Technical University of Denmark

DTU

Design of structurally colored surfaces based on scalar diffraction theory

Johansen, Villads Egede; Andkjær, Jacob Anders; Sigmund, Ole

DTU Library
Technical Information Center of Denmark

Design of structurally colored surfaces based on scalar diffraction theory

Villads Egede Johansen,* Jacob Andkjær, and Ole Sigmund

Technical University of Denmark, Department of Mechanical Engineering, 2800 Kgs. Lyngby, Denmark
*Corresponding author: vejo@mek.dtu.dk

In this paper we investigate the possibility of controlling the color and appearance of surfaces simply by modifying the height profile of the surface on a nanoscale level. The applications for such methods are numerous: new design possibilities for high-end products, color engraving on any highly reflective surface, paint-free text and coloration, UV-resistant coloring, etc. In this initial study, the main focus is on finding a systematic way to obtain these results. For now the simulation and optimization is based on a simple scalar diffraction theory model. From the results, several design issues are identified: some colors are harder to optimize for than others, and some can be produced by only a few height levels, whereas others require more complex structures. It is shown that a wide range of results can be obtained. © 2014 Optical Society of America

OCIS codes: (230.1950) Diffraction gratings; (050.2770) Gratings; (050.6624) Subwavelength structures; (240.6700) Surfaces; (330.1690) Color; (330.7326) Visual optics, modeling.
http://dx.doi.org/10.1364/JOSAB.31.000207

1. INTRODUCTION

Structural colors are colors caused by the interaction of light with small structures—most of them containing features comparable to the wavelength of light [1–3]. This makes it possible to obtain new colors (or appearances) that cannot be obtained through absorption-based methods like dyeing. Typical examples of structural colors include CDs, which contain gratings in the micrometer range, thin-film interference in soap bubbles due to their micrometer thickness, and the colorful appearance of peacock feathers due to even more complex microstructural effects [2,4].

Already in 1665 the effect of structural colors was described in the literature by Robert Hooke: "In this we are able from a colourless body to produce several coloured bodies, affording all the variety of Colours imaginable" [5]. This was only a few years after he formulated the wave theory of light, and since then many light experiments and color phenomena have been described. This is particularly true regarding the era after the introduction of Maxwell's equations. Furthermore, the increase in microscope resolution and computational power in recent decades has made it possible to study micro- and nanostructures and their interaction with light on an even more detailed level. This is evident in the detailed study of the microstructure of the Morpho butterfly (see [6–8]).

A lot of effort has been put into these investigations, and it has even been possible to successfully reproduce the color effect of the Morpho butterfly [9]. Few, though, have tried to approach structural colors from a different aspect than biomimicry. In this paper we will consider the inverse problem: if a color effect is desired, which grating structure can represent it? Based only on variations of the surface profile, this paper studies how well color effects can be obtained using gradient-based optimization of structures analyzed using scalar diffraction theory (SDT). More specifically, the results presented here examine whether it is possible to create a constant color within a certain angular interval.

The overall aim is to explore the limits of structural colors. Applications for structural coloring include altering the color of certain products even though they are made of the same material by, e.g., changing the casting mold (plastic coloration, preprinted text) [10], new color appearances for product design, extremely durable colors due to UV stability [11], security labeling [12], improved optical materials (antireflective surfaces, such as moth eye structures [11], color filters [13], and solar cell films [14]), and more eco-friendly production processes that avoid the use of paint.

It is important to stress that structuring an objects surface is just one out of many fundamentally different methods of obtaining structural colors. Whereas gratings are often mass produceable due to cast molding and etching technologies, other methods with different purposes have been investigated, and a wide range of colors have been produced. Examples are surface plasmons on structured surfaces with deposited metallic layers [15,16], controlling absorption of coated metallic nanoparticles by controlling size parameters [17], and plasmon effects in carbon nanostructures [18]. All of these methods are capable of producing a variety of colors, but with little control of their appearance with respect to angular distribution. However, in optics the control of angular distribution and intensities for different purposes has been studied. In [19] intensities for periodic narrowband structures were optimized using the nongradient-based particle swarm optimization method. In [13] surfaces were optimized for solar cell use. A number of methods for improving antireflective surfaces exist, one example of which is given in [20]. The design of such gratings can be seen as a special case (periodic and single frequency) of the optimization formulation

developed in this work. To the best of the authors' knowledge the only other works exploring the possibilities of systematic design of surfaces with angular distributed colors are [21] and [22]. These papers consider more complex distributions of material based on the topology optimization approach (c.f. [23]).

The rest of the paper is organized as follows. Section 2 describes the implementation of the physics and how to obtain a far-field response for the reflection of a surface, Section 3 describes how to convert light spectra to RGB colors, Section 4 states the proposed optimization problem for color optimization, Section 5 presents results from the optimization, and Section 6 draws conclusions from the study.

2. SCALAR DIFFRACTION THEORY

All simulations in this paper are based on SDT. This theory is widely used for calculating diffraction gratings and surface scattering problems such as estimating reflection from rough surfaces. The simplicity of the mathematical expressions in SDT means that calculations can be made rapidly, and they are thus well suited for use in optimization problems where repetitive simulations are required. SDT—and its shortcomings—is well described in the literature, and the most central result in this theory is the relation between far-field irradiance, E, and a complex amplitude distribution, U, emerging from the surface of an aperture (which later will be used as a model for the actual surface profile) [24–26]:

$$E(x', y') = \frac{1}{\lambda^2 z^2} \left| \mathcal{F}\{U(x,y)\}\left(\frac{x'}{\lambda z}, \frac{y'}{\lambda z}\right) \right|^2, \quad (1)$$

where λ is the wavelength and the coordinate components x, y, z, x', y' are as shown in Fig. 1. $\mathcal{F}$ represents the Fourier transform and is defined as

$$\mathcal{F}\{U(x,y)\}(\xi,\eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x,y)e^{-2\pi i(x\xi+y\eta)}\mathrm{d}x\mathrm{d}y. \quad (2)$$

The strength of SDT lies in the fact that many different physical setups can be described by defining U the right way. For the purpose of this paper, U will describe the complex phase
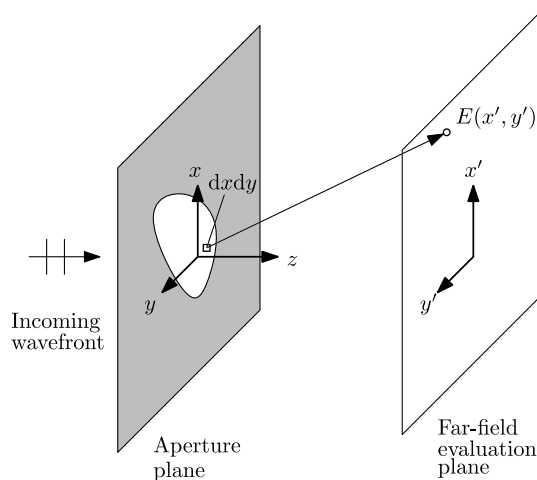
modulation of a plane wave incident on a structured surface (see Fig. 2 for a 2D example). Using this setup for a material with constant reflection R, the complex amplitude distribution is then described as

$$U(x,y) = Ae^{2\pi i 2(h_{\mathrm{ref}}-h(x,y))/\lambda} = Ae^{-4\pi i h(x,y)/\lambda}e^{4\pi i h_{\mathrm{ref}}/\lambda}, \quad (3)$$

where $A = R$ in the region of interest and $A = 0$ (nonreflecting/totally absorbing) elsewhere. Since the phase information will be of no interest in the far field, the constant phase part with h_{ref} is dropped for convenience. Furthermore, this paper will not consider material properties, but just a full reflecting surface in free space. Therefore the amplitude will be dropped as well, so that the complex amplitude function is described as

$$U(x,y) = \mathrm{rect}_{A_s}(x,y)e^{-4\pi i h(x,y)/\lambda}, \quad (4)$$

where $\mathrm{rect}_{\Omega}(x)$ is a function that is 1 for $x \in \Omega$ and 0 otherwise. A_s is the region of interest.

