

# Generation of Yield-Aware Pareto Surfaces for Hierarchical Circuit Design Space Exploration

Saurabh K Tiwary  
Carnegie Mellon University  
Pittsburgh, PA USA  
stiwary@ece.cmu.edu

Pragati K Tiwary  
BIT Mesra  
Ranchi, Jharkhand India  
pragati398@bitmesra.ac.in

Rob A Rutenbar  
Carnegie Mellon University  
Pittsburgh, PA USA  
rutenbar@ece.cmu.edu

## ABSTRACT

Pareto surfaces in the performance space determine the range of feasible performance values for a circuit topology in a given technology. We present a non-dominated sorting based global optimization algorithm to generate the nominal Pareto front efficiently using a simulator-in-a-loop approach. The solutions on this Pareto front combined with efficient Monte Carlo approximation ideas are then used to compute the *yield-aware* Pareto fronts. We show experimental results for both the nominal and yield-aware Pareto fronts for power and phase noise for a voltage controlled oscillator (VCO) circuit. The presented methodology computes yield-aware Pareto fronts in approximately 5-6 times the time required for a single circuit synthesis run and is thus practically efficient. We also show applications of yield-aware Pareto fronts to find the optimal VCO circuit to meet the system level specifications of a phase locked loop.

## Categories and Subject Descriptors

B.7.2 [Integrated Circuits]: Design Aids

## General Terms

Algorithms Design

## Keywords

Yield, performance space, Pareto surfaces, optimization

## 1. INTRODUCTION

Recent advances in design automation and increased computational power has led to a gradual transition from “hand-calculation” based analog circuit design to a simulation-based sizing methodology [3]. Simulation based synthesis uses efficient global optimization techniques to visit many circuit candidates, and fully evaluates each candidate via detailed simulation. This methodology works very well for circuits in the range of a few hundred devices. However, for larger circuits, the simulation time required for a single simulation is too large to do a practical simulator-in-the-loop circuit sizing. Also, with large number of design variables, the design space in which to search for optimal design points becomes too large for these circuits to be handled by the tools available today. To handle this problem, methods for modeling the performance space of circuits have been presented [10] [14] [5]. Using

these performance models, efficient system level design space explorations can be done for topology selection [9] as well as circuit synthesis [13]. Of late, much work has been done in the field of analog circuit macromodeling which aims at simulating these circuits faster [16] [8] [12]. These macromodels capture the important functional characteristics of the circuit while discarding the redundant information. They are faster to simulate than the original transistor level circuit and hence could be used to replace the circuits that they model in a system level simulation context. A class of models called *Pareto surfaces* generated in the performance space of the circuits has often been used with these macromodeling tools for hierarchical synthesis of larger systems [13]. These Pareto surfaces can be generated by both stochastic [19] [17] and deterministic algorithms [6] [15].

Pareto surfaces represent the best performance that can be obtained from a given circuit topology across its complete design space. These surfaces are generated for the nominal values of the process parameters and are often used for optimizing the performance parameters of the circuit in order to meet the system level specifications [13]. However, with reduction in feature sizes, the effect of process variations on circuit performance is becoming more pronounced. Such a system level design methodology gives us no idea of how the circuit would perform once fabricated with the optimized performance parameters in the presence of process variations. In the worst case, it is likely that the nominal design point lies in the tail of the performance distribution when Monte Carlo analysis is performed for the optimized circuit candidate across process variations. This, in turn, would result in poor *yield* (fraction of circuits that meet the system level specifications across all the circuits that are fabricated). Therefore, a *yield-aware* Pareto surface (points on the Pareto surface guaranteeing a fixed yield number) would be more useful for hierarchical synthesis applications.

