, y^i_{c_L}\}$ is the output vector resulting from evaluating the objective functions and constraints at $\mathbf{x}_i$.

4.2 Acquisition Function

The state-of-the-art multi-objective optimization method MESMO approach for solving MOO problems [4] proposed to select the input that maximizes the information gain about the optimal **Pareto front** for evaluation. However, this approach did not address the challenge of handling black-box constraints which can be evaluated only through expensive function evaluators. To overcome this challenge, the MESMOC [5] algorithm was introduced which utilizes MESMO's powerful acquisition function while being able to incorporate constraint functions into the optimization by maximizing the information gain between the next candidate input for evaluation $\mathbf{x}$ and the optimal constrained Pareto front $\mathcal{Y}^*$:

$$\alpha(\mathbf{x}) = I(\{\mathbf{x}, \mathbf{y}\}, \mathcal{Y}^* \mid D) = H(\mathbf{y} \mid D, \mathbf{x}) - \mathbb{E}_{\mathcal{Y}^*}[H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{Y}^*)] \tag{1}$$

In this case, the output vector $\mathbf{y}$ is $K + L$ dimensional: $\mathbf{y} = (y_{f_1}, y_{f_2}, \cdots, y_{f_K}, y_{c_1}, \cdots, y_{c_L})$ where $y_{f_j} = f_j(x) \forall j \in \{1, 2, \cdots, K\}$ and $y_{c_i} = C_i(x) \forall i \in \{1, 2, \cdots, L\}$. Consequently, the first term in Equation (1), entropy of a factorizable $(K + L)$-dimensional Gaussian distribution $P(\mathbf{y} \mid D, \mathbf{x})$, can be computed in closed form as shown below:

$$H(\mathbf{y} \mid D, \mathbf{x}) = \frac{(K + C)(1 + \ln(2\pi))}{2} + \sum_{j=1}^{K} \ln(\sigma_{f_j}(\mathbf{x})) + \sum_{i=1}^{L} \ln(\sigma_{c_i}(\mathbf{x})) \tag{2}$$

where $\sigma^2_{f_j}(\mathbf{x})$ and $\sigma^2_{c_i}(\mathbf{x})$ are the predictive variances of j^{th} function and i^{th} constraint GPs respectively at input $\mathbf{x}$. The second term in Equation (1) is an expectation over the Pareto front $\mathcal{Y}^*$. We can approximately compute this term via Monte-Carlo sampling as shown below:

$$\mathbb{E}_{\mathcal{Y}^*}[H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{Y}^*)] \simeq \frac{1}{S} \sum_{s=1}^{S} [H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{Y}^*_s)] \tag{3}$$

where S is the number of samples and $\mathcal{Y}^*_s$ denote a sample Pareto front. There are two key algorithmic steps to compute this part of the equation: 1) How to compute constrained Pareto front samples $\mathcal{Y}^*_s$?; and 2) How to compute the entropy with respect to a given constrained Pareto front sample $\mathcal{Y}^*_s$? We provide solutions for these two questions below.

1) Computing constrained Pareto front samples via cheap multi-objective optimization. To compute a constrained Pareto front sample $\mathcal{Y}^*_s$, we first sample functions and constraints from the posterior GP models via random Fourier features [40, 64] and then solve a cheap constrained multi-objective optimization over the K sampled functions and L sampled constraints.

Cheap MO solver. We sample $\tilde{f}_i$ from GP model $\mathcal{GP}_{f_j}$ for each of the K functions and $\tilde{C}_j$ from GP model $\mathcal{GP}_{c_j}$ for each of the L constraints. A *cheap* constrained multi-objective optimization problem over the K sampled functions $\tilde{f}_1, \tilde{f}_2, \cdots, \tilde{f}_k$ and the L sampled constraints $\tilde{C}_1, \tilde{C}_2, \cdots, \tilde{C}_L$ is solved to compute the sample Pareto front $\mathcal{Y}^*_s$. Note that we refer to this optimization problem as cheap because it is performed over sampled functions and constraints, which are cheaper to evaluate than performing expensive function evaluations. We employ the popular constrained NSGA-II algorithm [20, 30] to solve the constrained MO problem with cheap sampled objective functions and constraints.

2) Entropy computation with a sample constrained Pareto front. Let $\mathcal{Y}^*_s = \{\mathbf{v}^1, \cdots, \mathbf{v}^l\}$ be the sample constrained Pareto front, where l is the size of the Pareto front and each $\mathbf{v}^i$ is a $(K + L)$-vector evaluated at the K sampled functions and L sampled constraints $\mathbf{v}^i = \{v^i_{f_1}, \cdots, v^i_{f_K}, v^i_{c_1}, \cdots, v^i_{c_L}\}$. The following inequality holds for each component y_j of the $(K + L)$-vector $\mathbf{y} = \{y_{f_1}, \cdots, y_{f_K}, y_{c_1}, \cdots y_{c_L}\}$ in the entropy term $H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{Y}^*_s)$:

$$y_j \leq \max\{v^1_j, \cdots v^l_j\} \quad \forall j \in \{f_1, \cdots, f_K, c_1, \cdots, c_L\} \tag{4}$$

The inequality essentially says that the j^{th} component of $\mathbf{y}$ (i.e., y_j) is upper-bounded by a value obtained by taking the maximum of j^{th} components of all l $(K+L)$-vectors in the Pareto front $\mathcal{Y}^*_s$. This inequality had been proven by a contradiction for MESMO [4] for all objective functions $j \in \{f_1, \cdots, f_K\}$. We assume the same for all constraints $j \in \{c_1, \cdots, c_L\}$.

By combining the inequality (4) and the fact that each function is modeled as an independent GP, we can approximate each component y_j as a truncated Gaussian distribution since the distribution of y_j needs to satisfy $y_j \leq \max\{v^1_j, \cdots v^l_j\}$. Let $y_s^{c_i *} = \max\{v^1_{c_i}, \cdots v^l_{c_i}\}$ and $y_s^{f_j *} = \max\{v^1_{f_j}, \cdots v^l_{f_j}\}$. Furthermore, a common property of entropy measure allows us to decompose the entropy of a set of independent variables into a sum over entropies of individual variables [15]:

$$H(\mathbf{y} \mid D, \mathbf{x}, \mathcal{Y}^*_s) = \sum_{j=1}^{K} H(y_{f_j} | D, \mathbf{x}, y_s^{f_j *}) + \sum_{i=1}^{L} H(y_{c_i} | D, \mathbf{x}, y_s^{c_i *}) \tag{5}$$