A. Perception of Radiated Intensity
The expression in Eq. (1) is not easy to relate to the visual appearance of a surface. First, appearance is more naturally perceived with respect to viewing angle (in degrees), and second, irradiance is not proportional to the response of the human eye. Instead radiance will be used, as it possesses this property [27]. The coordinate system is first scaled and transformed such that

$$\hat{x} = x/\lambda, \qquad \hat{y} = y/\lambda, \qquad \hat{z} = z/\lambda \quad (5)$$

and

$$\alpha = \hat{x}/\hat{r}, \qquad \beta = \hat{y}/\hat{r}, \qquad \gamma = \hat{z}/\hat{r}, \quad (6)$$

where $\hat{r} = \sqrt{\hat{x}^2 + \hat{y}^2}$. α and β are referred to as direction cosines, and their relation to a conventional spherical coordinate system is [28]

$$\alpha = \sin\theta\cos\phi, \quad (7)$$

$$\beta = \sin\theta\sin\phi, \quad (8)$$

$$\gamma = \cos\phi. \quad (9)$$

Fig. 1.   Geometry for the scalar diffraction setup.
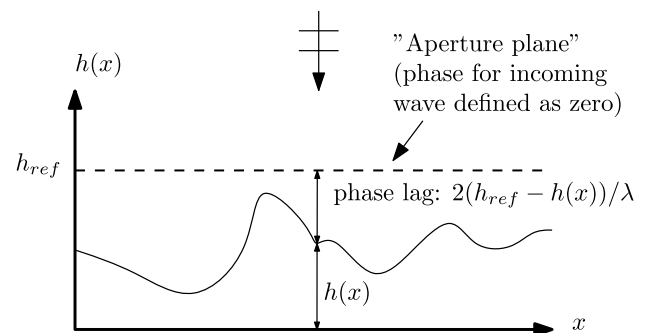

Fig. 2.   Illustration of how a reflective grating can be turned equivalent to a complex distribution on an aperture plane. The phase lag is due to the extra distance traveled by the wave.

Using this transformation, the radiance L (and not irradiance) is given as [28,29]

$$L(\alpha, \beta) = \frac{\lambda^2}{A_s}|\mathcal{F}\{U(\hat{x}, \hat{y}, 0)\}(\alpha, \beta)|^2$$
$$= \frac{1}{A_s}|\mathcal{F}\{U(x, y, 0)\}(\alpha/\lambda, \beta/\lambda)|^2, \qquad (10)$$

where A_s is the area of the aperture, and the last equality can be shown by substitution of variables in the Fourier integral. In spherical coordinates this means that the relation between radiance and irradiance far from the observer is given by $L(\theta, \phi) = r^2 E(\theta, \phi)/(A_s \cos \theta)$, where r is the distance from the source to the observation point. In two dimensions (the $\phi = 0$ plane) the expression in Eq. (10) simplifies to

$$L(\alpha) = \frac{1}{L_0}|\mathcal{F}\{U(x)\}(\alpha/\lambda)|^2, \qquad (11)$$

where L_0 is now the length of the aperture. This will be the starting point for all subsequent derivations.

B. Limitations of SDT

The advantage of SDT is the physical knowledge to be gained from its simplicity, and the disadvantage is that some physical properties are neglected due to its simplicity (see, e.g., [30] for a discussion on the interpretation of diffraction gratings). Inherent in the previous expressions is a small-angle (paraxial) approximation [28], meaning that the expressions lose their validity if energy ends up outside of the hemisphere above the surface. This is where $\theta \notin [-90^\circ, 90^\circ]$ (see, e.g., Fig. 5). This can in our case be corrected for [28] by normalizing L such that the energy when integrating over $\theta \in [-90^\circ, 90^\circ]$ corresponds to the incoming energy, since we assume full reflection. The extension is called nonparaxial SDT. Since we are optimizing for small angles and normal incoming light and do not want structures with features approaching the element resolution size of the model, which would give rise to this effect due to the high-spatial-frequency content [28], we have not found it necessary to do so—only when performing checks on the validity of the final result obtained. Other important limitations are that SDT takes only "first surface reflections" into account, which means that no multiple scattering or surface plasmon effects can be captured in this way, and that in two dimensions SDT best corresponds to TM-polarized light. This is because the TM polarization has the E-field going out of plane in Fig. 2, and the boundary conditions do not require zero field for electrical components parallel to a full reflecting plane. For components perpendicular to the plane, zero field is required, thus making TE polarization a worse approximation, since on nonhorizontal edges the boundary condition is violated.

C. Discretization of SDT

Normally, diffraction gratings described by SDT are evaluated using a fast Fourier transform (FFT). In this paper, another approach is proposed, which consists of representing the surface using rectangular sections that all have contributions that can be found analytically. This means that the formulation can be used for very large as well as small elements, always yielding a correct solution (to the SDT formulation) and not having to oversample or zero pad the FFT.

Furthermore, the gradients for the system can be found analytically. It can also be implemented with a constant memory consumption independent of the number of elements. The trade-off for this is speed. If a continuous function must be analyzed, it can be approximated by taking the average continuous value in an interval, as when discretizing any other physical structure (see Fig. 3).

For all simulations in this paper, the rectangular sections all have a constant width d and are divided into N elements. If $\{h_n\}_{n=0}^{N-1}$ are the corresponding N heights, as seen in Fig. 3, then the height function is

$$h(x) \approx \sum_{n=0}^{N-1} \text{rect}(x/d - n)h_n,$$

$$\text{where} \quad \text{rect}(x) = \begin{cases} 1 & \text{for } 0 < x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \qquad (12)$$