In this paper, we present an efficient, novel algorithm for generating a yield-aware Pareto surface for an arbitrary custom circuit. To the best of our knowledge, this is the first realistic approach successfully implemented and validated for this problem. Sec.2 gives some basic background on Pareto methods and the functionality missing in prior efforts. Sec.3 describes the elements of our yield-aware Pareto generation methodology: nominal “soft” Pareto generation; yield-aware Pareto for a particular design point; and optimization of yield-aware Pareto curve across the design space. Sec.4 shows experimental results for a voltage controlled oscillator (VCO) circuit. Nominal and yield-aware Pareto fronts for different yield targets of 20%, 50% and 80% are generated for the circuit efficiently. The total time taken for the front generation is only 5-6X the CPU time required for circuit synthesis runs. These Pareto fronts are then used to hierarchically synthesize a PLL. Sec.5 offers concluding remarks followed by acknowledgements in Sec.6.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2006, July 24–28, 2006, San Francisco, California, USA.  
Copyright 2006 ACM 1-59593-381-6/06/0007 ...\$5.00.

## 2. BACKGROUND

### 2.1 Circuit Sizing

A circuit sizing (synthesis) problem is to find the sizes/values of elements in a circuit block *e.g.*,  $W$  and  $L$  of transistors, *etc.*, such that the block meets a set of performance specifications. Formally, if  $\mathbf{p}$  is a vector of parameters values on which the performance function  $\mathbf{f}(\mathbf{p})$  depends, then the sizing problem is to find a set of values for  $\mathbf{p}$  represented by  $\mathbf{p}_0$  such that

$$\mathbf{f}(\mathbf{p}_0) < \mathbf{f}_g \quad (1)$$

Here,  $\mathbf{f}_g$  is the set of goal specifications for the circuit. Lower bounds on the performance parameters can be specified using this formulation by multiplying the goal value by -1.

If the statistical variation of the parameters is to be considered, *e.g.*, process variations, then the circuit sizing problem is rephrased as a yield improvement problem. The goal in such cases is to increase the *percentage* of circuit meeting the goal specifications ( $\mathbf{f}_g$ ) for a particular set of input parameters ( $\mathbf{p}_0$ ). A common method used to find the performance values for a particular set of input parameters is circuit simulation.

$$\mathbf{p} \dashv \text{circuit simulation} \mapsto \mathbf{f}(\mathbf{p}) \quad (2)$$

SPICE like circuit simulators are the most common tools used for such mapping. Being the most accurate tool for finding such a mapping, the process of circuit simulation is also the most expensive. Automatic circuit sizing tools are often designed in such a way so as to do minimal number of SPICE simulation. For a constrained range of input parameters, regression based approaches for modeling the performance function is often employed for fast evaluations. However, the performance function is very non-linear for large variations in the parameter space. A lot of effort has been devoted to better model the performance as a function of design parameters [5] [10]. Such efforts often require a lot of SPICE level circuit simulation upfront for building the models.

### 2.2 Pareto Optimal Surfaces

Often, a number of the elements in the objective function  $\mathbf{f}(\mathbf{p})$  compete against each other. Thus, the circuit optimization problem turns out to be a multi-objective optimization problem over the range of parameters ( $\mathbf{p}$ ) such that some sizing constraints ( $\mathbf{c}(\mathbf{p}) > \mathbf{0}$ ) are satisfied. For competing objective functions, it is not feasible to maximize the performance of all of them. In such case, we often have to consider *trade-offs* between the different performance parameters. Improving one might result in decreasing another competing performance function. In such cases, we strive for pareto optimality which ensures the best overall performance one could extract in the presence of competing performance functions.

Defining it formally, a set of performances  $\mathbf{f}_a$  is considered more optimal than another set  $\mathbf{f}_b$  if  $\mathbf{f}_a$  *dominates*  $\mathbf{f}_b$ .

$$\mathbf{f}_a \succ \mathbf{f}_b \Leftrightarrow \forall (f_{a_i} \geq f_{b_i}) \wedge \exists (f_{a_i} > f_{b_i}) \quad \text{for } i = \{1, \dots, n\} \quad (3)$$

A set of performances  $\mathbf{f}^*$  is considered pareto optimal if it is not dominated by any other set of performance parameters  $\mathbf{f}$ . The surface generated in the performance space by the complete set of pareto optimal points make up the *pareto optimal surface*.