The r.h.s is a summation over entropies of $(K + L)$-variables $\mathbf{y} = \{y_{f_1}, \cdots, y_{f_K}, y_{c_1}, \cdots y_{c_L}\}$. The differential entropy for each y_j is the entropy of a truncated Gaussian distribution [58] and is given by the following equations:

$$H(y_{f_j}|D,\mathbf{x},y_s^{f_j*}) \simeq \left[\frac{(1+\ln(2\pi))}{2} + \ln(\sigma_{f_j}(\mathbf{x})) + \ln\Phi(\gamma_s^{f_j}(\mathbf{x})) - \frac{\gamma_s^{f_j}(\mathbf{x})\phi(\gamma_s^{f_j}(\mathbf{x}))}{2\Phi(\gamma_s^{f_j}(\mathbf{x}))}\right] \tag{6}$$

$$H(y_{c_i}|D,\mathbf{x},y_s^{c_i*}) \simeq \left[\frac{(1+\ln(2\pi))}{2} + \ln(\sigma_{c_i}(\mathbf{x})) + \ln\Phi(\gamma_s^{c_i}(\mathbf{x})) - \frac{\gamma_s^{c_i}(\mathbf{x})\phi(\gamma_s^{c_i}(\mathbf{x}))}{2\Phi(\gamma_s^{c_i}(\mathbf{x}))}\right] \tag{7}$$

Consequently, we have:

$$H(\mathbf{y} \mid D,\mathbf{x},\mathcal{Y}_s^*) \simeq \sum_{j=1}^{K}\left[\frac{(1+\ln(2\pi))}{2} + \ln(\sigma_{f_j}(\mathbf{x})) + \ln\Phi(\gamma_s^{f_j}(\mathbf{x})) - \frac{\gamma_s^{f_j}(\mathbf{x})\phi(\gamma_s^{f_j}(\mathbf{x}))}{2\Phi(\gamma_s^{f_j}(\mathbf{x}))}\right] + \sum_{i=1}^{L}\left[\frac{(1+\ln(2\pi))}{2} + \ln(\sigma_{c_i}(\mathbf{x})) + \ln\Phi(\gamma_s^{c_i}(\mathbf{x})) - \frac{\gamma_s^{c_i}(\mathbf{x})\phi(\gamma_s^{c_i}(\mathbf{x}))}{2\Phi(\gamma_s^{c_i}(\mathbf{x}))}\right] \tag{8}$$

where $\gamma_s^{c_i}(x) = \frac{y_s^{c_i*}-\mu_{c_i}(\mathbf{x})}{\sigma_{c_i}(\mathbf{x})}$, $\gamma_s^{f_j}(x) = \frac{y_s^{f_j*}-\mu_{f_j}(\mathbf{x})}{\sigma_{f_j}(\mathbf{x})}$, and ϕ and Φ are the p.d.f and c.d.f of a standard normal distribution respectively. By combining equations (2) and (8) with equation (1), we get the final form of our acquisition function as shown below:

$$\alpha(\mathbf{x}) \simeq \sum_{i\in\mathcal{I}} AF(i,x) \text{ with } i \in \mathcal{I} \text{ and } \mathcal{I} = \{c_1 \cdots c_L, f_1 \cdots f_K\} \tag{9}$$

$$AF(i,x) = \sum_{s=1}^{S} \frac{\gamma_s^i(\mathbf{x})\phi(\gamma_s^i(\mathbf{x}))}{2\Phi(\gamma_s^i(\mathbf{x}))} - \ln\Phi(\gamma_s^i(\mathbf{x})) \tag{10}$$

4.3 Convex Combination for Preferences

We now describe how to incorporate preference specification (when available) into the acquisition function. The derivation of the acquisition function proposed in Equation 9 resulted in a function in the form of a summation of an entropy term defined for each of the objective functions and constraints as $AF(i,x)$. Given this expression, the algorithm will select an input while giving the same importance to each of the functions and constraints. However, as an example, in problems such as circuit design optimization, efficiency is typically the most important objective function. The designer would like to find a trade-off between the objectives. Nevertheless, candidate circuits with high voltage and very low efficiency might be useless in practice. Therefore, we propose to inject preferences from the designer into our algorithm by associating different weights to each of the objectives. A principled approach would be to assign appropriate preference weights resulting in a convex combination of the individual components of the summation $AF(i,x)$. Let p_i be the preference weight associated with each individual component. The preference-based acquisition function is defined as follows (see Algorithm 2):

$$\alpha_{pref}(\mathbf{x}) \simeq \sum_{i\in\mathcal{I}} p_i \times AF(i,x) \text{ with } i \in \mathcal{I} \quad s.t \sum_{i\in\mathcal{I}} p_i = 1 \tag{11}$$

It is important to note that in practice if a candidate design does not satisfy the constraints, the optimization will fail regardless of the preferences over objectives. Therefore, the cumulative weights assigned to the constraints have to be at least equal to the total weight assigned to the objective functions:

$$\sum_{i\in\{c_1\cdots c_L\}} p_i = \sum_{i\in\{f_1,\cdots,f_K\}} p_i = \frac{1}{2} \tag{12}$$

Given that satisfying all the constraints is equally important, the weights over the constraints would be equal. Finally, only the weights over the functions will need to be explicitly specified.

Algorithm 1 PAC-MOO Algorithm

Inputs: Input space $\mathcal{X}$, black-box functions $\{f_1, ..., f_K\}$, constraint functions $\{c_1, ..., c_L\}$, preferences $\mathbf{p} = \{p_{f_1}, \cdots, p_{f_K}, p_{c_1}, \cdots, p_{c_L}\}$, number of initial points $\mathcal{N}_0$, number of iterations T

Initialize Gaussian processes for functions $\mathcal{M}_{f_1}, \cdots, \mathcal{M}_{f_K}$ and constraints $\mathcal{M}_{c_1}, \cdots, \mathcal{M}_{c_L}$ by evaluating them on $\mathcal{N}_0$ initial design parameters
for each iteration t = $\mathcal{N}_0$ to T+$\mathcal{N}_0$ **do**
  if feasible design parameters $\mathbf{x}_{feasible} \notin \mathcal{D}$ **then**
    Select design parameters $\mathbf{x}_t \leftarrow \arg max_{\mathbf{x}\in\mathcal{X}} \ \alpha_{prob}(\mathbf{x})$ *# eq. 13*
  else
    Select design parameters $\mathbf{x}_t \leftarrow \arg max_{\mathbf{x}\in\mathcal{X}} \ \alpha_{pref}(\mathbf{x},\mathbf{p})$ in Algorithm 2 **s.t** $(\mu_{c_1} \geq 0, \cdots, \mu_{c_L} \geq 0)$
  end if
  Perform function evaluations using the selected design parameters
    $\mathbf{x}_t$: $\mathbf{y}_t \leftarrow (f_1(\mathbf{x}_t), \cdots, f_K(\mathbf{x}_t), C_1(\mathbf{x}_t), \cdots, C_L(\mathbf{x}_t))$
  Aggregate data: $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\mathbf{x}_t, \mathbf{y}_t)\}$
  Update models $\mathcal{M}_{f_1}, \cdots, \mathcal{M}_{f_K}$ and $\mathcal{M}_{c_1}, \cdots, \mathcal{M}_{c_L}$ using $\mathcal{D}$
end for
return the Pareto set of feasible design parameters from $\mathcal{D}$

