Research Article | Vol. 32, No. 21 / 7 Oct 2024 / Optics Express 37585
Optics EXPRESS

Selecting robust silicon photonic designs after Bayesian optimization without extra simulations

ZHENGQI GAO,[1,*] ZHENGXING ZHANG,[1] ZICHANG HE,[2] JIAQI GU,[3] DAVID Z. PAN,[3] AND DUANE S. BONING[1]

[1]Department of EECS, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Department of ECE, University of California, Santa Barbara, CA 93106, USA
[3]Department of ECE, The University of Texas at Austin, Austin, TX 78712, USA
[*]zhengqi@mit.edu

Abstract: Optimization methods are frequently exploited in the design of silicon photonic devices. In this paper, we demonstrate that pushing the objective function to its minimum during optimization often results in devices that gradually become more sensitive to perturbations of design variables. The dominant strategy of selecting the design with the smallest objective function can lead to fabrication failure or yield loss due to manufacturing process variations. To address this issue, we propose an intuitive selection criterion that can identify designs not only possessing small objective functions but that are also robust to variations. Our simulation results on the Y-splitter, direction coupler, and bent waveguide designs demonstrate that the proposed method can achieve 2x higher coverage of robust designs with almost negligible run time, compared to the two baseline methods.

1. Introduction

Due to its higher bandwidth and lower power consumption compared to electronic circuits, integrated silicon photonics has attracted a great deal of attention. Many research efforts have been devoted to this emerging field from various perspectives including material science [1,2], photonic circuit architecture [3–5], numerical methods [6,7], and optimization techniques [8–13].

Among these sub-fields, photonic device optimization (also known as inverse design [8–16]) is one appealing and active branch. It starts by parameterizing the shape of a photonic device via a multi-dimensional design variable (i.e., a column vector), and next defines an scalar objective function capturing the design goal. After minimizing the cost/objective function with respect to the design variable, the optimal design is chosen as the one with the smallest objective value among all designs visited by the optimization routine. Generally, the optimization techniques used in the literature of photonic device optimization can be classified into two categories: (i) gradient-free methods [17–19], and (ii) gradient-based methods [12,13]. Gradient-free methods usually rely on genetic algorithm, particle swarm optimization [17], or Bayesian optimization [16,20]. On the other hand, gradient-based methods attempt to make gradient information available during the optimization, so that gradient descent methods can be applied. The adjoint method is the mainstream choice in terms of how to calculate gradient [12,13].

However, an important piece is missing from the present discussion — **how robust is the optimized design under perturbations of design variables?** It is well-known that manufacturing process variations [21–25] (such as line edge roughness [25]) can cause the fabricated device shape to deviate from the desired design. Thus, to make the optimization method of practical utility, we must ensure that the optimized design is relatively robust to the perturbations of design variables; otherwise, it is likely that the optimized design will function erroneously due to the introduced variations during fabrication [20,26].

In this paper, we first demonstrate that when pushing the objective function to its minimum during an optimization, the optimized design gradually becomes more sensitive to perturbations

#531213 https://doi.org/10.1364/OE.531213
Journal © 2024 Received 29 May 2024; revised 5 Aug 2024; accepted 16 Sep 2024; published 2 Oct 2024

Optics EXPRESS

of design variables. Motivated by this observation, we propose an intuitive selection criterion that can identify designs not only possessing small objective functions, but that are also robust to variations. Our analysis and methods are presented under the framework of Bayesian optimization. As will be demonstrated later, this choice is natural as the surrogate model learned in Bayesian optimization enables us identify robust designs without any extra simulations. Our key is to utilize the surrogate model to calculate second-order derivatives of the *objective function* around the local minimum, which can be regarded as approximations to first-order derivatives of *device performances* (i.e., a robustness indicator). Numerical simulations on directional coupler, Y-splitter, and bent waveguide photonic devices verify our findings and demonstrate the efficacy of our method compared to two baseline methods.

The remainder of this paper is organized as follows. In Section 2, we briefly review Bayeisan optimization and motivate our research. Next, we propose our method in Section 3 and verify its efficacy on three key sillicon photonic devices (i.e., directional coupler, Y-splitter, and bent waveguide) in Section 4. Finally, we conclude with Section 5.

2. Preliminary

The problem of photonic device optimization can be formulated as:

$$\min_{\mathbf{w}\in\Omega} \quad L(\mathbf{w}) \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^d$ is a column vector representing the design variables, and $\Omega \subset \mathbb{R}^d$ is the feasible space that $\mathbf{w}$ can reside in. As an example, we can choose a series of grid points on the boundary of the silicon photonic device and make their coordinates be the design variables $\mathbf{w}$. Then $L(\cdot)$ is a user-defined scalar function capturing the design intention (e.g., transmission loss of the device, power splitting ratio, free spectral range), and evaluating a function value $L(\mathbf{w})$ needs to invoke a time-consuming physical simulation (e.g., FDTD, EME, or FDFD) once.

Bayesian optimization [16,20] is a black-box global optimization technique. Its major steps are summarized in Algorithm 1. In Step 4, a Gaussian process regression (GPR) surrogate model $\mathcal{GP}(\mathbf{w})$ is learned based on Γ, a set of pairs of design and objective values. When being fed an input $\mathbf{w}$, the GPR model returns a probabilistic prediction following a Gaussian distribution:

$$\mathcal{GP}(\mathbf{w}) \sim \mathcal{N}(\mu, \sigma^2) \tag{2}$$

where $\mu = \mu(\mathbf{w}) \in \mathbb{R}$ and $\sigma = \sigma(\mathbf{w}) \in \mathbb{R}^+$ are both functions of $\mathbf{w}$. Here $\mu(\mathbf{w})$ can be regarded as an approximation to the real objective function $L(\mathbf{w})$, and $\sigma(\mathbf{w})$ as the prediction uncertainty. In Step 5, we minimize a user-defined acquisition function. One common choice of the acquisition function is lower confidence bound (LCB):

$$\mathrm{LCB}(\mathbf{w}) = \mu(\mathbf{w}) - \rho \cdot \sigma(\mathbf{w}) \tag{3}$$

where ρ>0 is a user pre-defined constant. Refer to [27] for more details on Bayesian optimization.