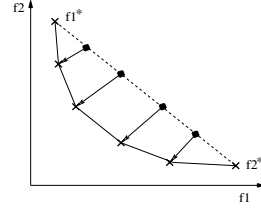
### 2.3 Finding the Pareto Optimal Surface

Many tools for circuit sizing [3] have focused on finding a single point on the pareto surface. This problem can be transformed into a scalar optimization problem by weighting the different performance parameters.

$$\max \mathbf{o}(\mathbf{p}) = \sum_{i=1}^n w_i f_i(\mathbf{p}) \quad \text{s.t. } \mathbf{c}(\mathbf{p}) > \mathbf{0} \quad (4)$$

A set of weights  $\mathbf{w} = \{w_1, \dots, w_n\}$  leads to a particular solution on the pareto front. The methods for obtaining the solution to this problem are fairly mature and various commercial softwares are also available for the purpose [2].

Considerable research has focussed on the automated generation of the pareto surfaces using deterministic and stochastic optimization techniques. One of the common methods for generating the points is by choosing different values for  $w_i$  in Eqn. 4. In such a case, the challenge is to find sets of weight values such that pareto front could be captured with least number of point evaluations of the surface. A method to automatically generate suitably spaced points on the pareto is the normal boundary intersection (NBI) method [6]. The basic idea of the algorithm is shown in Fig. 1. First, the global optimal points on the pareto are obtained for each



**Figure 1: Normal boundary intersection (NBI) method.**

of the performance dimensions ( $f1^*$  and  $f2^*$ ). The convex hull between the individual optima is described by

$$\mathbf{F} \cdot \mathbf{w} = w_1 f1^* + w_2 f2^* \quad \text{for } w_1, w_2 \geq 0; w_1 + w_2 = 1 \quad (5)$$

The convex hull is intersected at equidistant points as shown in Fig. 1. At each of these points the best performance set for the circuit is obtained in the direction of the normal to the convex hull. These points, obtained using standard circuit optimization techniques, are used to approximate the pareto surface.

Another method often employed for generating these pareto fronts is the non-dominated sorting genetic algorithm (NSGA) [7]. The basic idea is to start with a population of samples in the parameter space (*individuals*) are called *chromosomes*. During each generation, the individuals are sorted based on a *fitness* function (dominating points (Eqn. 3) are *fitter* for pareto surface generation). The second generation of population is generated by *crossover* and *mutation* of the chromosomes of the fitter individuals [7].

Simulated annealing [11] is another method quite extensively used for global optimization. We start at a randomly chosen point (*seed*) in the parameter space and then move to another point in the space with certain probabilities. The probability to move to the new point in the parameter space is higher if the performance values improve at this new point. Even in cases where the performance value degrades, the move is sometimes accepted (with low probability). This step ensures hill climbing properties of the algorithm.

### 2.4 Challenges in Generating Pareto Surfaces

The primary challenge that one faces while generating the pareto front is the curse of dimensionality. The number of points required to model a surface increases exponentially with the number of dimensions. The normal boundary intersection method, for example, requires proportional number of circuit optimization runs as the number of points required on the surface. Another problem is that the additional constraint of searching for optimal points in the direction perpendicular to the convex hull (Fig. 1) makes the optimization problem extremely difficult if we are using a stochastic optimizer to find the optimal point on the line. Also, for each point on the pareto front, one needs to run a complete sequence of the global optimization routine. This makes the method intractable for problems with high dimensionality.