Algorithm 2 Preference based Acquisition function (α_{pref})

$\alpha_{pref}(\mathbf{x},\mathbf{p})$
for Each sample $s \in \{1, \cdots, S\}$ **do**
  Sample functions $\tilde{f}_j \sim \mathcal{M}_{f_j}, \quad \forall j \in \{1, \cdots, K\}$
  Sample constraints $\tilde{C}_i \sim \mathcal{M}_{c_i}, \quad \forall i \in \{1, \cdots, L\}$
  Solve *cheap* MOO over $(\tilde{f}_1, \cdots, \tilde{f}_K)$ constrained by $(\tilde{C}_1, \cdots, \tilde{C}_L)$
  $\mathcal{Y}_s^* \leftarrow \arg max_{x\in\mathcal{X}}(\tilde{f}_1, \cdots, \tilde{f}_K)$ **s.t** $(\tilde{C}_1 \geq 0, \cdots, \tilde{C}_L \geq 0)$
end for
for $i \in \mathcal{I}$ and $\mathcal{I} = \{c_1 \cdots c_L, f_1 \cdots f_K\}$ **do**
  Compute $AF(i,x)$ based on S samples of $\mathcal{Y}_s^*$ via Equation 10
end for
Return $\sum_{i\in\mathcal{I}} p_i \times AF(i,x)$

4.4 Finding Feasible Regions of Design Space

The acquisition function defined in equation 11 will build constrained Pareto front samples $\mathcal{Y}_s^*$ by sampling functions and constraints from the Gaussian process posterior. The posterior of the GP is built based on the current training data $\mathcal{D}$. The truncated

Gaussian approximation defined in Equations 6 and 7 requires the upper bound $y_s^{f_j*}$ and $y_s^{c_i*}$ to be defined. However, in the early Bayesian optimization iterations of the algorithm, the configurations evaluated may not include any feasible design parameters. This is especially true for scenarios where the fraction of feasible design configurations in the entire design space is very small. In such cases, the sampling process of the constrained Pareto fronts $\mathcal{Y}_s^*$ is susceptible to failure because the surrogate models did not gather any knowledge about feasible regions of the design space *yet*. Consequently, the upper bounds $y_s^{f_j*}$ and $y_s^{c_i*}$ are not well-defined and the acquisition function in 11 is not well-defined. Intuitively, the algorithm should first aim at identifying feasible design configurations by maximizing the probability of satisfying all the constraints. We define a special case of our acquisition function for such challenging scenarios as shown below:

$$\alpha_{prob}(x) = \prod_{i=1}^{L} Pr(c_i(x) \geq 0) \tag{13}$$

This acquisition function enables an efficient feasibility search due to its exploitation characteristics [31]. Given that the probability of constraint satisfaction is binary (0 or 1), the algorithm will be able to quickly prune unfeasible regions of the design space and move to other promising regions until it identifies feasible design configurations. This approach will enable a more efficient search over feasible regions later and accurate computation of the acquisition function. The complete pseudo-code of PAC-MOO is given in Algorithm 1.

4.5 Regret Bound Analysis

The state-of-the-art MESMO approach [4] for multi-objective BO, possesses a regret analysis for multi-objective cumulative regret. Given well-fitted Gaussian processes for the objective functions and constraints, the cheap multi-objective algorithm will be able to generate designs that are Pareto optimal and satisfy the constraints with high probability. Under these conditions, the regret bound developed in MESMO also extends to MESMOC [5, 8]. Given the incorporation of the preference vector within the PAC-MOO framework, the theoretical regret bound takes on a distinct form. Specifically, the regret analysis for PAC-MOO yields a weighted summation of the regret for each individual objective, reflecting the preferences encoded by the user. This extension of regret formulation allows PAC-MOO to hold its regret bound in a diverse range of multi-objective optimization scenarios, where objectives may not possess uniform importance. Therefore, the regret analysis of PAC-MOO degenerates to MESMO's regret analysis in the absence of preferences and constraints. This theoretical insight underscores the versatility and practicality of PAC-MOO, as it extends the utility of prior regret analysis to settings where user preferences play a pivotal role in guiding the optimization process.

5 EXPERIMENTAL RESULTS

In this section, we present experimental evaluation of PAC-MOO and baseline methods on a synthetic constrained multi-objective optimization problem as well as two challenging real-world analog circuit design problems.

5.1 Experimental Setup

Baselines. We compare PAC-MOO with state-of-the-art constrained MOO evolutionary algorithms, namely, NSGA-II [20] and MOEAD [80]. We also compare to the constrained multi-objective optimization method, the Uncertainty aware search framework for multi-objective Bayesian optimization with constraints (USEMOC) algorithm [7]. We evaluated two variants of the USEMOC wrapper framework: USEMOC-EI and USEMOC-TS, using expected improvement (EI) and Thompson sampling (TS) acquisition functions. We also compare our method to the Max-value Entropy Search for Multi-Objective Bayesian Optimization with Constraints (MESMOC) [5] which is a special case of the PAC-MOO problem: equal preferences over all the black-box functions and without the approach to find feasible regions of input space (Section 4.4). Note that the acquisition function behind MESMOC is not well-defined when there are no training examples for feasible inputs. Hence, our PAC-MOO-0 (PAC-MOO with no preferences) mitigates the feasibility issue of MESMOC, by incorporating our proposed approach to handle lack of feasible training examples.