Finding the response of this structure is straightforward due to the linearity of the Fourier transform and the well-known solution of the Fourier transform of a rectangle:

$$L(\alpha) = \frac{1}{L_0}|\mathcal{F}\{\text{rect}_{L_0}(x)e^{-4\pi ih(x)/\lambda}\}(\alpha/\lambda)|^2$$

$$= \frac{1}{Nd}\left|\mathcal{F}\left\{\sum_{n=0}^{N-1}\text{rect}_{[0,1]}(x/d - n)e^{-4\pi ih_n/\lambda}\right\}(\alpha/\lambda)\right|^2$$

$$= \frac{1}{Nd}\left|\sum_{n=0}^{N-1}e^{-4\pi ih_n/\lambda}\mathcal{F}\{\text{rect}_{[0,1]}(x/d - n)\}(\alpha/\lambda)\right|^2$$

$$= \frac{1}{Nd}\left|\sum_{n=0}^{N-1}e^{-4\pi ih_n/\lambda}de^{-\pi id\alpha/\lambda}e^{2\pi ind\alpha/\lambda}\text{sinc}(d\alpha/\lambda)\right|^2$$

$$= \left|\underbrace{e^{-\pi id\alpha/\lambda}\text{sinc}(d\alpha/\lambda)}_{\text{const.shape}}\frac{1}{N}\sum_{n=0}^{N-1}\underbrace{e^{-4\pi ih_n/\lambda}}_{\text{height}}\underbrace{(e^{2\pi id\alpha/\lambda})^n}_{\text{translation}}\right|^2. \qquad (13)$$

The different products of this expression all have a physical interpretation. First, the response is shaped by $C(\alpha) = e^{-\pi id\alpha/\lambda}\text{sinc}(d\alpha/\lambda)$, which is due to the choice of rectangles as the interpolation function. If d is large, then the sinc function will give rise to unavoidable zeros in the angular domain (this is also in agreement with the physics), but for $d\alpha/\lambda \to 0$, the expression converges toward 1, thus not influencing the response if the approximation of h is well resolved. $H_n = e^{-4\pi ih_n/\lambda}$ describes the phase change to which the height at n gives rise, and $T_n(\alpha) = (e^{2\pi id\alpha/\lambda})^n$ modulates the phase change with a factor corresponding to the translation of the structure with respect to the origin. Note that $T_n = (T_1)^n$. It
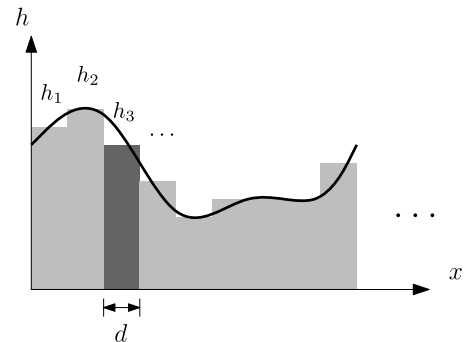

Fig. 3.   Discretization of a surface into rectangular bumps.

can furthermore be seen from this expression how all element contributions are independent of each other. This means that there is no coupling or other interaction between the elements, which is also an implicit assumption when using SDT. The expression can now more compactly be written as

$$L(\alpha) = \left| C(\alpha) \frac{1}{N} \sum_{n=0}^{N-1} H_n(\alpha) T_n(\alpha) \right|^2. \tag{14}$$

1. Implementation

To calculate the responses in an efficient way, two steps are taken. First, H_n is precalculated and stored in matrices for a wide range of equally spaced heights, so when a height is given, the response is calculated as a linear interpolation of the two closest heights. This could in general give rise to two problems: first, such interpolation makes the gradient piecewise linear, thus not giving continuous derivatives for the gradient-based optimization algorithm, and second, a linear interpolation of numbers on the complex unit circle will not yield correct values. In practice, though, no problems have been observed regarding these issues, as the interpolation points are quite close. For all of the experiments in this paper, a spacing of 1 nm is used. In the second step, it is not necessary to calculate all the translation matrices, because of their recursive relationship. For our choice of implementation it is favorable with respect to speed to implement the calculations using a recursive approach defined by

$$P = C \frac{1}{N} \hat{P}_N, \quad \text{where} \quad \hat{P}_n = \begin{cases} T_1 \hat{P}_{n-1} + H_{N-1-n} & \text{for } n > 0, \\ H_{N-1} & \text{for } n = 0, \end{cases} \tag{15}$$

where P satisfies $L = |P|^2$. This formula can be intuitively understood as starting with the rightmost rectangle at the origin, translating it d, adding the element that should be to the left of it, translating the new structure d, and then continuing this process until all elements have been "pushed" into the right order. It is interesting to note that the T_1 and $\hat{P}_n$ are matrices whose size is determined as the product of the number of wavelengths and the number of reflection angles. The H_n values depend on the number of *height discretization* steps and the number of wavelengths, and there is as such no dependence on the number of elements used. The calculations therefore have a constant memory consumption independent of the number of elements—except the array in which the element heights are stored. The solving time increases linearly due to the recursive summation. This is because there is no coupling between the elements.

2. Gradient Calculation

As the optimization is gradient based, the gradients of the reflected field with respect to height changes are needed. Due to the fact that every element is decoupled from every other, the gradients with respect to the height variation that are required for the optimization can simply be found as

$$\frac{\partial P}{\partial h_j} = C \frac{1}{N} \sum_{n=0}^{N-1} T_n \frac{\partial}{\partial h_j} e^{-4\pi i h_n / \lambda}$$

$$= -\frac{4\pi i}{N\lambda} C T_j e^{-4\pi i h_j / \lambda} = -\frac{4\pi i}{N\lambda} C T_j H_j. \tag{16}$$

3. COLOR CONVERSION

The color conversion is performed using

$$R = \int_0^\infty I(\lambda) \bar{r}(\lambda) d\lambda, \tag{17}$$

$$G = \int_0^\infty I(\lambda) \bar{g}(\lambda) d\lambda, \tag{18}$$

$$B = \int_0^\infty I(\lambda) \bar{b}(\lambda) d\lambda, \tag{19}$$

where $\bar{r}, \bar{g}, \bar{b}$ are the color-matching functions shown in Fig. [4] and $I(\lambda)$ is the light spectrum to be converted [31]. The color-matching functions are obtained by taking the CIE 1931 2° standard observer color-matching functions, $\bar{x}, \bar{y}, \bar{z}$, which are found in [32], and converting them to RGB color-matching functions by a linear transform with values as specified in [33], where the white reference used is the D65 illuminant that resembles the sun. This procedure is the same as in [22]. To simplify the notation, we will refer to colors as three-dimensional vectors, i.e., $C = (R, G, B)^T$.

A. Randomization

In the literature it has been shown how a random height variation of a repeated base structure will spread the color response to a wider angular range than if no randomization is introduced, where grating modes will be visible. This has in particular been studied for the Morpho butterfly, and numerical as well as experimental results can be found in [6–9]. For this paper we assume that the same kind of randomization can be added to the developed designs, thus making them usable for coloring a surface without strong diffraction effects. Performing the actual randomization is beyond the scope of this paper.