Stochastic algorithms that dynamically handle the pareto front like non-dominated sorting genetic algorithm (NSGA-II) [7] are very efficient in generating the pareto fronts. The problem with using such an algorithm with applications in modeling the pareto surface of a given circuit topology is that, the resulting pareto front has no notion of *yield* for the circuit. The fronts thus generated only consider the *nominal* design parameter values for the circuit. Once the circuit is fabricated for a particular set of design parameter values corresponding to a point on the pareto surface, it is probable that a high percentage of the fabricated circuit’s performance is inferior to that point on the pareto due to variations in process parameters. This probability is very high for “aggressively” generated pareto fronts. Often the performance function is highly non-linear with respect to the process parameters (e.g. *tox*, *vth* etc.) of the circuit. Thus, if we neglect the process variation information, the optimizer may choose designs whose nominal performance parameters lie on the pareto front but, this nominal design actually lies in the tail of the performance distribution across process variations.

The right – and so far, missing – solution is a yield-aware pareto surface. A trivial and brute-force approach would simply compute the yield for all circuit candidates through Monte Carlo sampling and thus, optimize the yield-aware pareto fronts. However, this method would be computationally very expensive if not prohibitive. The rest of the paper develops a new methodology that seeks to address this problem and generate yield-aware pareto surfaces efficiently.

### 3. YIELD-AWARE PARETO GENERATION

We have developed a *yield-aware* pareto front generation algorithm. The key steps are shown in Algorithm 1. The basic idea is to generate “soft” pareto surfaces for nominal design (without taking process variations into account). By “soft” we mean that we do not aggressively search for the actual pareto front using the stochastic algorithms described in Section 3.1. We stop the search after moderate convergence of the algorithm. Using points on this “soft” pareto, a fast *local* yield estimator algorithm is used to search for locally optimum pareto points. These points are then combined together to form the yield-aware pareto fronts. Before we discuss the details of our algorithm, let us revisit the nominal pareto generation problem.

---

#### Algorithm 1 Pseudo-code for yield-aware pareto front generation

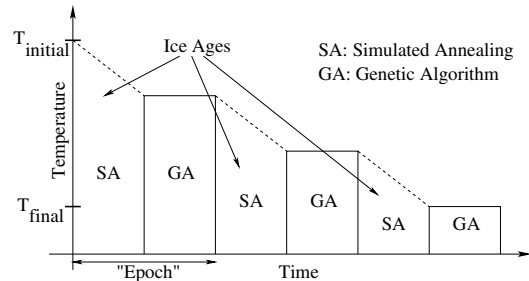
---

- 1: *Given*: A circuit topology, ranges for design parameters and distributions of process parameters
  - 2: Generate “soft” nominal pareto using NSGA with “Ice Ages”
  - 3: Choose equi-distant pts  $\{x_1, \dots, x_N\}$  on pareto front of order  $k$
  - 4: **for all**  $x_i$  as anchor points **do**
  - 5:   Generate *local* yield model
  - 6:   Locally optimize the yield front around anchor points
  - 7: **end for**
  - 8: Merge all the local yield fronts to generate a yield-aware pareto front for the given circuit topology
- 

#### 3.1 NSGA with “Ice Ages”

We use a combination of genetic algorithm (GA) [7] and simulated annealing (SA) [11] ideas to generate our nominal pareto front (Figure 2). Genetic algorithm and simulated annealing optimizations are sequentially used with underlying non-dominated sorting ideas to improve the pareto front. Each set of these sequences is termed an “epoch”. During the whole process, points are sorted using their non-domination rank and crowding distance

[7] for efficient, spread-out pareto surface generation with minimal number of points.