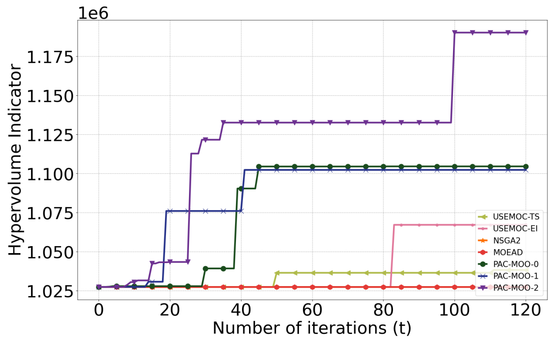
PAC-MOO: We employ a Gaussian process (GP) with squared exponential kernel for all our surrogate models. We evaluated several preference values for the efficiency objective function. PAC-MOO-0 refers to the preference being equal over all objectives and constraints. PAC-MOO-1 refers to assigning 80% preference to one preferred objective (e.g. efficiency in the HCR and SCVR problems) and equal importance to other functions and constraints, resulting in a preference value $p_i = 0.5 \times 0.8 = 0.4$ for the preferred objective. With PAC-MOO-2, we assign a total preference of 85% to the objective functions with 92% importance to one preferred objective resulting in a preference value of $p_i = 0.85 \times 0.92 = 0.782$. We assign equal preference to all other functions. With PAC-MOO-3, we assign a total of 0.65 preference to the objective functions and 0.35 to the constraints. Additionally, we provide 88% importance to the preferred objective resulting in a preference value of $p_i = 0.65 \times 0.88 = 0.572$. The implementation of our algorithm is included in the link below[2].

Evaluation Metrics: The *Pareto Hypervolume (PHV)* indicator is a commonly used metric to measure the quality of the Pareto front [84]. PHV is defined as the volume between a reference point and the Pareto front. After each expensive experiment (iteration), we measure the PHV for all algorithms and compare them. To demonstrate the efficacy of the preference-based PAC-MOO for real-world analog circuit design problems, we compare different algorithms using the maximum discovered efficiency of the optimized circuit configurations as a function of the number of circuit simulations.

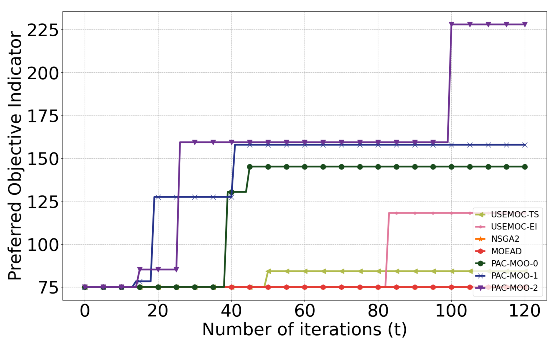
Benchmarks: We employ one synthetic and two challenging real-word engineering design problems to show the efficacy of the proposed PAC-MOO method.

1.*The OSY problem.* We solve the constrained multi-objective optimization OSY test problem [61] as a synthetic benchmark with a minor modification. To increase the complexity of the problem, each dimension in the input space is expanded to 1.5 times its original size, resulting in a search space approximately 11 times larger than the original problem. Additionally, we introduce a new

[2] https://github.com/Alaleh/PAC-MOO

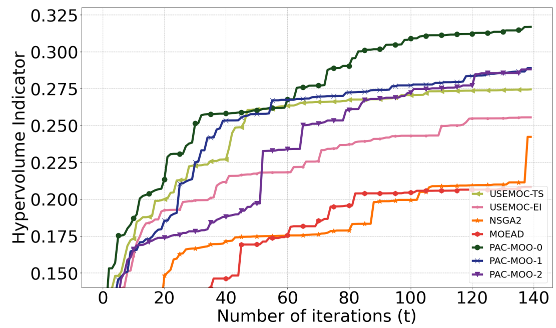


(a) Hypervolume - OSY

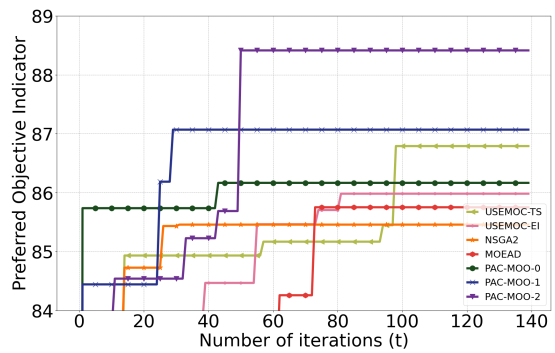


(b) Maximum discovered feasible value of F_1 - OSY

Figure 2: Hypervolume and preferred objective of the OSY benchmark with preferences defined to prioritize objective function F_1 vs. number of BO iterations.



(a) Hypervolume - SCVR



(b) Maximum discovered feasible efficiency - SCVR

Figure 3: Hypervolume and maximum discovered efficiency of a feasible design in the SCVR circuit with preferences vs. Number of circuit simulations.

constraint that renders any input outside the original input space as infeasible. Consequently, the modified OSY problem exhibits a significantly reduced rate of feasible points. This variant of the OSY problem is initialized with 12 random initial points and comprises 6 input dimensions, 2 objective functions which we are maximizing, and 7 constraints.

2. *Switched-Capacitor Voltage Regulator (SCVR) design optimization setup.* A flying-capacitor crossing technique (FCCT) is used in the multi-output SCVR to achieve dynamic capacitor optimization, as discussed in [83]. The constrained MOO problem for SCVR circuit design consists of 33 input design variables, 9 objective functions, and 14 constraints. Every baseline is initialized with 24 randomly sampled circuit configurations.

3. *High Conversion Ratio (HCR) design optimization setup.* For the HCR converter experiments, we use an inductor-first hybrid power stage, which has been previously introduced in [67]. The constrained MOO problem for HCR circuit design consists of 32 design variables, 5 objective functions, and 6 constraints. Considering that the fraction of feasible circuit configurations in the design space is extremely low (around 4%), every baseline is initialized with 32 initial feasible designs provided by a domain expert.

Table 1 consists of the details of each benchmark including the number of input dimensions, objective functions, and constraints.

In all our preference-based experiments, we assign a preference value to one high-priority objective and assign all other black-box functions (the rest of the objectives and the constraints) equal preference. The preferred objective is set to efficiency in the case of the two real-world analog circuit design problems and it is set to F_1 in the synthetic benchmark.

It is noteworthy that neither evolutionary algorithms nor the baseline BO method USEMOC are capable of handling preferences over objectives. This is an important advantage of our PAC-MOO algorithm, whose benefits we demonstrate through our experiments.