4. OPTIMIZATION PROBLEM

The purpose of optimizing the surface structures is to be able to obtain a prescribed angular color response with as strong an intensity as possible. To do so, we have implemented an algorithm for working on the following optimization problem:
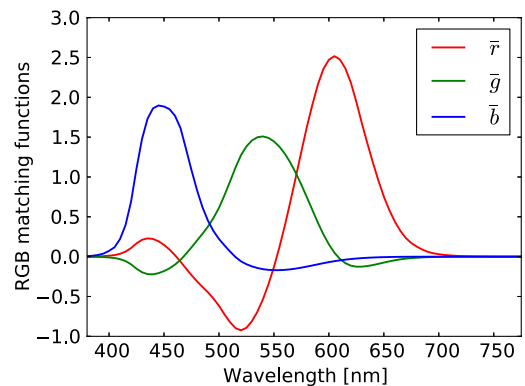

Fig. 4. CIE 1931 2° color-matching functions converted to RGB weighting functions with a D65 illuminant as reference.

$$\min_{\mathbf{h}} \; f(\boldsymbol{h}) = \max_{i=1,2,\ldots,M} \frac{-\|\boldsymbol{C}(\mathbf{h},\theta_i)\|^2}{\|\boldsymbol{C}_0(\theta_i)\|^2} + c_1 \frac{\|\nabla h\|^2}{(Nd)^2},$$

$$\text{s.t.} \;\; \boldsymbol{h}_{\min} \leq \boldsymbol{h} \leq \boldsymbol{h}_{\max},$$

$$f_i(\boldsymbol{h}) = \frac{\|\boldsymbol{C}(\mathbf{h},\theta_i) \times \boldsymbol{C}_0(\theta_i)\|^2}{\varepsilon^2 \|\boldsymbol{C}_0(\theta_i)\|^2} - 1 \leq 0, \tag{20}$$

where f is the objective function, the f_i denotes the M constraint functions, $\|\cdot\|$ is the Euclidean norm, $\boldsymbol{h} = (h_0, h_1, \ldots h_{N-1})$ is the design vector describing the heights of the profile, $\nabla \boldsymbol{h} = (h_1 - h_0, h_2 - h_1, \ldots h_{N-1} - h_{N-2})$ is a measure of the height differences between each element, $\boldsymbol{C}_0(\theta_i)$ is the prescribed color vector we want to obtain at the angle θ_i, M is the number of angles we consider in the optimization problem, $\boldsymbol{h}_{\min}$, $\boldsymbol{h}_{\max}$ are box constraints, and c_1 and ε are variables used to control the optimization problem, and they are described below.

The objective is formulated to maximize the worst-case (with respect to angle) desired color energy, subject to a constraint restricting energy in undesired color directions. This can be seen by considering

$$\|\boldsymbol{C} \times \boldsymbol{C}_0\|^2 = \|\boldsymbol{n} \sin\phi \|\boldsymbol{C}\| \cdot \|\boldsymbol{C}_0\| \|^2$$
$$= \|\boldsymbol{C}\|^2 \cdot \|\boldsymbol{C}_0\|^2 \cdot \sin^2\phi, \tag{21}$$

where ϕ is the angle between the two three-dimensional color vectors. Here it is seen how the squared cross product is proportional to the product of $\sin^2\phi$ and its own length squared. This means that the term approaches zero for vectors in the same direction or vectors with no intensity. Since we seek to maximize the color intensity, the algorithm will in practice try to satisfy the constraint by having a small angle between the vectors and not by keeping the intensity low. The variable ϵ is then used to control how small the angle should be without violating the constraint and has been set to $\epsilon = 0.01$ for all optimizations in this paper. This ensures that we obtain a color close to the prescribed color. The parameter c_1 is used for regularization by weighting the total variation of the design and is discussed in detail in the next section.

Since the optimization problems considered all require a symmetric color response, we have imposed a symmetry condition on the physical design.

The problem has been solved using the method of moving asymptotes [34]. This is a gradient-based optimization algorithm, and the gradient of the objective and the constraints can be calculated analytically using the chain rule, the linear color transform, and Eq. (16). Other nonlinear solvers handling inequality constraints could have been used as well.

A. Regularization Scheme
Various regularization techniques reviewed in [35] have been tried out, but based on experiments we prefer a penalization of the 2-norm of the design gradients, as proposed for other problems in [36]. The implementation of this is what can be seen as the last term in the objective function of Eq. (20). By doing so, a penalty for having huge jumps between neighboring elements is introduced, as well as a penalty for having sharp edges compared to soft edges. This is because the height differences are squared, and hence jumps between 0 and 1 are much more heavily penalized than those between 0 and 0.5—e.g., consider how three elements defined such that

$\boldsymbol{h} = (0, 0, 1)$ will be penalized more than $\boldsymbol{h} = (0, 0.5, 1)$. The obvious goal of removing small features is therefore present, and furthermore the preference for soft curves makes it easier for the optimizer to move elements from one level to another. This is important because we have found that the cost function often does not have monotonic behavior in going from one level to another, causing the (gradient-based) optimizer to maintain its "levels" when they are found. At the same time, no minimum length scale is imposed by this method, leaving the optimization algorithm free to create features of any desired size.

In order to avoid any bias in the results caused by the regularization, a continuation approach is used. In contrast to normal continuation approaches, where an optimization is initialized with a large penalty/filter parameter and then it is gradually decreased, here the design is started with c_1 from Eq. (20) set to zero, and then gradually c_1 is increased when the optimizer approaches a convergent design (for these specific cases, a Karush–Kuhn–Tucker norm smaller than 0.1 has been used as the criterion) to a large value of c_1 (between 30 and 40) and afterward gradually decreased—using the same scheme—until $c_1 = 0$ and the optimizer is then continued until convergence. Due to the color constraint in the formulation of the optimization problem, we retain a design with the desired color, no matter how large we set c_1, and the maximum c_1 value is therefore a parameter defining how many features we are willing to remove. The maximum size of c_1 is found not to be very critical for the overall features and performance of the final design.

B. Setup
To obtain the following results, the setup is as shown in Fig. 5. That is, the optimization is performed for a 2 μm wide surface with light incident along the normal of the surface, which is the only contribution to the far field. The surroundings are considered absorbing, not present or in other ways nonreflecting. The small surface size is chosen to restrict the problem to a size that is easier for the optimization to handle, and because, if a surface is to be colored this way, the most feasible and production-friendly way would seem to be to find a small structure that can then be repeated with added randomness [9]. Another issue is that for larger structures, the incoherence of light would have to be taken into account [8]. The optimization is carried out for 20 equidistant wavelengths between 400 and 700 nm. It is found that fewer wavelengths make it difficult to resolve especially green colors well. The spacing between angles specified for optimization is 1°, and the element spacing, d, is set to 10 nm.
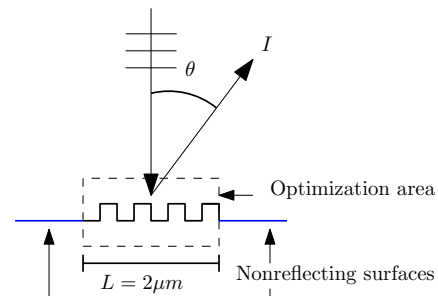

Fig. 5. Physics for the optimization problem.

5. RESULTS

The optimization is performed for two different setups: one for a simple specular reflection, and another for a constant color in an angular range of −10° to 10°. The first setup is used to prove the validity of the procedure by comparing it to analytical calculations—and also to reveal certain features of the optimized designs—whereas the second setup tests its usefulness for designing more complex color effects that cannot be predicted by intuition or simple formulas.