Consider the optimization of a Y-splitter as illustrated in Fig. 1. Our design goal is to make the normalized powers (with respect to the power of incident wave at port 1) transmitted at port 2 and port 3 both close to 0.5, and the power reflected back at port 1 close to 0. We desire that our design satisfies the above requirements in a given wavelength range $1.5\mu\mathrm{m} \le \lambda \le 1.6\mu\mathrm{m}$. To this end, the objective function is defined as:

$$L = \frac{1}{3N_{\mathrm{wav}}} \sum_{n=1}^{N_{\mathrm{wav}}} (p_n^{11})^2 + (p_n^{12} - 0.5)^2 + (p_n^{13} - 0.5)^2 \tag{4}$$

where p_n^{ij} represents the power of the fundamental TE mode from port i to port j at the n-th wavelength point. We run Bayesian optimization with $d = 13$, $N_{\mathrm{wav}} = 100$, $\rho = 0.3$, $N_{\mathrm{init}} = 30$, and $N_{\mathrm{max}} = 200$.

Optics EXPRESS

Algorithm 1. Bayesian Optimization

1: Set the number of initial simulations N_{init} and the number of total simulations $N_{\max}$. Define the design space Ω.
2: Uniformly sample N_{init} designs from Ω and invoke the simulator to evalute their function values, yielding $\Gamma = \{(\mathbf{w}_n, L_n = L(\mathbf{w}_n))|n = 1, 2, \cdots, N_{\text{init}}\}$.
3: **while** $n = N_{\text{init}} + 1 : 1 : N_{\max}$ **do**
4: Build a GPR model $\mathcal{GP}(\mathbf{w})$ based on Γ.
5: Minimizing Eq. (3) w.r.t. $\mathbf{w}$ gives a new design $\mathbf{w}_{\text{new}}$.
6: Invoke the simulator to evaluate $L(\mathbf{w}_{\text{new}})$.
7: Add $(\mathbf{w}_{\text{new}}, L(\mathbf{w}_{\text{new}}))$ into Γ.
8: **end while**
9: Return the design with the smallest L value in Γ.

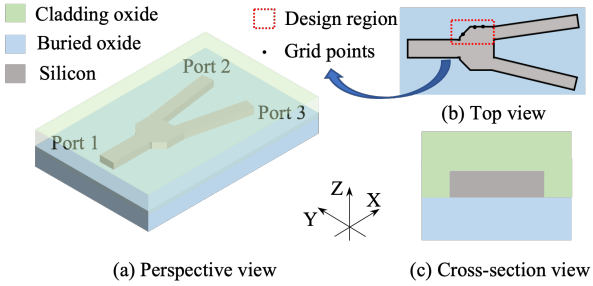

Fig. 1. A simplified schematic of a Y splitter, (a) perspective view, (b) top view, and (c) cross-section view.

After Bayesian optimization terminates, we deliberately select two optimized designs from Γ and plot their shapes and performance curves in Fig. 2. Although the top design possesses a slightly smaller objective value, its blue solid p^{12} curve (i.e., without perturbations) is similar to that of the design in the bottom row. Moreover, when perturbations are present, the top design degrades much more compared to the bottom design.

Figure 2 demonstrates a phenomenon that during optimization, the objective value is pushed to a minimum, but the optimized device becomes more vulnerable to perturbations. To further justify this finding, we perform two simulations for every design $\mathbf{w} \in \Gamma$ after Bayesian optimization finishes. One simulation is performed at the desired $\mathbf{w}$, recording $\{p_n^{12}|n = 1, 2, \ldots, N_{\text{wav}}\}$. The other is at the +5% relative perturbed $\mathbf{w}' = (1 + 5\%)\mathbf{w}$, recording $\{p_1^{12,\prime}|n = 1, 2, \ldots, N_{\text{wav}}\}$. Note that here +5% is just for illustration. In reality, this value should be set according to the variation of photonic devices that the specific manufacturing process introduces. Then we plot the maximum p^{12} change, i.e., $\max_{i=1}^{N_{\text{wav}}} |p_i^{12,\prime} - p_i^{12}|$, along with the objective value L for all 200 designs visited by Bayesian optimization in Fig. 3. As shown in the right figure, comparing the designs in the yellow and red circled regions, we find that they have similar objective values if viewing vertically, while those in red are much more robust to perturbations than those in yellow. Namely, the designs in the red circle are more desirable than those in the yellow circle. Strictly, the Pareto front [28] (i.e., the blue solid line in Fig. 3) shows the best values that we can achieve when desiring both performance change and objective value to be small. Thus, we look for an approach that can identify the designs on the Pareto front after Bayesian optimization finishes. Furthermore, since the Pareto front has a shape close to a 90 degree angle, it is even better if the proposed method can identify designs not only on the Pareto front, but also localized in the left bottom.

Optics EXPRESS

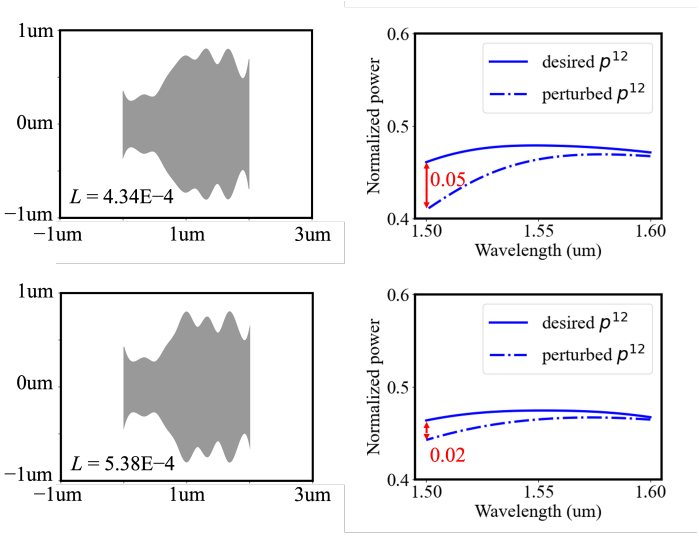

Fig. 2. We perform the iterative Bayesian optimization algorithm shown in Algorithm 1 to optimize the Y splitter shown in Fig. 1. Two different designs obtained at iteration 192 (top row) and at iteration 75 (bottom row) of Bayesian optimization are plotted. Compared to the bottom design, the design in the top row appears later in the optimization routine and achieves a better objective function value, but it is more susceptible to perturbations. For clarity, p^{11} and p^{13} are omitted. The dashed p^{12} curves are obtained by perturbing all design variable relatively by 5%, i.e., $\mathbf{w} \leftarrow (1 + 5\%)\mathbf{w}$.