Table 1: Benchmark details

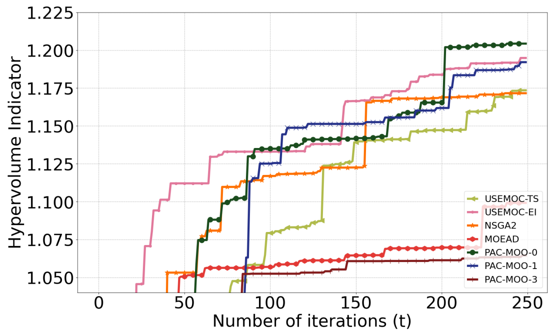
Problem Name	input dimensions	K	L
OSY	6	2	7
SCVR	33	9	14
HCR	32	5	6

5.2 Results and Discussion

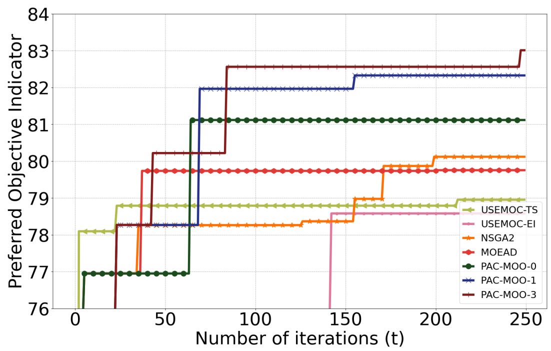
Hypervolume of Pareto set vs. the number of BO iterations. Figures 2a, 3a, and 4a show the results for pareto hypervolume of Pareto set as a function of the number of BO iterations for SCVR, HCR, and OSY problems, respectively. An algorithm is considered relatively better if it achieves higher hypervolume with a lower number of BO iterations. We make the following observations.

1) PAC-MOO with no preferences (i.e., PAC-MOO-0) outperforms all the baseline methods. This is attributed to the efficient information-theoretic acquisition function and the exploitation approach to finding feasible regions in the circuit design space.

2) At least one version of USEMOC performs better than all evolutionary baselines: USEMOC-EI for both SCVR and HCR designs, and the OSY problem. These results demonstrate that BO



(a) Hypervolume - HCR



(b) Maximum discovered feasible efficiency - HCR

Figure 4: Hypervolume and maximum discovered efficiency of a feasible design in the HCR circuit with preferences vs. Number of simulations

methods have the potential for accelerating analog circuit design optimization over evolutionary algorithms.

3) The performance of PAC-MOO with preference (i.e., PAC-MOO-1, 2, 3) may slightly decline in some cases in terms of the hypervolume. This is because the hypervolume metric evaluates the quality of the general Pareto front, while our algorithm focuses on specific regions of the Pareto front through preference specification. Although this behavior is expected, we have observed that the PHV of PAC-MOO-1 and PAC-MOO-2 remains competitive and even outperforms PAC-MOO-0, especially in the case of the OSY problem. The performance of PAC-MOO degrades only when a significantly high preference is given to one objective function and when there is a large number of objective functions available (e.g., PAC-MOO-3 in the HCR problem).

Value of the preferred objective vs. the number of BO iterations. Since efficiency is the most important objective for both SCVR and HCR circuits, we evaluate PAC-MOO by giving higher preference to efficiency over other objectives. In the case of the synthetic OSY constrained MOO problem, we assign a higher preference to the first objective function, F_1. Figures 2b, 3b, and 4b show the results for maximum discovered feasible value of the preferred objective as a function of the number of BO iterations.

1) As intended by design, PAC-MOO with preferences outperforms all baseline methods, including PAC-MOO without preferences, by finding feasible input designs with higher values of the preferred objective function,.

2) In the real-world analog circuit design problems, the improvement in maximum efficiency of uncovered circuit configurations for PAC-MOO with preferences comes at the expense of loss in hypervolume metric as shown in Figure 3a and Figure 4a.

3) In problems involving a small number of objective functions (e.g., 2 objective functions in the case of the OSY problem), the increase in the value of the preferred objective can outweigh the negative effects of emphasizing specific regions of the Pareto front through objective preferences. This effect is particularly noticeable when considering the hypervolume. As a result, PAC-MOO with preferences (PAC-MOO-2) can achieve a higher hypervolume value compared to PAC-MOO without preferences (PAC-MOO-0). This trend is evident in Figure 2a and Figure 2b.

Complexity Analysis. When comparing PAC-MOO to conventional methods, such as evolutionary baselines, it is evident that PAC-MOO introduces slightly more computational complexity. This increased complexity arises from the utilization of the sophisticated information-theoretic approach, which can result in slightly slower computational performance. Nonetheless, it's imperative to contextualize this additional computational time within practical applications such as circuit design optimization. In such practical scenarios, the additional computational overhead becomes almost negligible. For instance, in the context of the HCR problem, a single simulation takes approximately 20 minutes, whereas one iteration of PAC-MOO consumes around 30 seconds of time. This relatively minor increase in computational time becomes inconsequential, especially when considering the substantial benefits PAC-MOO offers. It consistently outperforms its counterparts, even with a slight increase in time, establishing itself as a compelling choice for real-world expensive multi-objective optimization tasks.

6 SUMMARY

Motivated by challenges in hard engineering design optimization problems (e.g., large design spaces, expensive simulations, a small fraction of configurations are feasible, and the existence of preferences over objectives), this paper proposed a principled and efficient Bayesian optimization algorithm referred to as PAC-MOO. The algorithm builds Gaussian process based surrogate models for both objective functions and constraints and employs them to intelligently select the sequence of input designs for performing experiments. The key innovations behind PAC-MOO include a scalable and efficient acquisition function based on the principle of information gain about the optimal constrained Pareto front; an effective exploitation approach to find feasible regions of the design space; and incorporating preferences over multiple objectives using a convex combination of the corresponding acquisition functions. Experimental results on one synthetic constrained multi-objective optimization problem with a small region of feasible points and two challenging real-world analog circuit design optimization problems demonstrated that PAC-MOO outperforms baseline methods in finding a Pareto set of feasible points with high hyper-volume using a small number of BO iterations. With preference specification, PAC-MOO was able to find design parameters that optimize the preferred objective functions better.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support from National Science Foundation (NSF) grants IIS-1845922, SII-2030159, and CNS-2308530.

REFERENCES

[1] Majid Abdolshah, Alistair Shilton, Santu Rana, Sunil Gupta, and Svetha Venkatesh. 2019. Multi-objective Bayesian optimisation with preferences over objectives. *Advances in neural information processing systems* 32 (2019).

[2] Alaleh Ahmadianshalchi, Syrine Belakaria, and Janardhan Rao Doppa. 2022. Preference-Aware Constrained Multi-Objective Bayesian Optimization For Analog Circuit Design. *Advances in Neural Information Processing Systems Workshop in ML for Systems* (2022).

