

Improving thin-film manufacturing yield with robust optimization

Jonathan R. Birge,^{1,*} Franz X. Kärtner,² and Omid Nohadani³

¹Research Laboratory of Electronics and Department of Electrical and Computer Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

²Research Laboratory of Electronics and Department of Electrical and Computer Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

³School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907, USA

*Corresponding author: birge@mit.edu

Received 2 August 2010; accepted 26 August 2010;
posted 13 September 2010 (Doc. ID 132773); published 14 October 2010

A novel robust optimization algorithm is demonstrated that is largely deterministic, and yet it attempts to account for statistical variations in coating. Through Monte Carlo simulations of manufacturing, we compare the performance of a proof-of-concept antireflection (AR) coating designed with our robust optimization to that of a conventionally optimized AR coating. We find that the robust algorithm produces an AR coating with a significantly improved yield. © 2010 Optical Society of America
OCIS codes: 310.4165, 310.5696.

1. Introduction

Manufacturing errors often limit the performance of thin-film devices [1]. This is especially the case for certain emerging applications of thin-film filters that require extremely high precision responses, such as narrow bandpass filters for telecom and dispersion compensating mirrors for ultrafast optics, to name just a few. There has been a significant amount of recent activity toward developing design approaches that mitigate the effects of manufacturing errors. These efforts may be broadly grouped into two classes: methods that aim to predict the most manufacturable design among a preexisting set of generated stacks, and methods that seek to directly produce designs that are more robust to manufacturing tolerances. Chief among the former class is a recent work by Tikhonravov and Trubetskov, which very accurately simulates the full manufacturing

process [2], including the feedback effects of optical monitoring that can lead to statistical correlations between layers. This approach has the benefit of incorporating as much information as possible about the actual production process, with the downside being that such simulations are not amenable themselves to being directly optimized. On the other hand, approaches that seek to directly minimize some statistic of the coatings produced must, by necessity, involve some assumptions and approximations about the manufacturing, as it is not currently computationally feasible to optimize a full simulation of the coating process. An example of work in this vein is the stochastic optimization algorithm developed by Yakovlev and Tempea, which has proven effective in improving the sensitivity of chirped mirrors [3]. However, their approach, being nondeterministic, is computationally more involved than standard deterministic gradient descent-based optimization.

As disclosed in recent papers [4,5], we have developed an approach to robust optimization that is entirely deterministic. It involves minimizing the

maximum merit function found within a region of uncertainty surrounding a trial solution, where possible errors may occur. That is, we seek a solution which minimizes the worst-case error caused by manufacturing errors limited by some bound. While this “min–max” method is computationally efficient, by necessity it involves significant simplifying assumptions on the nature of the errors. Furthermore, minimizing the worst-case error does not necessarily result in a design that is manufacturable in a real-world context.

In a commercial thin-film coating facility, the manufacturing yield is likely the most relevant quantity to optimize. Were computational resources unconstrained, the best design approach would be to perform a stochastic optimization of the yield predicted by a full simulation of the manufacturing process. In practice, however, even a local optimization of the yield is intractable for anything but the simplest of coatings, as the yield predicted by Monte Carlo simulation is a piecewise constant function of the design parameters, making computation of a meaningful gradient difficult. Moreover, simply computing the yield itself involves the significant burden of Monte Carlo simulations with large n , especially when comparing trial solutions with either low or high yields [6].

In a previous work, we developed a method to perform deterministic robust optimization for a chirped-mirror design by minimizing the worst-case merit function found within a subset of the hypersphere surrounding a point in the design space. This subset was chosen in an *ad hoc* attempt to reconcile the performance of the deterministic algorithm with the statistical reality of actual manufacturing processes [5]. In the final analysis, however, the algorithm is still limited by the fact that it involves a deterministic gradient search in the space surrounding a trial solution and is susceptible to being unduly affected by isolated spikes in the merit function. While the approach did succeed in improving the expected variance of a chirped dispersion compensating mirror pair, subsequent study indicated that the yield was not improved except when relatively large layer errors were assumed.

In light of the above, we address the issue of yield by adding a postprocessing step to the robust optimization routine. This final step consists of a one-dimensional global search along a path in the state space between the min–max optimum and the nominal optimum to find the intermediate design with the best yield. In this paper, we refer to yield as the fraction of designs that satisfy some defined acceptable standard. We find that yields are improved significantly by this refinement, and yet the step does not add significantly to the run time of the algorithm because only one dimension is searched.