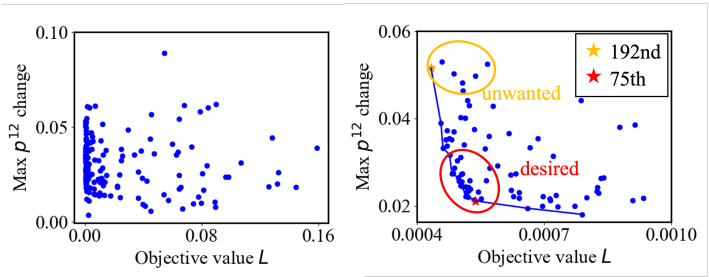

Fig. 3. Scatter plot of max p^{12} change along with objective value L for all 200 designs in Γ[20]. The right figure is a zoomed-in version of the left figure when $L <= 0.0010$. The blue solid line shows the Pareto front [28], and the designs located on this line are known as Pareto optimal. Additionally, the designs obtained at the 75th and 192nd iterations in Fig. 2 are also displayed on this scatter plot.

Optics EXPRESS

When Fig. 3 is available, we can easily achieve this goal. However, building such a figure requires us to run extra simulations after Bayesian optimization finishes. Namely, recall that in Algorithm 1, we only know the objective values for designs $\mathbf{w} \in \Gamma$, but not their performance changes. Thus, if we wish to obtain the vertical coordinates of points in Fig. 3, we have to run extra simulations at the perturbed values for every design in Γ after Bayesian optimization finishes, which will double the algorithm run time. This naturally raises a question: **How can we identify robust optimized designs without extra simulations?** In other words, under the framework of Algorithm 1, how can we pinpoint designs on the Pareto front (or in the left bottom red circle), with the limitation that we can only see horizontal coordinates in Fig. 3? The naive way used in Step 8 of Algorithm 1 gives us the iteration-192 design (with the smallest objective value, but the largest sensitivity), which does not work well as seen in the top row of Fig. 2. Consequently, an efficient selection method to identify robust designs in Γ is our goal in this paper.

3. Proposed method

In this section, we propose an intuitive and efficient method to address the aforementioned problem. To begin, we assume that the objective function has the following form:

$$L = \frac{1}{N}\sum_{n=1}^{N}(p_n - t_n)^2 \tag{5}$$

where t_n is a constant specifying a given performance target (e.g., 0 and 0.5 in Eq. (4)), p_n is some performance of the photonic device relying on the design variable $\mathbf{w} \in \mathbb{R}^d$, and N is the number of terms. We emphasize that this mean square error form of the objective function is generic and can encode many design goals such as power splitting, transmission loss, wavelength-division multiplexing, and resonance. In practical scenarios, the number of terms N can be significantly large, as these p_n values are associated with multiple ports (e.g., three ports in Eq. (4)) and numerous frequency/wavelength points (e.g., N_{wav} in Eq. (4)). Thus, we assume that the GPR model in Algorithm 1 only models the mapping from $\mathbf{w}$ to L, instead of $\mathbf{w}$ to each p_n, to save computational resources.

To compare the robustness of different designs, we need a robustness measure. For instance, we can use the following sensitivity measure:

$$\sum_{i=1}^{d}\sum_{n=1}^{N}(\frac{\partial p_n}{\partial w_i})^2 \tag{6}$$

as an indicator for robustness. To understand this quantity, we start from a first-order Taylor expansion of the performance vector $\mathbf{p} = [p_1, p_2, \ldots, p_N] \in \mathbb{R}^N$ at the desired $\mathbf{w} \in \mathbb{R}^d$:

$$\mathbf{p}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{p}(\mathbf{w}) + \mathbf{J} \cdot \Delta\mathbf{w} \tag{7}$$

where $\mathbf{J} \in \mathbb{R}^{N\times d}$ is the Jacobian matrix, and its entry on the i-th column ($i = 1, 2, \ldots, N$) and j-th row ($j = 1, 2, \ldots, d$) is $\partial p_i/\partial w_j$. Defining $\Delta\mathbf{p} = \mathbf{p}(\mathbf{w} + \Delta\mathbf{w}) - \mathbf{p}(\mathbf{w})$ and rearranging the above equation, we obtain:

$$\Delta\mathbf{p} = \mathbf{J} \cdot \Delta\mathbf{w} \tag{8}$$

When the L2 norm of $\Delta\mathbf{w}$ is fixed to a constant a (i.e., $||\Delta\mathbf{w}||_2 = a$), we have:

$$||\Delta\mathbf{p}||_2 = ||\mathbf{J} \cdot \Delta\mathbf{w}||_2 \le a||\mathbf{J}||_2 \le a||\mathbf{J}||_F = a[\sum_{i=1}^{d}\sum_{n=1}^{N}(\frac{\partial p_n}{\partial w_i})^2]^{\frac{1}{2}} \tag{9}$$

where $||\mathbf{J}||_2$ and $||\mathbf{J}||_F$ represent the induced L2 norm and Frobenius norm of the Jacobian matrix $\mathbf{J}$, respectively. In Eq. (9), we use Eq. (8) and the definition of induced L2 norm, and then use the inequality $||\mathbf{J}||_2 \le ||\mathbf{J}||_F$ and explicitly write out the Frobenius norm based on its definition.

Optics EXPRESS

Here Eq. (9) states that under a perturbation with fixed L2 norm $||\Delta\mathbf{w}||_2 = a$, the change of performance vector $||\Delta\mathbf{p}||_2$ is upper bounded by $a||\mathbf{J}||_F$. In other words, $||\mathbf{J}||_F$ could measure the robustness of $\mathbf{p}$ under perturbation of $\mathbf{w}$, pessimistically though. This motivates us to choose $\sum_{i=1}^{d}\sum_{n=1}^{N}(\frac{\partial p_n}{\partial w_i})^2$, the squared Frobenius norm of $\mathbf{J}$, as the robustness indicator. We emphasize that this choice is not unique, and that $[\sum_{i=1}^{d}\sum_{n=1}^{N}|\frac{\partial p_n}{\partial w_i}|^k]^{\frac{1}{k}}$ $(\forall k \in \{1, 2, \ldots\})$ could work as robustness indicator as well. However, our present choice is the most convenient to approximate as demonstrated later.

If we can evaluate this indicator for all designs, then we can address the problem by selecting designs with both small objective and small indicator value. However, the challenge here is that without extra simulation at the perturbed design point, we cannot evaluate this indicator value. Our key idea is simple: we seek to find a related metric that can be easily evaluated with the GPR model, and that is an approximation to the above robustness indicator.

To find such a metric, let us consider a well-optimized design $\mathbf{w}_{\text{opt}}$ that can achieve $L(\mathbf{w}_{\text{opt}}) \leq \epsilon$, where ϵ is a very small positive constant representing a user-defined "good design" threshold. Then for this design, we can derive:

$$\epsilon \geq L(\mathbf{w}_{\text{opt}}) = \frac{1}{N}\sum_{n=1}^{N}(p_n - t_n)^2 \geq \frac{1}{N}\max_n(p_n - t_n)^2 \tag{10}$$

by noticing that all terms inside the summation are no smaller than zero. This further gives:

$$\max_n |p_n - t_n| \leq \sqrt{N\epsilon}\ . \tag{11}$$

Intuitively, this says that for a well-optimized design, its residue term $|p_n - t_n|$ should be smaller than $\sqrt{N\epsilon}$ for all $n = 1, 2, \ldots, N$.