**Figure 2: Pareto front generation schedule.**

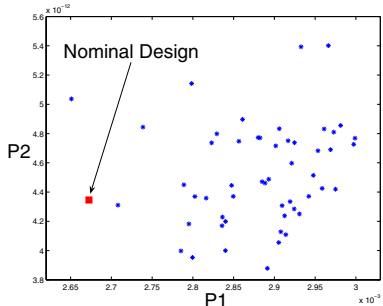
If we take a close look at GA, the two steps that bring in diversity to the population sample are: (a) crossover and (b) mutation. Crossover intermingles the parameter values of the highly ranked individuals in the population. Mutation, on the other hand, results in sampling of new regions of the parameter space that are close to the spatial location of the current individual. This step is similar to the choice of new points in the simulated annealing algorithm where we randomly pick a new candidate and probabilistically accept it after evaluating its set of performance functions. The heuristic of sequential application of GA and SA algorithm is based on the notion that GA based algorithms perform an excellent job of finding better candidates from a “mix” of the current population. However, the parameter space that they visit is rather constrained by the set of parameter values of the highly ranked individuals in the population. Passes of SA ensure a better stochastic search in the whole of the parameter space. Though, it might seem that the mutation step of GA is similar to a pure SA pass. However, the amount of effort that a standard GA implementation puts in mutation is quite less. Sequential runs of both of these algorithms provide the opportunity for new and (hopefully) attractive regions of the parameter space to be discovered (by SA) and the performance further improved amongst the candidate set (through GA).

This notion is similar to the concept of *ice ages* during the evolution of organisms on earth. During the warm periods, regular crossover and mutation of chromosomes occurred between the organisms leading to the development of better living creatures. However, during periods of extremely cold temperatures (*ice ages*), these superior organisms did not get the right environment to further their breed. Rather, some other inefficient (during the warm periods) organisms got the opportunity to mature, develop and compete against the superior organisms when the new “warm” periods ensued. We apply three such *epochs* (SA followed by GA) to generate our final pareto fronts.

#### 3.2 Yield-Aware Pareto Surfaces

As discussed in earlier sections, if we ignore the process variations during the generation of the pareto surface, the points on the pareto may have poor yield values even though they represent the best set of performances that the circuit can achieve for the nominal design. This is primarily because the global optimizer used to generate the pareto works extremely hard to push (improve) the nominal design. In such cases, it often pushes the nominal design towards the tail in the performance distribution curve to get the non-dominated point on the pareto. However if we plot the performance distribution across process variation for that particular nominal design corresponding to a point on the pareto, we would often find the nominal point lying along the edges of the resulting distribution (Fig. 3). This means that once the circuit is fabricated using the particular set of design parameters corresponding to the point on the pareto front, most of the circuits would have performances

much inferior to co-ordinates of the pareto point in the performance space.



**Figure 3: Distribution of process variation based Monte Carlo samples in the performance space for a design that was part of the pareto front. The nominal design lies at the edge of the distribution thereby resulting in very low yield if the corresponding co-ordinates of the nominal design are considered as performance specifications for the circuit.**

This behavior was observed in almost all the points on the pareto front that we generated for a particular circuit. Figure 3 serves as the motivation behind our work for generation of yield-aware pareto surfaces.

Formally, we define a *yield-aware pareto surface with a yield ‘Y’* as a surface in the performance space generated using a set of non-dominated performance points such that for each of these points, there exists atleast one set of design parameters that results in yield greater than ‘Y’. The yield being measured by setting the co-ordinates of the points in the performance space as the circuit’s performance specification.

A simple way of generating these fronts is to keep track of the points on the yield curves (for a specific yield number) for each set of design parameters during a global optimization routine and then trying to improve the front generated by these points. This can be done by generating Monte Carlo samples across process variations at each design point. However, the method is computationally prohibitive since it requires Monte Carlo simulations for each candidate point in the parameter space. In the next section, we describe an approximate but fast methodology for computing Monte Carlo distributions that can be used to characterize the yield surface for a particular design point.

### 3.3 Approximate Monte Carlo Simulation

Monte Carlo analysis for a particular design point (that we call *anchor points*) is generally very expensive since it requires evaluation of the performance function for large number of input process parameter sets. Latin Hypercube (LH) sampling [4] is the most common method to reduce the number of evaluations required while still ensuring reasonable accuracy in computing the performance distribution function.