Our robust optimization algorithm is designed to replace a gradient descent routine in the larger context of a design approach (such as needle optimization [7] or gradual evolution). We propose that

considering robustness explicitly, albeit in the context of approximations to the manufacturing process, is a more efficient and effective way to mitigate manufacturing errors than doing more accurate simulations after optimization to rank a set of potential designs.

2. Method

Our method begins by first performing a very brief standard gradient descent, with a termination criteria several orders of magnitude away from what would normally be used. (We use a relative merit function change of 1/1000.) This avoids the robust optimization becoming stuck in a local extrema too early, and yet gives it enough room to find a solution different from that which would normally be found by a standard nominal optimization.

We then seek a minimum worst-case design \mathbf{x} using the algorithm described in detail by Nohadani *et al.* [5]. This algorithm, at the core of our approach, probes the merit function in a bounded space of potential layer thickness errors. While this results in a much more computationally involved optimization, the result is more robust to significant perturbation as the full structure of the merit function is considered in a neighborhood around a nominal solution.

For the purposes of this min–max optimization, we model manufacturing errors as independent random sources of additive noise, because any known systematic errors, such as miscalibration, can be best addressed in the actual production. Therefore, we assume that, when manufacturing a mirror with layer thicknesses given by \mathbf{x} , statistically independent additive implementation errors $\Delta\mathbf{x} \in \mathbb{R}^n$ may be introduced due to variation in the coating process, resulting in an actual thickness of $\mathbf{x} + \Delta\mathbf{x}$. We assume a mean of zero and a variance on each layer that is motivated by actual manufacturing errors. Here, $\Delta\mathbf{x}$ resides within an uncertainty set $\mathcal{U} := \{\Delta\mathbf{x} \in \mathbb{R}^n \mid \|\Delta\mathbf{x}\|_2 \leq \Gamma\}$. Note that $\Gamma > 0$ is a scalar describing the size of perturbation against which the design needs to be protected. For this paper, we took the manufacturing uncertainty to be normally distributed with a standard deviation of $\sigma = 1$ nm. The *robust optimization method* seeks to minimize the worst-case cost via

$$\min_{\mathbf{x}} g(\mathbf{x}) \equiv \min_{\mathbf{x}} \max_{\Delta\mathbf{x} \in \mathcal{U}} f(\mathbf{x} + \Delta\mathbf{x}). \quad (1)$$

When implementing a certain design $\mathbf{x} = \hat{\mathbf{x}}$, the possible realization due to implementation errors $\Delta\mathbf{x} \in \mathcal{U}$ lies in the neighborhood set $\mathcal{N} := \{\mathbf{x} \mid \|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \Gamma\}$.

Therefore, $g(\hat{\mathbf{x}})$, is the maximum cost attained within \mathcal{N} . Let $\Delta\mathbf{x}^*$ be the worst implementation error at $\hat{\mathbf{x}}$, $\Delta\mathbf{x}^* = \arg \max_{\Delta\mathbf{x} \in \mathcal{U}} (\hat{\mathbf{x}} + \Delta\mathbf{x})$. Then, $g(\hat{\mathbf{x}})$ is given by $f(\hat{\mathbf{x}} + \Delta\mathbf{x}^*)$. Because we seek to navigate away from all the worst implementation errors, the inner maximization problem needs to be solved first. We conduct local gradient ascent searches to determine the worst configurations within \mathcal{N} .

One of the risks of this approach is that the algorithm compromises nominal performance too much to avoid a worst-case error that is very unlikely to occur in practice. In other words, the algorithm might try to avoid a very localized but high spike in the merit function space. One promising class of “rare” perturbations to eliminate from consideration are those that have strong resonances, manifesting as spikes in the autocorrelation of the thickness perturbation vector. Indeed, we found that, when the maximization was unconstrained, the maxima generally either fell into this class or consisted of shifts of all the layers in the same direction. We thus restrict the search to error vectors whose Fourier components have uniform amplitude [5] except for the dc component, which is constrained to zero. This has the additional benefit of allowing us to obtain the set of local maxima using a simple unconstrained optimization over the Fourier component phases. While this constraint cannot be rigorously justified beyond the fact that the optical density of a multilayer stack as a function of distance is closely related to the Fourier transform of the transmission function [8], we have empirically observed that the constraint is necessary for meaningful improvements in yield.