[3] Anne Auger, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. 2009. Theory of the Hypervolume Indicator: Optimal μ-Distributions and the Choice of the Reference Point. In *Proceedings of the Tenth ACM SIGEVO Workshop on Foundations of Genetic Algorithms* (Orlando, Florida, USA) *(FOGA '09)*. Association for Computing Machinery, New York, NY, USA, 87–102. https://doi.org/10.1145/1527125.1527138

[4] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2019. Max-value Entropy Search for Multi-Objective Bayesian Optimization. In *Conference on Neural Information Processing Systems*. 7823–7833.

[5] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2020. Max-value Entropy Search for Multi-Objective Bayesian Optimization with Constraints. *arXiv preprint arXiv:2009.01721* (2020).

[6] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2020. Multi-Fidelity Multi-Objective Bayesian Optimization: An Output Space Entropy Search Approach. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 10035–10043.

[7] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2020. Uncertainty aware search framework for multi-objective Bayesian optimization with constraints. *arXiv preprint arXiv:2008.07029* (2020).

[8] Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2021. Output space entropy search framework for multi-objective Bayesian optimization. *Journal of artificial intelligence research* 72 (2021), 667–715.

[9] Syrine Belakaria, Aryan Deshwal, Nitthilan Kannappan Jayakodi, and Janardhan Rao Doppa. 2020. Uncertainty-Aware Search Framework for Multi-Objective Bayesian Optimization. In *AAAI conference on artificial intelligence*.

[10] Syrine Belakaria, Derek Jackson, Yue Cao, Janardhan Rao Doppa, and Xiaonan Lu. 2020. Machine Learning Enabled Fast Multi-Objective Optimization for Electrified Aviation Power System Design. In *ECCE*.

[11] Stephen P Boyd, Seung-Jean Kim, Dinesh D Patil, and Mark A Horowitz. 2005. Digital circuit optimization via geometric programming. *Operations research* 53, 6 (2005), 899–932.

[12] Jürgen Branke. 2008. Consideration of partial user preferences in evolutionary multiobjective optimization. *Multiobjective optimization: Interactive and evolutionary approaches* (2008), 157–178.

[13] Wonje Choi, Karthi Duraisamy, Ryan Gary Kim, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. 2018. On-Chip Communication Network for Efficient Training of Deep Convolutional Networks on Heterogeneous Manycore Systems. *IEEE Trans. Computers* 67, 5 (2018), 672–686.

[14] Tinkle Chugh. 2022. R-MBO: a multi-surrogate approach for preference incorporation in multi-objective bayesian optimisation. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 1817–1825.

[15] Thomas M Cover and Joy A Thomas. 2012. *Elements of information theory*. John Wiley and Sons.

[16] Sourav Das, Janardhan Rao Doppa, Partha Pratim Pande, and Krishnendu Chakrabarty. 2017. Design-Space Exploration and Optimization of an Energy-Efficient and Reliable 3-D Small-World Network-on-Chip. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 36, 5 (2017), 719–732.

[17] Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. 2020. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. *Advances in Neural Information Processing Systems* 33 (2020), 9851–9864.

[18] Samuel Daulton, Sait Cakmak, Maximilian Balandat, Michael A Osborne, Enlu Zhou, and Eytan Bakshy. 2022. Robust multi-objective bayesian optimization under input noise. In *International Conference on Machine Learning*. PMLR, 4831–4866.

[19] Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. 2022. Multi-objective Bayesian optimization over high-dimensional search spaces. In *Uncertainty in Artificial Intelligence*. PMLR, 507–517.

[20] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation* 6, 2 (2002), 182–197.

[21] Maria del Mar Hershenson, Sunderarajan S Mohan, Stephen P Boyd, and Thomas H Lee. 1999. Optimization of inductor circuits via geometric programming. In *Proceedings of the 36th annual ACM/IEEE Design Automation Conference*. 994–998.

[22] Aryan Deshwal, Syrine Belakaria, Ganapati Bhat, Janardhan Rao Doppa, and Partha Pratim Pande. 2021. Learning Pareto-Frontier Resource Management Policies for Heterogeneous SoCs: An Information-Theoretic Approach. In *(DAC)*.

[23] Aryan Deshwal, Nitthilan Kannappan Jayakodi, Biresh Kumar Joardar, Janardhan Rao Doppa, and Partha Pratim Pande. 2019. MOOS: A Multi-Objective Design Space Exploration and Optimization Framework for NoC enabled Manycore Systems. *ACM TECS* (2019).

[24] Aryan Deshwal, Cory Simon, and Janardhan Rao Doppa. 2021. Bayesian optimization of nanoporous materials. *Molecular Systems Design and Engineering, Royal Society of Chemistry* (2021).

[25] Liang Du, Ruobin Gao, Ponnuthurai Nagaratnam Suganthan, and David Z.W. Wang. 2022. Bayesian optimization based dynamic ensemble for time series forecasting. *Information Sciences* 591 (2022), 155–175. https://doi.org/10.1016/j.ins.2022.01.010

[26] R. J. Duffin and E. L. Peterson. 1973. Geometric programming with signomials. *Journal of Optimization Theory and Applications* 11, 1 (Jan. 1973), 3–35. https://doi.org/10.1007/bf00934288

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

[28] David Eriksson, Michael Pearce, Jacob Gardner, Ryan D Turner, and Matthias Poloczek. 2019. Scalable Global Optimization via Local Bayesian Optimization. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/6c990b7aca7bc7058f5e98ea909e924b-Paper.pdf

[29] Mourad Fakhfakh, Yann Cooren, Amin Sallem, Mourad Loulou, and Patrick Siarry. 2010. Analog circuit design optimization through the particle swarm optimization technique. *Analog integrated circuits and signal processing* 63, 1 (2010), 71–82.

[30] Paul Feliot, Julien Bect, and Emmanuel Vazquez. 2017. A Bayesian approach to constrained single-and multi-objective optimization. *Journal of Global Optimization* 67, 1-2 (2017), 97–133.

[31] Jacob R Gardner, Matt J Kusner, Zhixiang Eddie Xu, Kilian Q Weinberger, and John P Cunningham. 2014. Bayesian Optimization with Inequality Constraints.. In *ICML*, Vol. 2014. 937–945.

[32] Eduardo C Garrido-Merchán, Daniel Fernández-Sánchez, and Daniel Hernández-Lobato. 2023. Parallel predictive entropy search for multi-objective Bayesian optimization with constraints applied to the tuning of machine learning algorithms. *Expert Systems with Applications* 215 (2023), 119328.

[33] Eduardo C Garrido-Merchán and Daniel Hernández-Lobato. 2019. Predictive entropy search for multi-objective Bayesian optimization with constraints. *Neurocomputing* 361 (2019), 50–68.

[34] Finley J Gibson, Richard M Everson, and Jonathan E Fieldsend. 2022. Guiding surrogate-assisted multi-objective optimisation with decision maker preferences. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 786–795.