Next, we calculate the second derivative of L with respect to the i-th design variable (where $i = 1, 2, \ldots, d$):

$$\frac{\partial^2 L}{\partial w_i^2} = \frac{2}{N}\sum_{n=1}^{N}\left[(\frac{\partial p_n}{\partial w_i})^2 + (p_n - t_n)\frac{\partial^2 p_n}{\partial w_i^2}\right]\ . \tag{12}$$

If we move the first term on the right-hand side to the left and denote an intermediate variable $r_i = \frac{\partial^2 L}{\partial w_i^2} - \frac{2}{N}\sum_{n=1}^{N}(\frac{\partial p_n}{\partial w_i})^2$ for calculation simplicity, then we can derive:

$$\begin{aligned} |r_i| &= \left|\frac{\partial^2 L}{\partial w_i^2} - \frac{2}{N}\sum_{n=1}^{N}(\frac{\partial p_n}{\partial w_i})^2\right| = \left|\frac{2}{N}\sum_{n=1}^{N}(p_n - t_n)\frac{\partial^2 p_n}{\partial w_i^2}\right| \\ &\leq \frac{2}{N}\sum_{n=1}^{N}\left|(p_n - t_n)\frac{\partial^2 p_n}{\partial w_i^2}\right| = \frac{2}{N}\sum_{n=1}^{N}|p_n - t_n| \cdot \left|\frac{\partial^2 p_n}{\partial w_i^2}\right| \\ &\leq \frac{2}{N}\sum_{n=1}^{N}\sqrt{N\epsilon}\cdot\left|\frac{\partial^2 p_n}{\partial w_i^2}\right| = \frac{2\sqrt{\epsilon}}{\sqrt{N}}\sum_{n=1}^{N}\left|\frac{\partial^2 p_n}{\partial w_i^2}\right|\ . \end{aligned} \tag{13}$$

Built upon the above derivation, we can obtain:

$$\left|\sum_{i=1}^{d} r_i\right| \leq \sum_{i=1}^{d}|r_i| \leq \frac{2\sqrt{\epsilon}}{\sqrt{N}}\sum_{i=1}^{d}\sum_{n=1}^{N}\left|\frac{\partial^2 p_n}{\partial w_i^2}\right| \leq 2d\sqrt{N\epsilon}B \tag{14}$$

Optics EXPRESS

where we have defined B as the maximum of all second derivatives, i.e., $B = \max_{i,n} |\frac{\partial^2 p_n}{\partial w_i^2}|$. Now if we write out the summation of r_i, we observe:

$$\begin{aligned}\sum_{i=1}^{d} r_i &= \sum_{i=1}^{d} \left[\frac{\partial^2 L}{\partial w_i^2} - \frac{2}{N} \sum_{n=1}^{N} (\frac{\partial p_n}{\partial w_i})^2 \right] \\ &= \sum_{i=1}^{d} \frac{\partial^2 L}{\partial w_i^2} - \frac{2}{N} \sum_{i=1}^{d} \sum_{n=1}^{N} (\frac{\partial p_n}{\partial w_i})^2 \ .\end{aligned} \tag{15}$$

Combining this with Eq. (14), we obtain the key inequality:

$$\left| \sum_{i=1}^{d} \frac{\partial^2 L}{\partial w_i^2} - \frac{2}{N} \sum_{i=1}^{d} \sum_{n=1}^{N} (\frac{\partial p_n}{\partial w_i})^2 \right| \leq 2d\sqrt{N\epsilon}B \tag{16}$$

which further simplifies to:

$$| \underbrace{\frac{N}{2} \sum_{i=1}^{d} \frac{\partial^2 L}{\partial w_i^2}}_{\text{SED metric}} - \underbrace{\sum_{i=1}^{d} \sum_{n=1}^{N} (\frac{\partial p_n}{\partial w_i})^2}_{\text{robustness indicator}} | \leq d\sqrt{\epsilon} N^{\frac{3}{2}} B \ . \tag{17}$$

Since the GPR model $\mathcal{GP}(\mathbf{w})$ built in Bayesian optimization is an approximation to the mapping $L(\mathbf{w})$, the first term on the left-hand side of Eq. (17) can be approximated using the GPR model by numerical differentiation. Namely, we can approximate the second derivative of a scalar function $g(w)$ by $\frac{\partial^2 g}{\partial w^2} \approx \frac{g(x+\Delta x)+g(x-\Delta x)-2g(x)}{\Delta x^2}$ with $\Delta x \to 0$. Alternatively, if the GPR model is implemented in a programming language which supports automatic differentiation (such as Python with Pytorch, Jax, or Autograd add-on), then automatic differentiation is available. For later simplicity, we will denote this term as SED, short for sum of second-order derivatives.

The second term on the left-hand side of Eq. (17) is the sum of all squared first-order derivatives of p_n with respect to w_i, our desired robustness indicator. In other words, Eq. (17) provides a remarkable relationship between what we can calculate and what we desire: their discrepancy is upper bounded and related to the "good design" threshold ϵ, the number of terms N in the objective function, the dimension d of the design variable, and the maximum second derivative absolute value B. Most importantly, if $\epsilon \to 0$, the discrepancy goes to zero and we can regard $\text{SED} = \frac{N}{2} \sum_{i=1}^{d} \frac{\partial^2 L}{\partial w_i^2}$ as the robustness indicator. Motivated by this finding, we propose our robust design selection method in Algorithm 2.

Algorithm 2. Robust Design Selection

1: Set the value of ϵ, and the level of Pareto front N_{pl}.
2: Build the set $\Theta_{\text{cand}} = \{\mathbf{w} | (\mathbf{w}, L(\mathbf{w})) \in \Gamma, L(\mathbf{w}) \leq \epsilon\}$.
3: For all $\mathbf{w} \in \Theta_{\text{cand}}$, invoke $\mathcal{GP}(\mathbf{w})$ to numerically approximate $\text{SED}(\mathbf{w}) = \frac{N}{2} \sum_{i=1}^{d} \frac{\partial^2 L}{\partial w_i^2}$.
4: Update $\Theta_{\text{cand}} = \{\mathbf{w} \in \Theta_{\text{cand}} | \text{SED}(\mathbf{w}) \geq 0\}$.
5: **while** $n = 1 : 1 : N_{\text{pl}}$ **do**
6: Identify Pareto optimal designs $\mathbf{w} \in \Theta_{\text{cand}}$ according to the $(L(\mathbf{w}), \text{SED}(\mathbf{w}))$ metric pair.
7: Collect the Pareto optimal designs into a set Θ_n.
8: Update $\Theta_{\text{cand}} = \Theta_{\text{cand}} / \Theta_n$.
9: **end while**
10: Return $\{\Theta_1, \Theta_2, \cdots, \Theta_{N_{\text{pl}}}\}$.