Once worst-case neighbors are identified, a direction is sought along which an updated neighborhood would not include these worst-case scenarios any longer. This direction is a vector that spans the largest angle $\Theta \geq 90^\circ$ to all worst implementation errors at $\hat{\mathbf{x}}$ in the set of worst implementation errors $\mathcal{U}^*(\hat{\mathbf{x}}) = \arg \max_{\Delta \mathbf{x} \in \mathcal{U}} f(\hat{\mathbf{x}} + \Delta \mathbf{x})$. To navigate away from the elements in $\mathcal{U}^*(\hat{\mathbf{x}})$, a descent direction \mathbf{d}^* can be found efficiently by solving the following second-order cone problem:

$$\begin{aligned} & \underset{\mathbf{d}, \beta}{\text{minimize}} \quad \beta \\ & \text{subject to} \quad \|\mathbf{d}\|_2 \leq 1 \\ & \quad \mathbf{d} \cdot \Delta \mathbf{x}^* \leq \beta \quad \forall \Delta \mathbf{x}^* \in \mathcal{U}^*(\hat{\mathbf{x}}) \\ & \quad \beta \leq -\epsilon, \end{aligned} \quad (2)$$

where ϵ is a small positive scalar, β is an auxiliary variable, and $\|\cdot\|_2$ is the vector length. A feasible solution to Eq. (2), \mathbf{d}^* , forms the maximum possible angle θ_{\max} with all $\Delta \mathbf{x}^*$. This angle is always greater than 90° due to the constraint $\beta \leq -\epsilon < 0$, as illustrated in Fig. 1. This constraint guarantees that \mathbf{d}^* will provide an updated design neighborhood that excludes all known $\Delta \mathbf{x}^*$. When Eq. (2) is infeasible, then $\hat{\mathbf{x}}$ is a robust local minimum. Reference [4] provides a detailed discussion of the actual implementation.

The last step is to perform a standard gradient descent minimization using the actual merit function to convergence. The complete path through the configuration space that is taken by the nominal optimization is recorded. The yield predicted by a full Monte Carlo simulation of the manufacturing process is then maximized along the aforementioned path using a simple one-dimensional search algorithm.

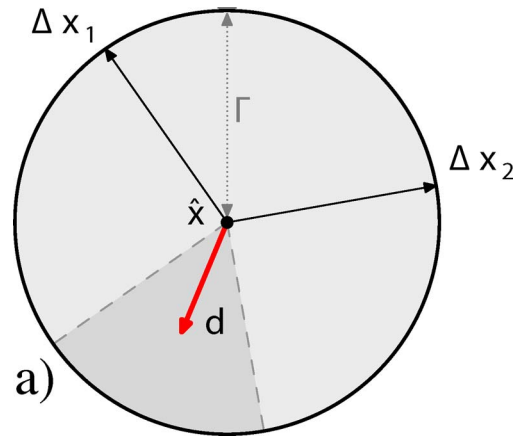


Fig. 1. (Color online) Two-dimensional illustration of the neighborhood. For a design $\hat{\mathbf{x}}$, all possible implementation errors $\Delta \mathbf{x} \in \mathcal{U}$ are contained in the shaded circle. The bold arrow \mathbf{d} shows a possible descent direction, and thin arrows $\Delta \mathbf{x}_i^*$ represent worst errors.

The latter constrained optimization is computationally feasible regardless of the complexity of the thin film being considered. The search simply occurs over the one-dimensional space parameterized by the position along the path from robust to nominal optimum, involving simulating the manufacturing statistics for several hundred candidate designs. We also note that because of the reduced complexity of this search, it is feasible to abandon the simple stochastic model inherent to the min-max stage and simulate the full manufacturing process when computing the simulated yield. It would be possible, for instance, to do a full simulation of an on-line calibration, as in Ref. [2].

Our motivation for this local postrefinement step is that minimizing the worst-case scenario is tantamount to maximizing the yield for the extreme pathological manufacturing error distribution concentrated at the extremes of the potential error bounds. On the other hand, a local nominal optimization maximizes the “yield” for the opposing extreme of zero errors. It is thus at least plausible that improvement can be made by searching between the two extremes, although we hasten to note that we can provide no rigorous justification for why this postprocessing approach works as well as it does nor a theoretical proof that it will do so for any design.