It should be noted that since the level of reflected energy is determined *a priori* by the SDT formulation (e.g., no extra absorption or transmission effects possible, such as surface plasmons), then the colors suppressed in a certain angular interval will be present outside this interval. Roughly speaking, the complementary color to C_0 will be spread out in the rest of the angular domain. This means that a structure scattering blue around the specular direction will have the red and green color components "smeared" out in the rest of the angular domain, as seen in Fig. 6. Furthermore, all color plots in the following take the form seen in Fig. 6, with curves for R, G, and B on a background colored by the combined RGB value of the curves to illustrate the actual color.

A. Designing Specular Reflection

As a benchmark case for the proposed optimization procedure, optimization of the specular reflection for light incident normal to a structure is used. For transmission gratings (by scaling the height, the results can be used for transmission gratings as well), this type of structure is referred to as zero-order diffraction (ZOD) and finds its use in, e.g., color filters.

1. Analytical Solution for the Binary Grating

Using SDT, the expression for zero-order reflection simplifies a great deal, since this corresponds to putting $\alpha = 0$ in Eq. (13):

$$L(0) = \left| \frac{1}{N} \sum_{n=0}^{N-1} e^{-4\pi i h_n / \lambda} \right|^2. \tag{22}$$

Note how the elements lose their dependence on position, since the translation product becomes unity. This is due to the fact that we consider light incident normal to the surface and because of the simple formulation of SDT. In practice the positioning of the elements can be of importance, since the geometry should not violate the basic assumptions of SDT. This means that, e.g., a structure with only one big trench will

be better than one with many small trenches, since the electromagnetic boundary condition on the vertical edges are ignored by SDT and the geometry will contain more high-spatial-frequency content [28].

A particularly simple and relevant case is when we restrict the heights to take one of two values. This is known as a binary phase grating and is easy to produce, e.g., by using etching processes, and is therefore widely used. For this, assume that we allow only two values for h_n, which can be defined such that h_1 is the height difference between the two elements and $h_0 = 0$ [see Fig. 7(c)]; then the response at $\theta_{\text{out}} = 0°$ will be expressed by the following simple analytical expression:

(a)

(b)

(c)
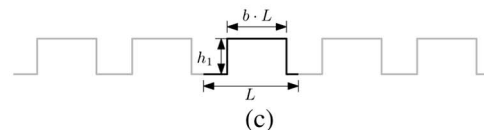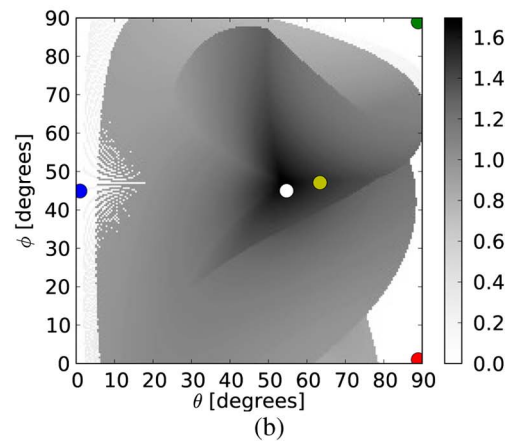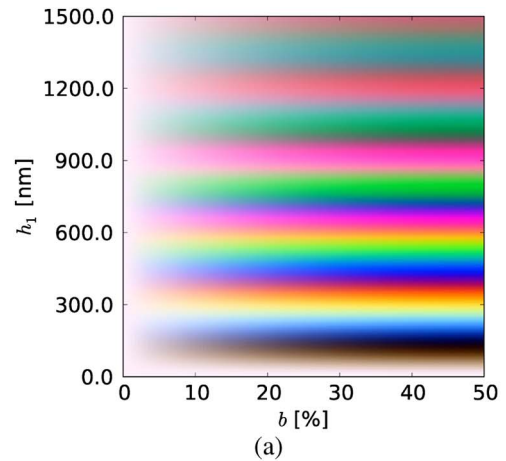
Fig. 7. Analysis of possible colors produceable from the specular mode of a binary phase grating. The maximum height has been limited to 1500 nm. (a) Map of binary phase grating zero-order colors as a function of height and duty cycle (a few RGB values have been cut, as they were negative or exceeded 1) and (b) possible "color directions" for the colors in (a), where the color vector has been transformed into polar coordinates and the gray scale indicates the (largest obtained) value of the length, r, of the color vector. The red, green, and blue points have been indicated, as well as white and black (which overlap, since they only vary in intensity), which are represented by a white circle, and the obtainable test reference is indicated with a yellow circle. (c) Example of how a structure can be realized for certain parameters of h_1 and b.
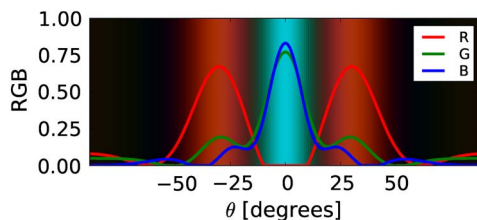

Fig. 6. Reflection for a full reflective surface to illustrate how the color not scattered in the desired directions will be scattered in other directions. In this plot blue and green are scattered close to the specular direction, where red is suppressed, meaning that red will have to be scattered somewhere else.

$$L(b, h_1) = \left| \frac{\sum_{n \in H_0} e^{-4\pi i h_0/\lambda}}{\sum\limits_{n \in H_0} 1} + \frac{\sum_{n \in H_1} e^{-4\pi i h_1/\lambda}}{\sum\limits_{n \in H_1} 1} \right|^2$$

$$= |(1 - b) + b e^{-4\pi i h_1/\lambda}|^2, \qquad (23)$$

where b is the duty cycle (or filling fraction) of h_1. Since whether or not h_1 is defined as a positive or negative quantity with respect to h_0 (making h_0 the bump if h_1 is negative) gives rise to a change only in the far-field phase—and not the amplitude—the function $L(b, h_1)$ for $b \in [0, 1]$ is symmetric around $b = 0.5$, and we therefore need to consider only the interval $b \in [0, 0.5]$. By converting the spectra obtained with different choices of h_1 and b to colors, it is possible to map the space of realizable colors for this setup. The result can be seen in Fig. 7(a). Furthermore, converting the set of color vectors from Cartesian coordinates to spherical coordinates using $r = \sqrt{R^2 + G^2 + B^2}$, $\theta = \arccos B/r$, $\phi = \arctan G/R$, we can plot the possible color directions for binary gratings to obtain an idea of how well we can fill the color space using the specular mode of binary gratings. This mapping procedure makes the color a two-dimensional quantity, which is similar to plotting x and y from a CIE xyY color space. The result is shown in Fig. 7(b). From these plots it can be seen that not all colors are realizable using this approach [as indicated in Fig. 7(b); e.g., pure red, green, and blue cannot be achieved]. To verify the optimization algorithm for more complex designs, a realizable color using a binary grating and red, green, and blue will be the targets. As a more or less random choice, the yellowish color for $h_0 = 275$ nm, $b = 25\%$ was chosen as a reference color for the optimization, which gives rise to $(R, G, B) = (0.89, 0.96, 0.65)$, or $(r, \theta, \phi) = (1.46, 63.4°, 47.1°)$ for specular reflection. In addition to the pure red, green, and blue color, this was used as the optimization goal $\boldsymbol{C}_0$.