Optics EXPRESS

In Step 1, we set the value of ϵ and define the level of Pareto front [28] N_{pl} (e.g., $N_{\text{pl}} = 2$). In Step 2, we narrow down consideration only focusing on those designs with performance $L \leq \epsilon$. In Step 3, we use the GPR model $\mathcal{GP}(\mathbf{w})$ to numerically approximate SED($\mathbf{w}$). Next in Step 4, due to the choice of ϵ as well as numerical error, we might get SED($\mathbf{w}$)<0 for some $\mathbf{w} \in \Theta_{\text{cand}}$. Ideally, this would not happen since we hope SED($\mathbf{w}$) works as an approximation to $\sum_{i=1}^{d}\sum_{n=1}^{N}(\frac{\partial p_n}{\partial w_i})^2$, which is always non-negative. As a simple heuristic remedy in cases where this does occur, we discard those designs with negative SED($\mathbf{w}$) values. Then at Step 5, we go into an iteration, where in the n-th iteration, a set Θ_n containing a few designs will be returned. In Step 6, we select the Pareto optimal [28] designs from Θ_{cand} based on the metric $(L(\mathbf{w}), \text{SED}(\mathbf{w}))$. Intuitively, this is to say scatter all the values $\{(L(\mathbf{w}), \text{SED}(\mathbf{w}))|\mathbf{w} \in \Theta_{\text{cand}}\}$ with L and SED as the horizontal and vertical coordinates, respectively, onto a figure, and then choose all $\mathbf{w}$ whose coordinates (L, SED) are located at the most bottom left (see Fig. 3 as an example). Since our method involves approximations, it is useful to perform several rounds (i.e., N_{pl}>1) and identify several levels of Pareto optimal designs. We will explain in details the benefit of multiple round later in Section 4.

Several important things need to be clarified here. First, there are two main sources of approximations which might introduce error: (i) the SED metric shown in Eq. (17) is calculated based on the GPR model $\mathcal{GP}(\mathbf{w})$ while the GPR model itself is an approximation to the real function mapping $L(\mathbf{w})$, and (ii) the SED metric is employed to approximate the desired robustness indicator. The first issue (i) is mitigated as the number of samples in Γ increases, while the second issue (ii) is alleviated by selecting a small ϵ.

Second, in implementing the calculation of SED($\mathbf{w}$), the constant factor $\frac{N}{2}$ can be neglected. Because our primary concern is to compete the values of SED($\mathbf{w}_1$) with SED($\mathbf{w}_2$) for two different designs $\mathbf{w}_1$ and $\mathbf{w}_2$, any positive constant scaling factor will not affect the result, as only the relative order matters.

Third, we discard those designs with negative SED($\mathbf{w}$) values in Step 4, because the robustness indicator in Eq. (17) is always non-negative, and as its approximation, the SED metric should also be non-negative. However, a bad approximation could happen when ϵ or B is large according to Eq. (17), and such negative values respectively correspond to a design with large objective value or large second-order derivatives. Note that the remedy strategies are not unique. For instance, we could adopt SED($\mathbf{w}$) = $\sum_{i=1}^{d}\max(0, \frac{\partial^2 L}{\partial w_i^2})$ in the definition to guarantee positiveness. In all simulation examples considered in this paper, we empirically found that the discarding strategy works better than setting negative second-order derivatives to zero.

Fourth, finding Pareto optimal designs in Step 6 is motivated by Fig. 3. The desired designs are usually located on the Pareto front in the scatter plot of performance p^{12} change versus objective value L. Without extra simulations, we do not know the values of performance p^{12} change, but instead, we have the SED metric. Thus, we fall back to identifying those Pareto optimal designs on the scatter plot of SED versus L. This is reasonable as long as SED is a good approximation to p^{12} change, which indeed is the case, as implied by Eq. (17).

Fifth, our derivations rely heavily on the objective function having a mean square error form as shown in Eq. (5). For instance, the second-order derivatives shown in Eq. (12) only hold true under a mean square error form. Although the specifics will differ, our derivations can be extended to objective functions like $L = \frac{1}{N}\sum_{n=1}^{N}|p_n - t_n|^k$ for any positive integer k, where $k = 2$ corresponds to the case considered in this paper. Dealing with other objective function forms might require substantially different derivation flows. Again, we re-iterate that the objective function shown in Eq. (5) is generic and can encode many design goals.

Sixth, we emphasize that the proposed robust selection scheme is intended to work with gradient-free optimization methods that utilize surrogate models such as Bayesian optimization. For gradient-based optimization methods, we have the gradient values $\frac{\partial p_n}{\partial w_i}$ so that we can directly calculate the robustness indicator in Eq. (17) without needing the SED metric.

Optics EXPRESS

Last, but not least, we want to mention that incorporating robust design selection into Bayesian optimization is a natural choice since the surrogate GPR model enables us to conveniently calculate derivatives. This is impossible for other gradient-free optimization methods without a surrogate model. Moreover, the design variable in Bayesian optimization is chosen as a series of grid points on the boundary of a silicon photonic device, so that their perturbations have a physical meaning — corresponding to shape perturbation. In contrast, for gradient-based adjoint optimization, their design variable represents the permittivity of the design region. Although the derivative of device performance with respect to design variable is available, it does not directly correspond to any process variation (e.g., device width/height variation) introduced during manufacturing.

4. Numerical results

To verify our method, we perform numerical simulation on Y-splitter, directional coupler, and bent waveguide devices. All photonic devices are defined in 3D and simulated using Lumerical FDTD on a RedHat Linux server with 3TB memory and 16 CPU cores. The algorithm is implemented using Python. In the Y-splitter example, we will show that the calculated SED metric might be negative and our heuristic remedy (i.e., discarding designs with negative SED value) is sufficient. Later in the directional coupler and bent waveguide examples, we explain why we need Algorithm 2 to be iterative and have several levels of Pareto front (i.e., $N_{pl}>1$).