3. Results

To demonstrate the efficacy of our approach, we designed an antireflection (AR) coating for the visible, intended to be robust against manufacturing errors of 2.5 nm rms deviation and maximize yield for an average reflectance of 1% or below. The core min-max optimization was performed with a shell radius of 4.3 nm, empirically determined to be appropriate for layer thickness tolerances of the order of 2 nm. The starting point was a series of 21 thin layers of around $1/8$ of a wavelength, alternating between SiO_2 and Nb_2O_5 , beginning with the latter against a fused silica substrate. The merit function used

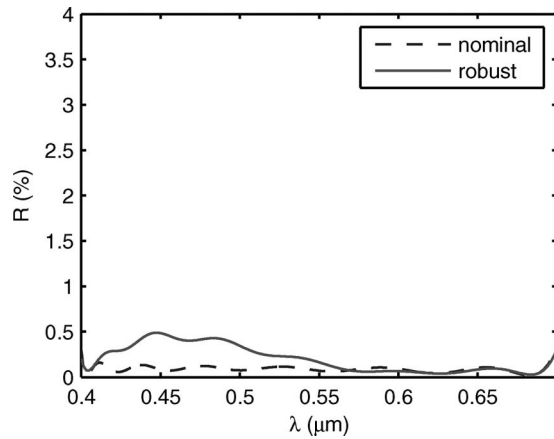


Fig. 2. Reflectances of the nominal- and robust-optimized AR coatings. The nominal design has an average reflectance of roughly 0.1%, and the robust design has a reflectance of around 0.2%.

was the total optical power reflected, assuming a uniform source spanning 400 to 700 nm in the wavelength domain. A quasi-Newton optimization was used for the nominal optimization, as well as the inner maximization used in the robust problem [Eq. (1)].

The robust algorithm was compared to the results from a single gradient descent from the same starting point, with no opportunity for additional layers. As our robust optimization algorithm is designed to replace a gradient descent routine in the larger context of a design approach (such as needle optimization [7] or gradual evolution), we can only fairly compare the results from the robust algorithm with a single gradient descent. Our goal with the present work was to compare our robust routine with that of a gradient descent to determine if any advantage in yield was attainable.

For the final yield maximization, we modeled the manufacturing process as producing layers with normally distributed errors, hard limited to always have a positive layer thickness. We note that this model was chosen for simplicity, and that it is also possible to specify a manufacturing model that involves non-identically distributed errors, or even correlated errors. This final stage typically only took around 5 min on an eight-core PC, leaving significant room

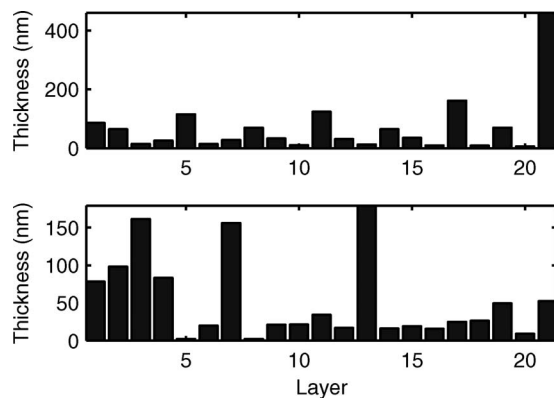


Fig. 3. Layer thicknesses of nominal-optimized (top) and robust-optimized (bottom) designs.

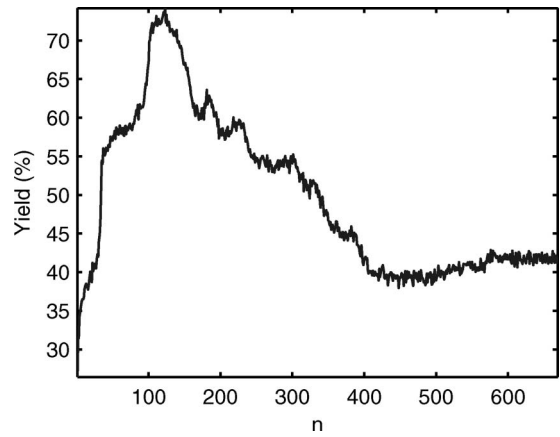


Fig. 4. Yield as a function of the final postoptimization from the robust point to the nominal optimum, parameterized by iteration number. The last stage of our robust optimization routine is to search this function for the maximum yield point typically found relatively close to the robust optimum despite having a significantly better yield.

to introduce extra complexity in the manufacturing model. For this purely numerical test, however, there was no motivation to do so.