2. Detailed Analysis

Going into more detail with the color space, it is possible to find the same yellowish response (generated by a different spectrum) at $(h_0, b) \approx (550$ nm$, 11\%)$, and so it should be noted that two binary solutions exist. If we instead consider the pure red, green, and blue color and try to find the points in the color space matching them best—that is, giving the minimum valid value of ε in Eq. (20)—we find, respectively, that $(h_0, b) \approx (121$ nm$, 50\%)$ with $\varepsilon \approx 0.02$, $(h_0, b) \approx (129$ nm$, 50\%)$ with $\varepsilon \approx 0.10$, and $(h_0, b) \approx (146$ nm$, 50\%)$ with $\varepsilon \approx 8.3 \cdot 10^{-4}$. Looking at Fig. 7(a), the colors look very dark in that area,

which means that the low intensity is helping satisfying the constraint. These solutions are not what the optimization algorithm will aim for, since they have extremely low intensities and will therefore not satisfy the parallel requirement in Eq. (20). If we exclude heights below 200 nm and find the best match for heights above that, we get, respectively, $(h_0, b) \approx (365$ nm$, 50\%)$ with $\varepsilon \approx 0.23$, $(h_0, b) \approx (790$ nm$, 50\%)$ with $\varepsilon \approx 0.16$, and $(h_0, b) \approx (432$ nm$, 50\%)$ with $\varepsilon \approx 0.09$. This study confirms the results in Fig. 7(b), where it is seen that even though a clear blue does not exist in the color space of simple binary gratings, we can find solutions close to it. Furthermore, it is interesting to see that green needs a fairly large difference in height. This can probably be explained by the fact that green is in the middle of the spectrum, and therefore both the lower and upper parts of the spectrum need to be suppressed, whereas red and blue each occupy their part of the spectrum and we just need to suppress the rest.

3. Optimization without Regularization

Running the optimization formulated in Eq. (20) for $\boldsymbol{C}_0 = (0, 0, 1)$ (blue) at $\theta_0 = 0°$, with $\varepsilon = 10^{-2}$, no regularization ($c_1 = 0$), and an initial guess with random heights varying from 250 to 1250 nm, gives the result obtained in Fig. 8. Note that not starting from zero leaves space for the optimization to decrease the heights if necessary. The structure is seen to have very fine one-element features. This is undesirable, as the small features scatter energy in evanescent modes, and paraxial SDT does not take this into account, leading to nonphysical solutions. Besides this, there will be multiple reflections/scattering and internal scattering that are not taken into account, and the result will probably be very different for TE polarization due to the small features. We therefore obtain structures that are hardly realizable when the regularization term is not included. Analyzing the structure using nonparaxial SDT, the color response changes significantly, as seen in Fig. 8, and it is clear that the design will not satisfy our original objective.

To solve all of these problems at once, regularization is used—as discussed in Section 4.A—to improve the quality of design with respect to the length scale.

4. Optimization with Regularization

Using the regularization scheme from Section 3.A and rerunning the optimization, the results presented in Fig. 9 are obtained for blue as well as for the other three prescribed colors. In the figure, $|f|$ indicates the final objective value and N_{it} the number of iterations used. The number of
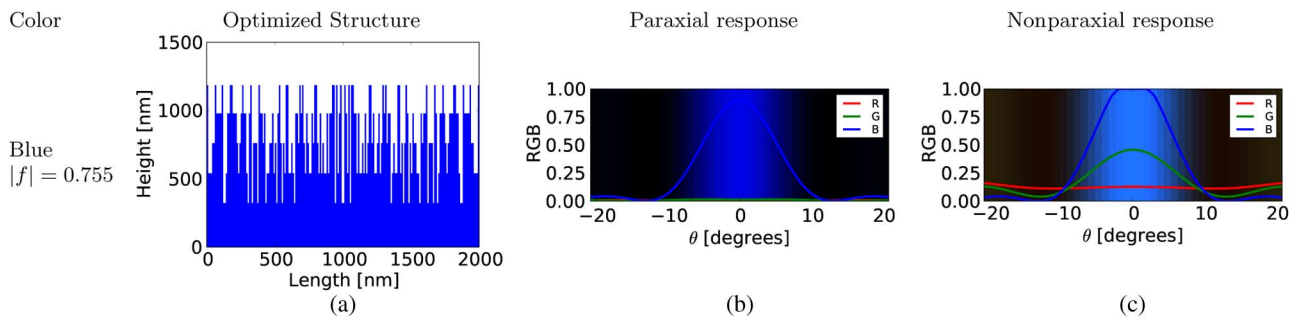

Fig. 8.   Optimization of ZOD for blue color without any regularization. (a) The obtained design, (b) the color response seen by the optimization algorithm, and (c) an analysis of the final structure using nonparaxial SDT. This shows that SDT does not capture the physics well, due to scattering from small feature sizes.
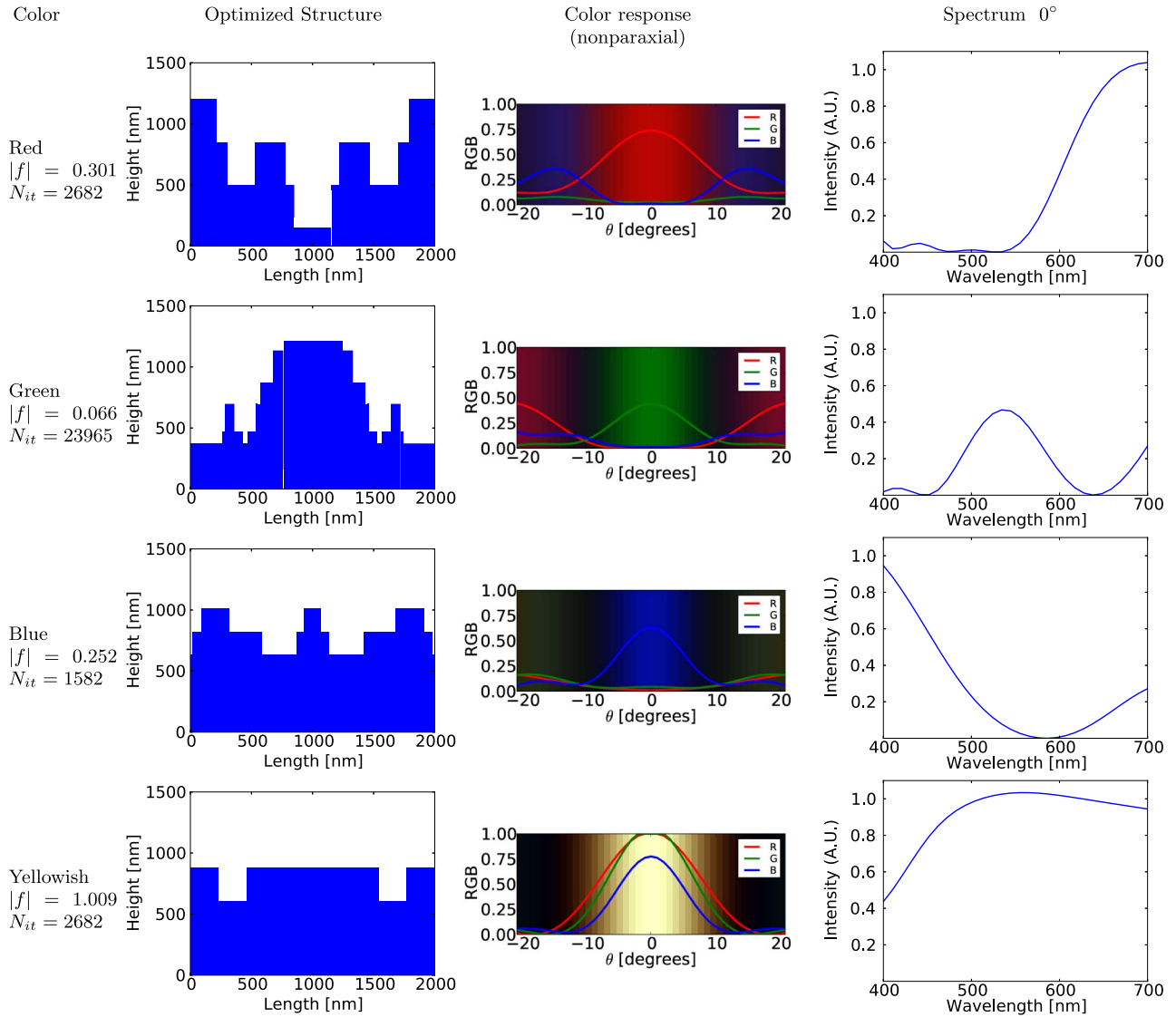
Fig. 9.   Optimization of ZOD using regularization for different colors. All parameters except C_0 and c_1 are unchanged from the previous example. The color coordinates for the prescribed design colors are (1, 0, 0), (0, 1, 0), (0, 0, 1), and (0.89, 0.96, 0.65), respectively.