4.1. Y-Splitter

In this subsection, we verify our method on the Y-splitter example shown in Fig. 1. We choose $\epsilon = 0.0010$ following Fig. 3. We choose $N_{\text{pl}} = 1$ in this example, so that only the first-level of Pareto optimal set Θ_1 will be returned. As shown in Fig. 4, we use the GPR model $\mathcal{GP}(\mathbf{w})$ to numerically approximate SED($\mathbf{w}$) for those designs $\mathbf{w}$ with $L(\mathbf{w}) \leq \epsilon$. We observe that SED<0 usually occurs when L is relatively large, which coincides with our previous expectation. After discarding those designs with SED<0, we identify the Pareto front and construct Θ_1. Next, as shown in Fig. 5, we scatter all identified designs onto the golden performance change figure. We observe that a large portion of our points are located at the most bottom left as desired. Remarkably, most of our identified designs (i.e., red dots) are located on the real Pareto front (i.e., the blue solid line in Fig. 3), justifying SED as a substitution of the unknown max p^{12} change, and demonstrating that our methods works well in this example.

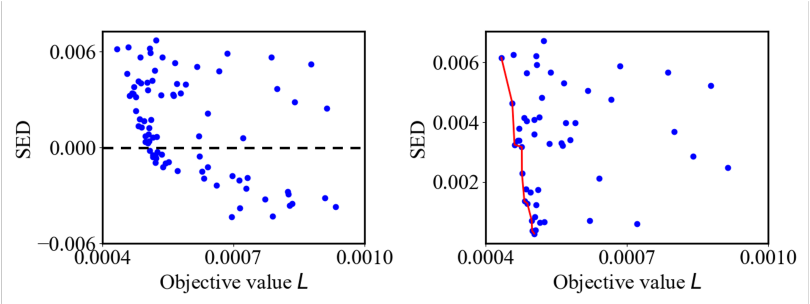

Fig. 4. Left: Scatter plot of $(L(\mathbf{w}), \text{SED}(\mathbf{w}))$ for all $\mathbf{w}$ with $L(\mathbf{w}) \leq \epsilon = 0.001$. Right: Keep those designs $\mathbf{w}$ with $\text{SED}(\mathbf{w}) \geq 0$. The red solid line shows the Pareto optimal designs we find, making up of Θ_1.

To better understand the effectiveness of our method, it is educational to compare it with other baseline methods. Under the condition of no extra perturbed simulations, there are two reasonable strategies: (i) choosing designs with smallest objective values, and (ii) randomly

Optics EXPRESS

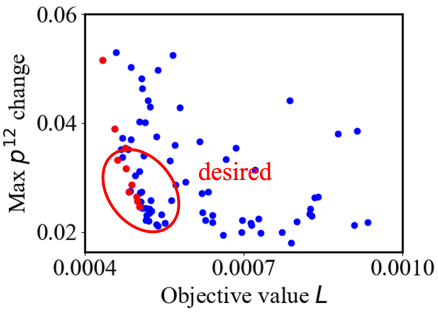

Fig. 5. Scatter plot of our identified designs $\mathbf{w} \in \Theta_1$ (i.e., the red dots) onto the golden performance change figure (i.e., the right side in Fig. 3)

choosing designs with objective values smaller than ϵ. We carry out these two strategies and scatter the resulting designs by each of these alternatives onto the golden performance change figure, respectively. The results are shown in Fig. 6. Clearly, counting the red, cyan, and pink dots inside the desired region, we see that our proposed method performs the best. Moreover, when the number of simulations $N_{\max}$ becomes even larger (i.e., pushing objective function harder to the minimum), then the cyan dots in the left of Fig. 6 will be more squeezed out of the desired region. On the other hand, randomly selecting designs bears too much uncertainty. Table 1 summarizes performance of different methods on this Y-splitter example. Our method identifies 10 designs, and seven of them are located in the desired region as shown in Fig. 5. Moreover, since only the GPR model is used to approximate the robustness indicator, the algorithm finishes in less than 10 seconds. Selection based on smallest objective value and random selection also run fast, but with fewer designs residing in the desired region. Alternatively, the golden method needs to invoke one perturbed simulation for each design with objective value L smaller than ϵ. In our experiment, 83 designs satisfy the requirement $L \leq \epsilon$, and each simulation takes about 150 seconds, resulting in an overall run time of about 3.45 hours. We conclude this example by emphasizing that since in Fig. 1, we choose the design variable $\mathbf{w}$ as the Y-coordinates of several grid points in the design region, perturbing $\mathbf{w}$ corresponds to width variation in manufacturing. Our method identifies optimized designs robust to such.

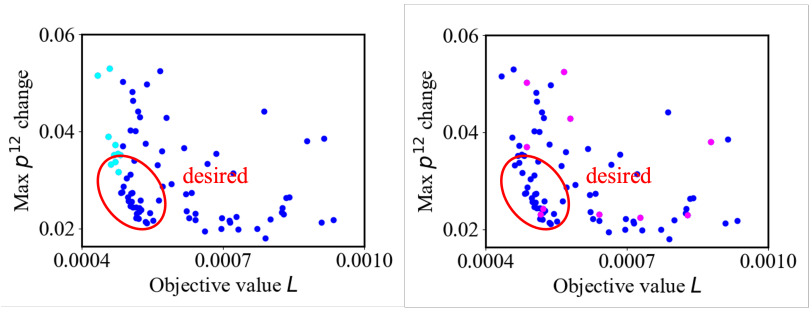

Fig. 6. Left: Select the designs with the smallest objective values. Right: Randomly select designs. For fairness, the number of cyan (or pink) dots here is the same as that of red dots in Fig. 5.

4.2. *Directional coupler*

In this subsection, we consider applying our method to the design of a 50% : 50% directional coupler (DC) shown in Fig. 7. We use the first entry of $\mathbf{w}$ to represent the height h, and its

Optics EXPRESS

Table 1. Performances of different methods on Y-splitter

	Ours	Golden	Smallest	Random
# Extra Sims.	0	83	0	0
Runtime	< 10 s	3.45 hr	< 10 s	< 10 s
Coverage	7 / 10	10 / 10	3 / 10	2.8 ± 0.7[a] / 10

[a]For fair comparison, the coverage metric for random selection is calculated after running the experiments independently five times, and reported in the format 'mean'±'std'.

remaining $(d-1)$ entries are used to encode the design flexibility in the XY plane. Specifically, we uniformly choose $(d-1)$ grid points on the top left arm. Their X-coordinates are fixed once the grids are generated, while their Y-coordinates are regarded as our design variables. The objective function in this example is defined as:

$$L = \frac{1}{2N_{\text{wav}}} \sum_{n=1}^{N_{\text{wav}}} \left[(p_n^t - 0.5)^2 + (p_n^c - 0.5)^2 \right] \tag{18}$$

where p_n^t and p_n^c represent, respectively, the normalized power of the fundamental TE mode at the through and cross port at the n-th sampled wavelength point. We implement Bayesian optimization with $d = 11$, $\rho = 0.3$, $N_{\text{init}} = 40$, and $N_{\max} = 600$. In this example, we set $\epsilon = 0.0014$ and choose $N_{\text{pl}} = 3$ to explain the loop of Algorithm 2. The results are shown in Figs. 8–10.