[35] Georges GE Gielen, Herman CC Walscharts, and Willy MC Sansen. 1990. Analog circuit design optimization based on symbolic simulation and simulated annealing. *IEEE Journal of Solid-State Circuits* 25, 3 (1990), 707–713.

[36] David E. Goldberg. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning* (1st ed.). Addison-Wesley Longman Publishing Co., Inc., USA.

[37] Tomasz Golonek and Jerzy Rutkowski. 2007. Genetic-algorithm-based method for optimal analog test points selection. *IEEE Transactions on Circuits and Systems II: Express Briefs* 54, 2 (2007), 117–121.

[38] Kyle J Goodrick, Emiliano Dall'Anese, and Dragan Maksimović. 2021. Systematic Design and Optimization of Large DC Distribution Architectures Using Simulated Annealing. In *2021 IEEE Applied Power Electronics Conference and Exposition (APEC)*. IEEE, 134–141.

[39] Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah, and Ryan Adams. 2016. Predictive entropy search for multi-objective Bayesian optimization. In *Proceedings of International Conference on Machine Learning (ICML)*. 1492–1501.

[40] José Miguel Hernández-Lobato, Matthew W Hoffman, and Zoubin Ghahramani. 2014. Predictive entropy search for efficient global optimization of black-box functions. In *Advances in Neural Information Processing Systems*. 918–926.

[41] Jiangli Huang, Fan Yang, Changhao Yan, Dian Zhou, and Xuan Zeng. 2021. A Robust Batch Bayesian Optimization for Analog Circuit Synthesis via Local Penalization. In *Proceedings of the 26th Asia and South Pacific Design Automation Conference* (Tokyo, Japan) *(ASPDAC '21)*. Association for Computing Machinery, New York, NY, USA, 146–151. https://doi.org/10.1145/3394885.3431543

[42] Jiangli Huang, Shuhan Zhang, Cong Tao, Fan Yang, Changhao Yan, Dian Zhou, and Xuan Zeng. 2021. Bayesian Optimization Approach for Analog Circuit Design Using Multi-Task Gaussian Process. In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. 1–5. https://doi.org/10.1109/ISCAS51556.2021.9401205

[43] Nitthilan Kanappan Jayakodi, Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. 2020. Design and Optimization of Energy-Accuracy Tradeoff Networks for Mobile Platforms via Pretrained Deep Models. *ACM Transactions on Embedded Computing Systems (TECS)* 19, 1 (2020), 4:1–4:24.

[44] Biresh Kumar Joardar, Ryan Gary Kim, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. 2018. Learning-based Application-Agnostic 3D NoC Design for Heterogeneous Manycore Systems.

IEEE Trans. Comput. 68, 6 (2018), 852–866.

[45] Donald R. Jones, Matthias Schonlau, and William J. Welch. 1998. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* 13, 4 (1998), 455–492. https://doi.org/10.1023/a:1008306431147

[46] J. Kennedy and R. Eberhart. [n. d.]. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*. IEEE. https://doi.org/10.1109/icnn.1995.488968

[47] Jintae Kim, Jaeseo Lee, Lieven Vandenberghe, and Chih-Kong Ken Yang. 2004. Techniques for improving the accuracy of geometric-programming based analog circuit design optimization. In *IEEE/ACM International Conference on Computer Aided Design, 2004. ICCAD-2004*. IEEE, 863–870.

[48] Jong-Hwan Kim, Ji-Hyeong Han, Ye-Hoon Kim, Seung-Hwan Choi, and Eun-Soo Kim. 2012. Preference-Based Solution Selection Algorithm for Evolutionary Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation* 16, 1 (2012), 20–34. https://doi.org/10.1109/TEVC.2010.2098412

[49] Ryan Gary Kim, Wonje Choi, Zhuo Chen, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. 2017. Imitation Learning for Dynamic VFI Control in Large-Scale Manycore Systems. *IEEE Trans. Very Large Scale Integr. Syst.* 25, 9 (2017), 2458–2471.

[50] Ryan Gary Kim, Janardhan Rao Doppa, Partha Pratim Pande, Diana Marculescu, and Radu Marculescu. 2018. Machine Learning and Manycore Systems Design: A Serendipitous Symbiosis. *IEEE Computer* 51, 7 (2018), 66–77.

[51] Zhiyuan Jerry Lin, Raul Astudillo, Peter Frazier, and Eytan Bakshy. 2022. Preference exploration for efficient bayesian optimization with multiple outcomes. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 4235–4258.

[52] Jialin Lu, Shuhan Zhang, Fan Yang, Dian Zhou, and Xuan Zeng. 2020. A Mixed-Variable Bayesian Optimization Approach for Analog Circuit Synthesis. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. 1–4. https://doi.org/10.1109/ISCAS45731.2020.9181162

[53] Wenlong Lyu, Pan Xue, Fan Yang, Changhao Yan, Zhiliang Hong, Xuan Zeng, and Dian Zhou. 2018. An Efficient Bayesian Optimization Approach for Automated Optimization of Analog Circuits. *IEEE Transactions on Circuits and Systems I: Regular Papers* 65, 6 (2018), 1954–1967. https://doi.org/10.1109/TCSI.2017.2768826

[54] Wenlong Lyu, Fan Yang, Changhao Yan, Dian Zhou, and Xuan Zeng. 2018. Batch Bayesian Optimization via Multi-objective Acquisition Ensemble for Automated Analog Circuit Design. In *ICML*.

[55] Wenlong Lyu, Fan Yang, Changhao Yan, Dian Zhou, and Xuan Zeng. 2018. Multi-objective Bayesian Optimization for Analog/RF Circuit Synthesis. In *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*. 1–6. https://doi.org/10.1109/DAC.2018.8465872

[56] Sumit K. Mandal, Ganapati Bhat, Janardhan Rao Doppa, Partha Pratim Pande, and Ümit Y. Ogras. 2020. An Energy-aware Online Learning Framework for Resource Management in Heterogeneous Platforms. *ACM Trans. Design Autom. Electr. Syst.* 25, 3 (2020), 28:1–28:26.

[57] Sumit K. Mandal, Ganapati Bhat, Chetan Arvind Patil, Janardhan Rao Doppa, Partha Pratim Pande, and Ümit Y. Ogras. 2019. Dynamic Resource Management of Heterogeneous Mobile Platforms via Imitation Learning. *IEEE Trans. Very Large Scale Integr. Syst.* 27, 12 (2019), 2842–2854.

[58] Joseph Victor Michalowicz, Jonathan M Nichols, and Frank Bucholtz. 2013. *Handbook of differential entropy*. Chapman and Hall/CRC.