iterations are for most designs a couple thousand, which is reasonable, since design evaluation is fast. For more computationally heavy optimizations, iterations may be decreased by tuning the regularization parameters. The final results were generated using nonparaxial SDT to show that the results are still satisfactory. A much simpler structure for blue is now obtained at the expense of a decrease in objective value. The structure is divided into three levels, and no single element features are present. From Fig. 7(b) it may be seen how one level would not be enough to generate blue. For red, the same features may be observed, using only four levels. However, for green, it seems that many levels are needed to obtain a good result, but the number of iterations also indicates that the algorithm has had difficulties obtaining a good result. An explanation for this could be that the green color is in the middle of the spectrum, and the grating therefore should act more as a bandpass filter, while red and blue each occupy their end of the visible spectrum, meaning that their behavior resembles more that of a low- or high-pass filter, which is normally easier to realize. For the test color,

it may be seen that the optimizer found a binary solution (without regularization, a multilevel structure with worse performance was found). Its height and duty cycle are $h_1 = 278$ nm, $b = 23\%$, which is close to the reference $h_1 = 275$ nm, $b = 25\%$. The small deviations are probably due to the fact that this solution still satisfies the color constraint but obtains a better objective value.

Running the optimization for the yellowish color with the starting guess being $(h_0, b) \approx (550 \text{ nm}, 11\%)$, which is shown in Section 5.A.2 to also be a valid solution, we end up with a slightly perturbed version of the starting guess and an objective of $|f| = 0.946$, which is slightly lower than what is observed in Fig. 9. This confirms how multiple solutions might exist. In practice several initial guesses have been tried for each color, but have ended up having more or less the same efficiency.

In general the optimization is found to work well, and a wide range of ZOD designs can be created using this method. The designs all have large flat areas divided into levels, even though there is nothing in the formulation preferring that kind of design (either with or without regularization), and from this
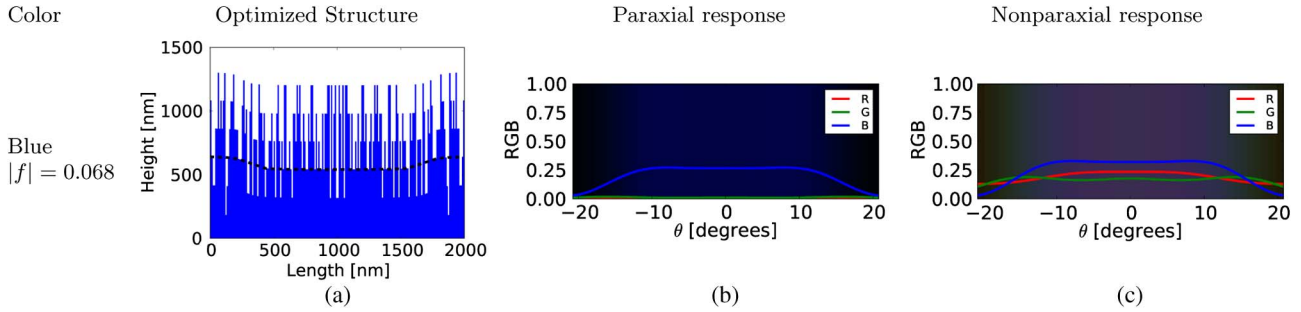
Fig. 10. Optimization for blue for $\theta_{out} \in [-10^{\circ}, 10^{\circ}]$ with no regularization (all other parameters are the same as for the ZOD example). (a) The design obtained, with the dashed line indicating one of the envelopes seen in the structure, (b) the color response seen by the optimization algorithm, and (c) an analysis of the final structure using nonparaxial SDT.

it may be concluded that at least for the colors considered in the examples (and a number of other colors tried by the authors), a few levels are enough to tune a surface to a certain color in the specular mode. It is also worth noting that we have observed that for some colors the regularization method—in addition to yielding a feasible solution—also leads to a better objective value than without regularization.

B. Wide-Angle Color Optimization

Optimizing for wide-angle color response without regularization encounters the same problems as for the ZOD case: performing the same optimization run—now using an angular interval of $\theta_{out} \in [-10^{\circ}, 10^{\circ}]$—yields the result shown in Fig. 10. Again, SDT did not resolve the color very well, and the structure has one-element features, so regularization is again required.

One interesting feature to take note of in Fig. 10 is that it resembles the result in Fig. 8, but with a nonconstant envelope for each level. For the colors where the nonregularized ZOD optimization results in only a few flat levels of height difference, the wide-angle optimization often turns out to be at the same levels, but with a nonconstant envelope. Extracting the

envelope and simulating this alone shows that the envelope shape has the effect of scattering C_0 in a wider angle than a flat surface would. This is illustrated in Fig. 11, where one of the envelopes from Fig. 10 was extracted and a much flatter response may be seen in the blue region compared to a flat surface. This indicates that some designs can be obtained when choosing a color by specifying a height difference, and then finding an appropriate envelope shape of these heights to spread out the color in the specified angular domain.