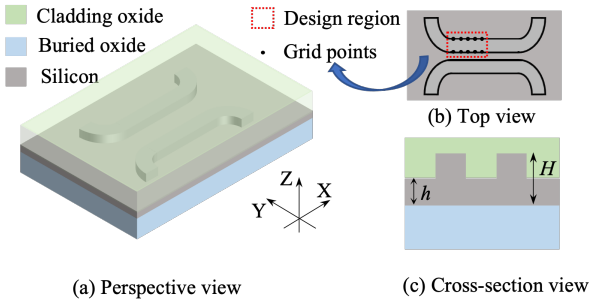

Fig. 7. A simplified schematic of a rib-waveguide-based directional coupler (DC); (a) perspective view; (b) top view; and (c) cross-section view.

As shown in Fig. 8, we see that the red solid line is located at the most bottom left, and then the yellow and green lines follow. The motivation to use a multiple-level Pareto front is two fold. First, as shown in this example, with only one level, there are five identified designs (red dots). The characteristics of the robust design selection problem, i.e., the trade-off between objective value and robustness metric, indicates that there are multiple designs of interest. When sent to manufacturing, we can utilize this property and simultaneously fabricate a few designs on one single wafer [23]. However, if we use only one single level Pareto front, we only have five designs in this case, which might be fewer than wanted. The second reason is that since our method is still an approximation, one single level Pareto front might be insufficient and miss some good robust designs. These can be captured by succeeding levels. Compared to the strategies shown in Fig. 10, our method again performs the best in this example.

4.3. *Bent waveguide*

In this subsection, we consider the design of a bent waveguide as shown in Fig. 11. A bent waveguide is usually used in photonic circuits to route light, and low loss is highly desired. In this

Optics EXPRESS

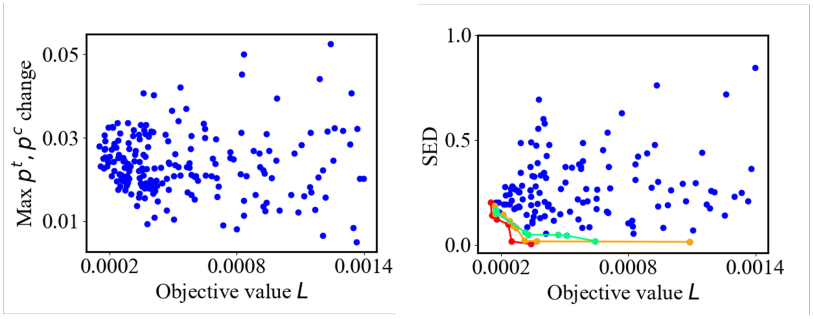

Fig. 8. Left: Scatter plot of max p^t and p^c change under a relative +5% change (i.e., $\mathbf{w} \leftarrow (1 + 5\%)\mathbf{w}$) along with objective value L for all 600 designs in Γ. Right: We calculate the SED metric and keep those designs $\mathbf{w}$ with SED($\mathbf{w}$) ≥ 0. The red, yellow, and green solid lines are, respectively, the identified first, second, and third-level Pareto front, and these designs make up Θ_1, Θ_2, and Θ_3.

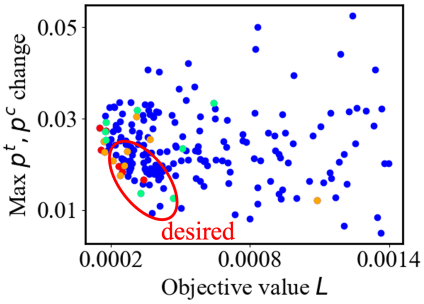

Fig. 9. Scatter plot of designs $\mathbf{w} \in \Theta_1, \Theta_2, \Theta_3$ onto the golden performance change figure (i.e., the left side in Fig. 8). Designs from Θ_1, Θ_2, and Θ_3 are represented by 6 red, 9 yellow, and 8 green dots, respectively.

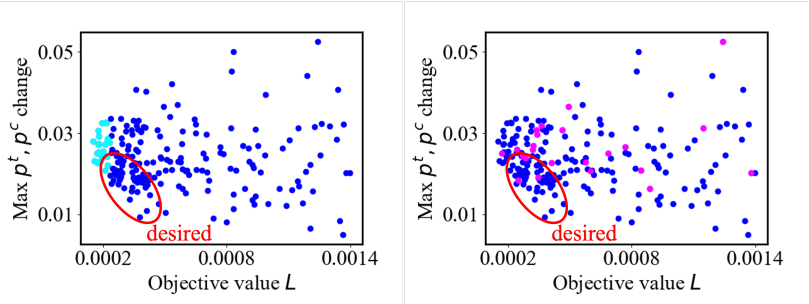

Fig. 10. Left: Select the designs with the smallest objective values. Right: Randomly select designs. For fairness, the number of cyan (or pink) dots here is the same as the sum of red, yellow, and green dots in Fig. 9.

Optics EXPRESS

example, we fix the width of the bending and evenly choose d grid points in polar coordinates. Specifically, the angles of grid points are determined once we set $d = 6$, and their radii are the design variables. The objective function in this example is defined as:

$$L = \frac{1}{N_{\text{wav}}} \sum_{n=1}^{N_{\text{wav}}} (p_n - 1)^2 \tag{19}$$

where p_n represents the normalized power of the fundamental TE mode at the output port at the n-th sampled wavelength point. We implement Bayesian optimization with $d = 6$, $\rho = 0.3$, $N_{\text{init}} = 40$, and $N_{\text{max}} = 160$. We set $\epsilon = 0.003$ and choose $N_{\text{pl}} = 2$. The results are shown in Table 2. As in the previous examples, compared to the strategies of minimum selection or random selection, our method again performs the best.

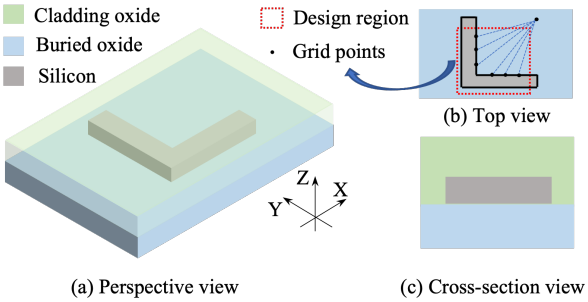

Fig. 11. A simplified schematic of a rib-waveguide-based directional coupler (DC); (a) perspective view; (b) top view; and (c) cross-section view.