We began with a standard optimization to a high precision relative termination of 10^{-8} . This nominally optimized design has an average reflectance of 0.1%. We then applied the robust algorithm, starting with a random perturbation of the same starting point (to avoid falling into a local minima close to our first design). As expected, the robust design has a worse nominal performance, with an average reflectance of roughly 0.2%. The theoretical performance (i.e., assuming no deposition errors) of the two designs is compared in Fig. 2. The significant increase in reflectance for the robust design in the absence of manufacturing defects is the ubiquitous “price of robustness.” The resulting layer thicknesses for both designs are shown in Fig. 3, demonstrating that the outcomes are nonintuitive.

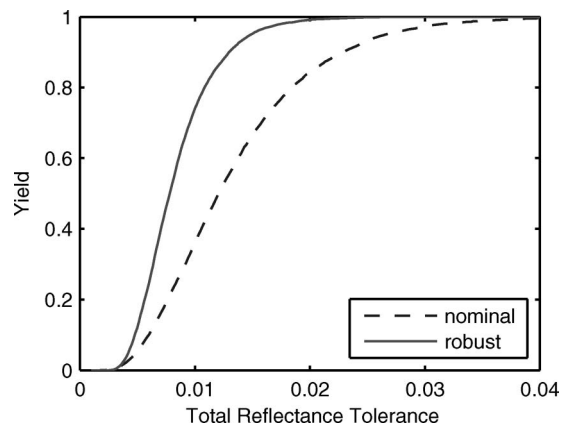


Fig. 5. Manufacturing yield of both robust- and nominal-optimized designs as a function of performance cutoff, assuming layer thickness errors of 2.5 nm rms. At the design cutoff of 1%, the robust design has a yield of 74% versus 36% for the nominally optimized design we started with.

To illustrate the significant improvement in yield that is found by searching along the path intermediate to the robust and nominal extrema, in Fig. 4 we show the simulated yield as a function of optimization iteration. In Fig. 5 we show the expected yield (i.e., the percentage of coatings that possess a minimum performance) for a range of target reflectivities, assuming a rms layer variation of 2.5 nm. Despite being designed to have more than twice the reflectivity, the robust design yields significantly more than the nominal stack above a tolerance of about 0.1% average reflectance. The yield for the robust stack at the design cutoff reflectance of 1% is more than double that of the nominal design.

4. Conclusions

Our robust optimization method explicitly seeks to minimize the worst-case performance as a starting point for a more computationally tractable search for designs with better manufacturing yield. We show, in the case of an AR coating, that this approach is capable of significant improvement in manufacturing yields compared to straight gradient descent optimization. As our approach is ultimately intended to replace gradient descent optimization within the context of a larger design evolution, the next step will be to integrate our scheme into a design framework such as gradual evolution or needle optimization. Doing so, and actually testing the results in a manufactured design, would be a necessary ultimate validation of our approach.

The authors gratefully acknowledge support from the National Science Foundation (NSF) under contract AST-0905592, the Defense Advanced Research Projects Agency (DARPA), HR0011-05-C-0155, and the National Aeronautics and Space Administration (NASA) under grant NNX10AE68G.

References

1. B. Sullivan and J. Dobrowolski, "Deposition error compensation for optical multilayer coating. I. Theoretical description," *Appl. Opt.* **31**, 3821–3835 (1992).
2. A. Tikhonravov and M. Trubetskov, "Computational manufacturing as a bridge between design and production," *Appl. Opt.* **44**, 6877–6884 (2005).
3. V. Yakovlev and G. Tempea, "Optimization of chirped mirrors," *Appl. Opt.* **41**, 6514–6520 (2002).
4. D. Bertsimas, O. Nohadani, and K. M. Teo, "Robust nonconvex optimization for simulation-based problems," *Oper. Res.* **58**, 161 (2010).
5. O. Nohadani, J. R. Birge, F. X. Kärtner, and D. J. Bertsimas, "Robust chirped mirrors," *Appl. Opt.* **47**, 2630–2636 (2008).
6. A. V. Tikhonravov, M. K. Trubetskov, and T. V. Amotchkina, "On the reliability of computational estimations used for choosing the most manufacturable design," in *Optical Interference Coatings*, OSA Technical Digest (Optical Society of America, 2010), paper TuA3.
7. A. Tikhonravov, M. Trubetskov, and G. DeBell, "Application of the needle optimization technique to the design of optical coatings," *Appl. Opt.* **35**, 5493–5508 (1996).
8. P. Verly, "Fourier transform technique with refinement in the frequency domain for the synthesis of optical thin films," *Appl. Opt.* **35**, 5148–5154 (1996).