Work-in-Progress: On Leveraging Approximations for Exact System-level Design Space Exploration

Kai Neubauer, Christian Haubelt
University of Rostock, Germany
{kai.neubauer,christian.haubelt}@uni-rostock.de

Philipp Wanko, Torsten Schaub
University of Potsdam, Germany
{torsten,wanko}@cs.uni-potsdam.de

***Abstract*—In order to find good design points for embedded systems, an efficient exploration of the design space is imperative. The ever-increasing complexity of embedded systems, however, results in a deterioration of the overall exploration performance. The DSE essentially consists of two parts: (1) the search for feasible solutions and (2) the evaluation of found feasible solutions. While the search has been massively improved by ASPmT-based strategies, the evaluation emerges as the main bottleneck. Tragically, evaluating bad solutions takes as much time as evaluating good ones. Hence, in this paper we study the utilization of approximations in the evaluation process integrated in an ASPmT-based DSE to identify bad solutions more quickly while still retaining the exact Pareto-front.**

***Index Terms*—Design Space Exploration, System-level Design, Approximation**

I. Introduction and Related Work

Exploring the design space of embedded systems involves a number of steps that influence the overall performance and can be conceptually seen as a filtering process (see [1]) as depicted in Fig. 1. From the set of all possible solutions X, a feasibility filter returns design points that comply with given feasibility constraints, i.e., feasible binding, routing, etc., forming the feasible set X_F. Afterwards, the validity filter yields valid design points X_V based on quality constraints, e.g., latency and power constraints. Finally, all dominated design points are sorted out by the Pareto filter, resulting in X_P. A solution x is dominated by solution y if all objective functions f_n evaluate y to be at least equivalent good as x and at least one objective evaluates y better than x. While the feasibility filter is usually based on structural rules that can be efficiently solved by symbolic techniques such as SAT and Answer Set Programming (ASP), the validity and Pareto filters depend on the evaluation of the design points w.r.t. constraint and objective functions like latency, power, and area requirements. Especially, non-linear objectives (e.g., latency) are hard to evaluate as often costly simulations have to be executed. Moving these filters into a background theory, resulting in an ASP modulo Theory (ASPmT)-based design space exploration (DSE) [2], has been shown to speed up the search. However, the evaluation of feasible implementations remains a bottleneck.

In order to speed up this process, Piscitelli et al. [3] propose to use a hybrid approach combining fast analytical estimations with costly but accurate simulations. Here, the application mapping is first encoded into a Kahn Process Network (KPN) and analyzed analytically. Whenever the resulting KPN contains cycles, the estimation becomes inaccurate and the costly simulation is executed to obtain the real objective value. The authors, however, propose to relax this requirement by only evaluating a specific percentage of found solutions using simulations while using the estimation for the rest. Hence, some non-dominated designs might be excluded from the final solution set. More recently, Zhang et al. [4] propose to approximate objective and constraint functions using stochastic simulations. After an initial set of design points, they determine the probability of feasibility as well as the expected hypervolume improvement to select the next regions to explore. Again, this approach only creates an approximate Pareto front of the problem, i.e., some Pareto-optimal design points might not be found.

This work was funded by the German Science Foundation (DFG) under grants HA 4463/4-1 and SCHA 550/11-1.

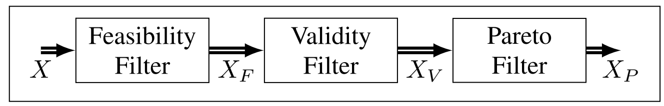

Fig. 1. Design Space Exploration as filtering process obtaining the Pareto front X_P from the design space X.

In comparison, we present an approximation-based methodology that is able to obtain the true Pareto front. For this purpose, we to use approximations during objective calculation in the background theory of the ASPmT-based DSE with the goal to prune regions from the search space that do not contain non-dominated solutions. Only in cases where approximation does not permit an early decision, the more time-consuming accurate objective calculation is invoked.

II. Approximations for DSE

Typically, during the exploration, one or more solutions from the design space X are taken and evaluated for feasibility, validity, and non-dominance regarding already found solutions. In the beginning of the DSE, many of the newly found (valid) design points are considered to be non-dominated and the overall quality improves steadily. With each improvement, however, the number of better solutions is reduced resulting in stagnation after these initial improvements. Still, in this phase, each novel solution from the design space is evaluated with the same costly evaluation functions.

Instead, we propose to utilize fast estimation functions that accelerate deciding whether a design point is dominated by already found solution or not. The idea is that the estimated quality vectors are compared against the already found

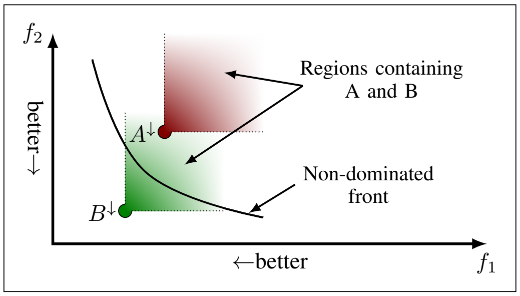

Fig. 2. Under-approximations in minimization problems. The shaded red and green regions are possible locations of the exact values A and B, respectively.