Table 2. Performances of different methods on bent waveguide

	Ours	Golden	Smallest	Random
# Extra Sims.	0	63	0	0
Runtime	< 10 s	32.5 mins	< 10 s	< 10 s
Coverage	10 / 18	18 / 18	7 / 18	6.5 ± 2.1[a] / 18

[a]For fair comparison, the coverage metric for random selection is calculated after running the experiments independently five times, and reported in the format 'mean'±'std'.

5. Conclusions

In this paper, we demonstrate that during optimization of photonic devices, as the objective value is gradually pushed to a minimum, the design can become more sensitive to perturbations. We propose a robustness metric SED, short for sum of second-order derivatives, that can be numerically approximated using the Gaussian process regression model provided in Bayesian optimization. Next, using the pairs of SED and objective values, we extract Pareto optimal designs that are shown to possess small objective value and good robustness simultaneously. Our simulation results verify that the proposed method can achieve high coverage of robust designs with almost negligible run time. The suggested selection criterion can be seen as a straightforward extension to identify robust designs after the completion of a conventional Bayesian optimization.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Optics EXPRESS

References

1. W. Bogaerts, L. Van Iseghem, P. Edinger, *et al.*, "Low-power electro-optic actuators for large-scale programmable photonic circuits," in *Conference on Lasers and Electro-Optics (CLEO)*, (IEEE, 2021).
2. P. Edinger, A. Y. Takabayashi, and C. Errando-Herranz, "Silicon photonic microelectromechanical phase shifters for scalable programmable photonics," Opt. Lett. **46**(22), 5671–5674 (2021).
3. W. Bogaerts, D. Pérez, and J. Capmany, "Programmable photonic circuits," Nature **586**(7828), 207–216 (2020).
4. T. Kim, P. Bhargava, and C. V. Poulton, "A single-chip optical phased array in a wafer-scale silicon photonics/CMOS 3D-integration platform," IEEE J. Solid-State Circuits **54**(11), 3061–3074 (2019).
5. Y. Shen, N. C. Harris, and S. Skirlo, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**(7), 441–446 (2017).
6. Z. Zhu and T. G. Brown, "Full-vectorial finite-difference analysis of microstructured optical fibers," Opt. Express **10**(17), 853–864 (2002).
7. T. W. Hughes, I. A. D. Williamson, M. Minkov, *et al.*, "Wave physics as an analog recurrent neural network," Sci. Adv. **5** (2019).
8. M. H. Tahersima, K. Kojima, and T. Koike-Akino, "Deep neural network inverse design of integrated photonic power splitters," Sci. Rep. **9**(1), 1368 (2019).
9. W. Ma, Z. Liu, and Z. A. Kudyshev, "Deep learning for the design of photonic structures," Nat. Photonics **15**(2), 77–90 (2021).
10. K. Kojima, M. H. Tahersima, and T. Koike-Akino, "Deep neural networks for inverse design of nanophotonic devices," J. Lightwave Technol. **39**(4), 1010–1019 (2021).
11. Z. Liu, D. Zhu, L. Raju, *et al.*, "Tackling photonic inverse design with machine learning," Adv. Sci. **8**(5), 2002923 (2021).
12. A. Y. Piggott, J. Lu, and K. G. Lagoudakis, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," Nat. Photonics **9**(6), 374–377 (2015).
13. T. W. Hughes, M. Minkov, I. A. Williamson, *et al.*, "Adjoint method and inverse design for nonlinear nanophotonic devices," ACS Photonics **5**(12), 4781–4787 (2018).
14. C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, *et al.*, "Adjoint shape optimization applied to electromagnetic design," Opt. Express **21**(18), 21693–21701 (2013).
15. J. Peurifoy, Y. Shen, and L. Jing, "Nanophotonic particle simulation and inverse design using artificial neural networks," Sci. Adv. **4** (2018).
16. Z. Gao, Z. Zhang, and D. S. Boning, "Automatic synthesis of broadband silicon photonic devices via Bayesian optimization," J. Lightwave Technol. **40**(24), 7879–7892 (2022).
17. Y. Zhang, S. Yang, and A. E.-J. Lim, "A compact and low loss Y-junction for submicron silicon waveguide," Opt. Express **21**(1), 1310–1316 (2013).
18. P. Sanchis, P. Villalba, and F. Cuesta, "Highly efficient crossing structure for silicon-on-insulator waveguides," Opt. Lett. **34**(18), 2760–2762 (2009).
19. Y. Zhang, S. Yang, and A. E.-J. Lim, "A CMOS-compatible, low-loss, and low-crosstalk silicon waveguide crossing," IEEE Photon. Technol. Lett. **25**(5), 422–425 (2013).
20. Z. Gao, Z. Zhang, and D. S. Boning, "Automatic design of a broadband directional coupler via Bayesian optimization," in *Conference on Lasers and Electro-Optics*, (Optica Publishing Group, 2022), p.JW3B.156.
21. Y. Xing, J. Dong, and S. Dwivedi, "Accurate extraction of fabricated geometry using optical measurement," Photonics Res. **6**(11), 1008–1020 (2018).
22. Y. Xing, J. Dong, U. Khan, *et al.*, "Correlation between pattern density and linewidth variation in silicon photonics waveguides," Opt. Express **28**(6), 7961–7968 (2020).
23. Z. Zhang, S. I. El-Henawy, C. A. R. Ocampo, *et al.*, "Inference of process variations in silicon photonics from characterization measurements," Opt. Express **31**(14), 23651–23661 (2023).
24. Y. Xing, J. Dong, U. Khan, *et al.*, "Capturing the effects of spatial process variations in silicon photonic circuits," ACS Photonics **10**, 928–944 (2022).
25. Z. Zhang, M. Notaros, and Z. Gao, "Impact of process variations on splitter-tree-based integrated optical phased arrays," Opt. Express **31**(8), 12912–12921 (2023).
26. Z. He and Z. Zhang, "PoBO: A polynomial bounding method for chance-constrained yield-aware optimization of photonic ICs," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **41**(11), 4915–4926 (2022).
27. B. Shahriari, K. Swersky, and Z. Wang, "Taking the human out of the loop: A review of Bayesian optimization," Proc. IEEE **104**(1), 148–175 (2016).
28. Z. Gao, J. Tao, F. Yang, *et al.*, "Efficient performance trade-off modeling for analog circuit based on Bayesian neural network," in *Intl. Conf. on Computer-Aided Design (ICCAD)*, (2019).