[59] Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. 1978. The Application of Bayesian Methods for Seeking the Extremum. *Towards Global Optimization* 2, 117-129 (1978), 2.

[60] Gaurav Narang, Aryan Deshwal, Janardhan Rao Doppa, Partha Pratim Pande, Raid Ayoub, and Mike Kishinevsky. 2023. Dynamic Power Management in Large Manycore Systems: A Learning-to-Search Framework. *ACM Transactions on Design Automation of Electronic Systems (TODAES)* (2023).

[61] Andrzej Osyczka and Sourav Kundu. 1995. A new method to solve generalized multicriteria optimization problems using the simple genetic algorithm. *Structural optimization* 10 (1995), 94–99.

[62] Yu Pang, Kartazyna Radecka, and Zeljko Zilic. 2010. Optimization of imprecise circuits represented by Taylor series and real-valued polynomials. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 29, 8 (2010), 1177–1190.

[63] Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. 2020. A flexible framework for multi-objective bayesian optimization using random scalarizations. In *Uncertainty in Artificial Intelligence*. PMLR, 766–776.

[64] Ali Rahimi and Benjamin Recht. 2008. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*. 1177–1184.

[65] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* 104, 1 (2016), 148–175. https://doi.org/10.1109/JPROC.2015.2494218

[66] Shinya Suzuki, Shion Takeno, Tomoyuki Tamura, Kazuki Shitara, and Masayuki Karasuyama. 2020. Multi-objective Bayesian optimization using Pareto-frontier entropy. In *International Conference on Machine Learning*. PMLR, 9279–9288.

[67] Nghia Tang, Wookpyo Hong, Bai Nguyen, Zhiyuan Zhou, Jong-Hoon Kim, and Deukhyoun Heo. 2020. Fully integrated switched-inductor-capacitor voltage regulator with 0.82-A/mm 2 peak current density and 78% peak power efficiency. *IEEE Journal of Solid-State Circuits* 56, 6 (2020), 1805–1815.

[68] Rajesh Amratlal Thakker, M Shojaei Baghini, and Mahesh B Patil. 2009. Low-power low-voltage analog circuit design using hierarchical particle swarm optimization. In *2009 22nd international conference on VLSI design*. IEEE, 427–432.

[69] Lothar Thiele, Kaisa Miettinen, Pekka J Korhonen, and Julian Molina. 2009. A preference-based evolutionary algorithm for multi-objective optimization. *Evolutionary computation* 17, 3 (2009), 411–436.

[70] Hakki Mert Torun, Madhavan Swaminathan, Anto Kavungal Davis, and Mohamed Lamine Faycal Bellaredj. 2018. A global Bayesian optimization algorithm and its application to integrated system design. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 26, 4 (2018), 792–802.

[71] Konstantinos Touloupas, Nikos Chouridis, and Paul P. Sotiriadis. 2021. Local Bayesian Optimization For Analog Circuit Sizing. In *2021 58th ACM/IEEE Design Automation Conference (DAC)*. 1237–1242. https://doi.org/10.1109/DAC18074.2021.9586172

[72] Peter J. M. van Laarhoven and Emile H. L. Aarts. 1987. *Simulated Annealing: Theory and Applications*. Springer Netherlands. https://doi.org/10.1007/978-94-015-7744-1

[73] Revna Acar Vural and Tulay Yildirim. 2010. Component value selection for analog active filter using particle swarm optimization. In *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, Vol. 1. IEEE, 25–28.

[74] Jianzhou Wang, Wendong Yang, Pei Du, and Yifan Li. 2018. Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. *Energy* 148 (2018), 59–78.

[75] Ye Wang, Michael Orshansky, and Constantine Caramanis. 2014. Enabling efficient analog synthesis by coupling sparse regression and polynomial optimization. In *Proceedings of the 51st Annual Design Automation Conference*. 1–6.

[76] Yu Wang, Hailian Yin, Shuai Zhang, and Xiongqing Yu. 2014. Multi-objective optimization of aircraft design for emission and cost reductions. *Chinese Journal of Aeronautics* 27, 1 (2014), 52–58.

[77] Christopher KI Williams and Carl Edward Rasmussen. 2006. *Gaussian processes for machine learning*. Vol. 2. MIT Press.

[78] Xiaoxuan Yang, Syrine Belakaria, Biresh Kumar Joardar, Huanrui Yang, Janardhan Rao Doppa, Partha Pratim Pande, Krishnendu Chakrabarty, and Hai Helen Li. 2021. Multi-Objective Optimization of ReRAM Crossbars for Robust DNN Inferencing under Stochastic Noise. In *IEEE/ACM International Conference On Computer Aided Design, ICCAD 2021, Munich, Germany, November 1-4, 2021*. IEEE, 1–9.

[79] Xiaoxuan Yang, Syrine Belakaria, Biresh Kumar Joardar, Huanrui Yang, Janardhan Rao Doppa, Partha Pratim Pande, Krishnendu Chakrabarty, and Hai Helen Li. 2021. Multi-Objective Optimization of ReRAM Crossbars for Robust DNN Inferencing under Stochastic Noise. In *IEEE/ACM International Conference On Computer Aided Design, ICCAD 2021, Munich, Germany, November 1-4, 2021*. IEEE, 1–9.

[80] Qingfu Zhang and Hui Li. 2007. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation* 11, 6 (2007), 712–731. https://doi.org/10.1109/TEVC.2007.892759

[81] Shuhan Zhang, Wenlong Lyu, Fan Yang, Changhao Yan, Dian Zhou, and Xuan Zeng. 2019. Bayesian Optimization Approach for Analog Circuit Synthesis Using Neural Network. In *2019 Design, Automation Test in Europe Conference Exhibition (DATE)*. IEEE. https://doi.org/10.23919/date.2019.8714788

[82] Zhiyuan Zhou, Syrine Belakaria, Aryan Deshwal, Wookpyo Hong, Janardhan Rao Doppa, Partha Pratim Pande, and Deukhyoun Heo. 2020. Design of Multi-Output Switched-Capacitor Voltage Regulator via Machine Learning. In *DATE*.

[83] Zhiyuan Zhou, Syrine Belakaria, Aryan Deshwal, Wookpyo Hong, Janardhan Rao Doppa, Partha Pratim Pande, and Deukhyoun Heo. 2020. Design of multi-output switched-capacitor voltage regulator via machine learning. In *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 502–507.

[84] Eckart Zitzler. 1999. *Evolutionary algorithms for multiobjective optimization: methods and applications*. Ph. D. Dissertation. University of Zurich, Zürich, Switzerland.