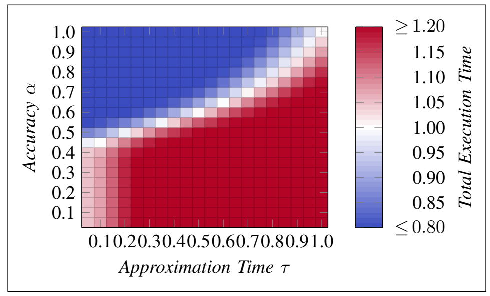

Fig. 3. Dependency of the accuracy and calculation time of the estimation function on the total execution time.

solutions and only if the results indicate a promising quality, the costly exact evaluation is performed. Nevertheless, this technique is only possible if the accuracy of the estimation is bounded. A precise restriction of the estimation to specific error bounds, as required in the work of Abraham et al. [1], is typically not viable for real-world objectives like latency. Thus, we relax the restriction of specific error bounds towards estimation consistency, called over- and under-approximations in the following. Given an exact objective function $f(X)$, an estimation is called over-approximation $f^{\uparrow}(X)$ if the estimated value is always larger than (or equal to) the exact value. Analogously, an estimation is called under-approximation $f^{\downarrow}(X)$ if the estimated value is always smaller than (or equal to) the exact value. Formally: $f^{\downarrow}(X) \leq f(X) \leq f^{\uparrow}(X)$. For example, in the area of embedded system-level design, a safe under-approximation for the latency evaluation is finding a schedule without considering communication delays and link congestion. This way, the number of variables and, thus, the evaluation time can be reduced.

The reasonable utilization of under- and over-approximation is mutually exclusive. While minimization problems mainly profit from under-approximation, over-approximations can be leveraged by maximization problems. Without loss of generality, the technique is depicted in Fig. 2 for the objective functions f_1 and f_2 that are to be minimized. In the first step, both design points **A** and **B** are estimated through $f_1^{\downarrow}$ and $f_2^{\downarrow}$. As can be seen, the under-approximated quality vector of **A** is already dominated by the previously found solutions in the non-dominated front. In combination with the fact that the exact value is always larger than the approximation, **A** can be discarded directly. On the other hand, the estimation of **B** dominates the current front and might be an Pareto-optimal design point. Hence, the value of **B** has to be determined through the exact objective functions. Afterwards, the dominance check can be performed for the exact value of **B**.

III. Accuracy vs. Approximation Time

As can be seen in the example above, the advantage of our approach helps to diminish the evaluation time whenever the estimated quality vector is already dominated by previously found deign points. On the other hand, two calculations are necessary (i.e., the estimated and exact) if the new design point is still non-dominated after approximation. One of the most important factors influencing the number of double calculation is the accuracy of the estimation itself. That is, if the estimation is very inaccurate w.r.t. the exact value, most of the evaluations have to be executed twice introducing an additional solving overhead compared to the original methodology without using function approximations.

In order to visualize this problem, we conducted a set of Pareto-filtering simulations with convex and concave Pareto sets. We generated a set X_V of 10,000 valid design points and filtered out the non-dominated set X_P. For this illustration, the accuracy[1] $\alpha \in \{0.05, \dots, 1.0\}$ and approximation time $\tau \in \{0.05,...,1.0\}$ were given as fractions of the corresponding exact calculations. For example, if the exact evaluation takes T time units and produces the quality vector (v_1, v_2), the estimation takes $\tau \cdot T$ time units and yields $(\alpha \cdot v_1, \alpha \cdot v_2)$. The characteristics of the results are similar for each investigated Pareto set and only differ in specific values. As depicted in Fig. 3, both the accuracy and the required approximation time are important for an overall improvement of the total execution time (i.e., the complete Pareto-filtering process). While the blue regions indicate an improvement, red regions indicate a deterioration compared to using exact evaluations only.

IV. Discussion

We presented a methodology on how to utilize constraint and objective function approximation in order to fasten exact design space exploration. The visualization presented in the previous section shows the most important requirements of the approach: in order to improve the overall filtering performance, the estimation functions have to be selected carefully to assure a certain accuracy. However, real DSE may perform differently as the parameters α and τ might not be statically determinable. In principle, the methodology presented here is independent of the utilized DSE technique. Yet, detailed implications for specific search strategies (e.g., population-based strategies) are not negligible and have to be studied further.

References

[1] S. G. Abraham *et al.*, "Fast design space exploration through validity and quality filtering of subsystem designs," *HP Laboratories Technical Report*, vol. 98, 2000.

[2] K. Neubauer *et al.*, "Exact multi-objective design space exploration using aspmt," in *Proceedings of DATE*, March 2018, pp. 257–260.

[3] R. Piscitelli and A. D. Pimentel, "Design space pruning through hybrid analysis in system-level design space exploration," in *Proc. of DATE*, 2012, pp. 781–786.

[4] J. Zhang *et al.*, "Estimation of the pareto front in stochastic simulation through stochastic kriging," *Simulation Modelling Practice and Theory*, vol. 79, pp. 69 – 86, 2017.

[1] For sake of brevity, we use an accuracy model that scales all objectives simultaneously.