In our methodology, we generate sets of 10 LH samples using variational information about the *process* parameters and we repeat this step 10 times, thereby generating 100 input parameter sets. We then find the performance distribution function for the circuit candidate using SPICE simulations. Due to the nonlinear nature of the circuits, the resulting distributions often do not conform to any standard distribution *e.g.*, Normal, Weibull *etc.* We use a transformation function to convert these distributions to a normal distribution. Each dimension (x) in the performance space is transformed into a new space (y) using either one of the *transformation equations* shown in Eqn. 6 such that the distribution function matches that of a *normal distribution* with the same mean and variance as that for the given data in the transformed space. This is an opti-

mization problem where values of the variables ( $a, b, c, d$ ) is to be determined such that our performance distribution data best *resembles* a Gaussian in the transformed space. The choice of these particular functions was made after experimentations which suggested that these equations provided the required fidelity for transforming arbitrary distributions, obtained during Monte Carlo simulations of the circuit, to a Normal distribution using minimal number of parameters. The optimization is performed using Brent-Powell based local optimization scheme [18]. The metric for *resemblance* is the sum of squared error between the distribution functions.

$$\begin{cases} y = a + b * (x + c)^d \\ y = a + b * \log(c * x + d) \end{cases} \quad (6)$$

Once the optimal transformation is obtained, we pick the set of 10 LH samples that best *resemble* the full performance distribution of the 100 LH samples in the transformed space. We call this set of 10 LH samples *optimal* latin hypercube sampling  $LHS_{opt}$  for the particular anchor point. We use  $LHS_{opt}$  to predict the distribution function of the circuit candidate in its performance space. For a design point close to the anchor point, we simulate the circuit for these 10  $LHS_{opt}$  points. The performance values obtained through simulations are then transformed into a new space by applying Eqn. 6 using the values for parameters  $a, b, c$  and  $d$  obtained through optimization in the previous steps. In the new transformed space, the performance distribution functions are expected to match a Normal distribution as was the case for the anchor point. Using the mean, variance and correlation matrix of these points in the transformed space, large number of pseudo points are generated for the underlying Normal distribution. The pseudo points are directly generated using the distribution functions and do not require actual circuit simulation. Hence, this step is extremely fast. Once, we have these large number of points, we can compute the yield of the circuit for that particular design point for any arbitrary performance goal. Alternatively, we can also compute the performance values which would give us a particular yield number for the circuit candidate under consideration.

The set of optimum LH samples works fine for approximating the performance distribution function in regions of the parameter space ( $\mathbf{p}$ ) that are *close* to the anchor point. We, therefore, need different  $LHS_{opt}$  for each region of the parameter space that we need to model.

### 3.4 Generating Yield-Aware Pareto

We first generate the nominal pareto front using the *new* non-dominated sorting based algorithm described in Sec. 3.1 (NSGA with ice ages). The primary pareto front thus generated is called the pareto front of order 1. A pareto front of order  $k$  is the front generated after removing all the points of pareto fronts of order 1 to  $(k - 1)$ .

We start with  $N$  equi-distant points on the  $k^{th}$  order front that we call *anchor points* (line 3 of Algorithm 1).  $k$  is usually chosen as 3 or 4 to give us a “soft” pareto front (line 2 of Algorithm 1). We then compute  $LHS_{opt}$  for these anchor points on the “soft” pareto front. Using  $LHS_{opt}$ , we compute the points in the performance space that would give a yield greater than some fixed quantity ‘Y’ (line 5 of Algorithm 1). The resulting set of points form a yield front. Figure 4 shows one such yield front computed for a particular design point with a yield of 80%.

We then try to improve this yield surface using local optimization techniques like Brent-Powell [18] (line 6 of Algorithm 1). At each yield evaluation for a particular design point during Brent-Powell optimization, we generate a set of points on the yield front (Figure 4). We use a non-dominated sorting algorithm to keep the yield points computed in the current evaluation and all previous