1. Designs with Regularization

To avoid small design features such as those in Fig. 10, we introduce regularization using the same scheme as in Section 4.A. The results are shown in Fig. 12. What is interesting to note here is that trends similar to the ones from the ZOD optimization can be found: the structure of the test color and blue are the simplest, having only three levels and some slight curvature within more or less the same height difference as for the ZOD result. For green the profile extends to a slightly larger height range but otherwise shows the same trends as blue. For red the optimized structure seems rather simple, but the response is not flat. This means that when the lowest
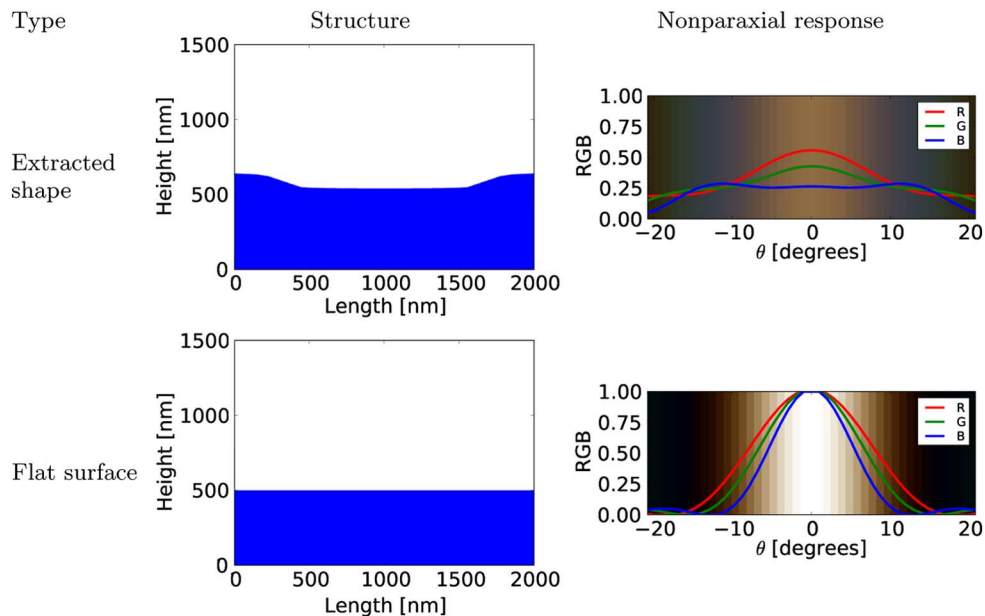

Fig. 11. Comparison of the envelope indicated by a dashed line in Fig. 10(a) and a flat surface having the same length. It is clearly seen how the shape scatters blue in an almost flat interval from -10° to 10° compared to the flat response.
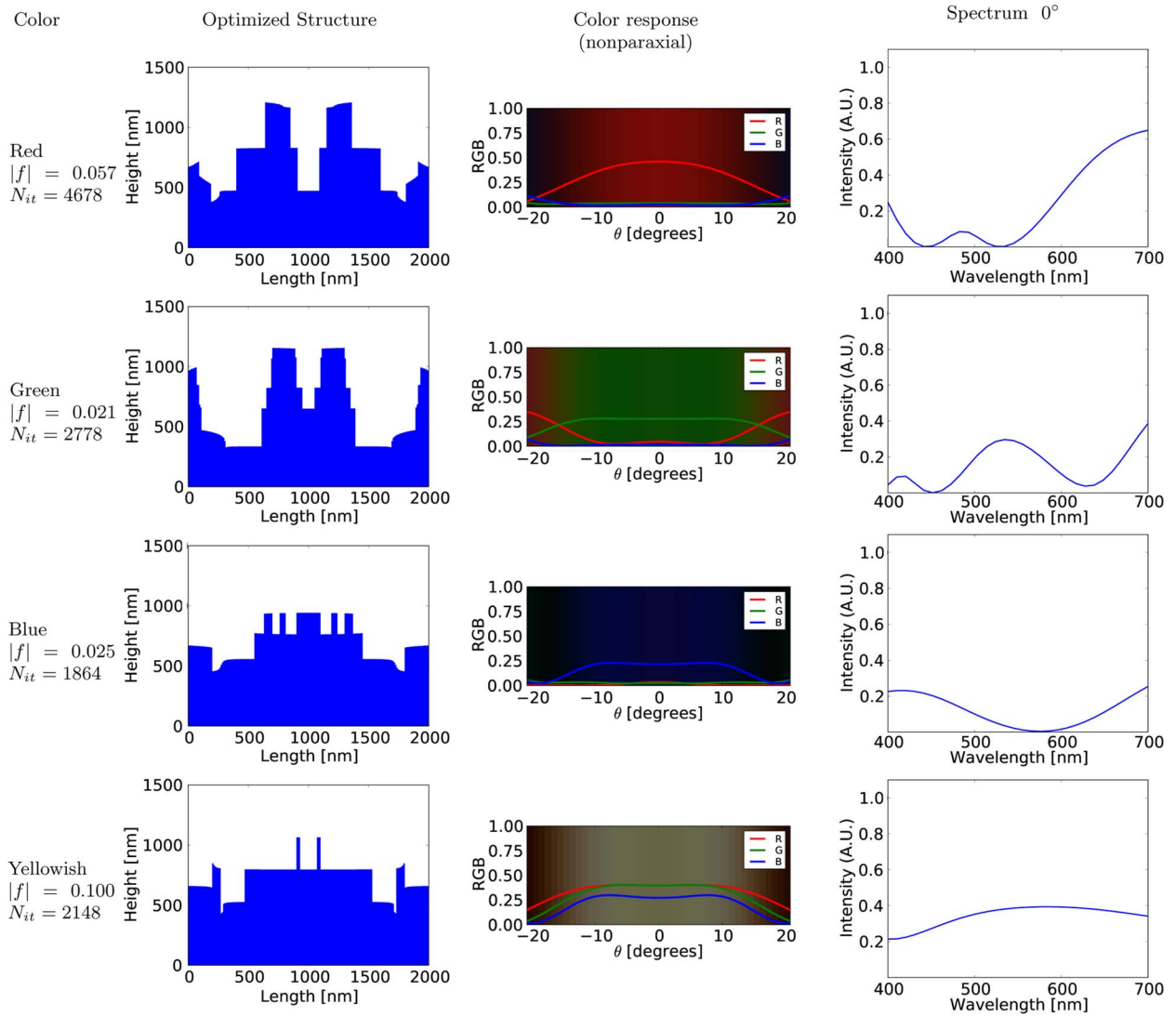
Fig. 12.   Optimization for different colors for $\theta_{out} \in [-10°, 10°]$ using regularization. Except for C_0 and c_1, all the parameters are unchanged from the previous example. The color coordinates for the prescribed design colors are (1, 0, 0), (0, 1, 0), (0, 0, 1), and (0.89, 0.96, 0.65), respectively.

intensities are raised to $\pm 10°$, the other intensities are raised as well. A totally flat response for red was probably not obtained because red light has the longest wavelengths, and thus in general it will scatter less for the 2 µm surface chosen for these experiments. To verify this, we have tried the same procedure with $d = 20$ nm, giving a design domain of 4 µm, and the result obtained here is indeed flat.

In general it seems that the more complex the ZOD design is, the more complex the wide-angle design is, and that overall the procedure seems to be able to find structures satisfying the constraints independently of the choice of color.

6. CONCLUSIONS

In this paper we have shown how to formulate a structural color optimization problem and use it for the optimization of fully reflecting surface profiles to create on-demand angular color effects. The modeling is carried out using SDT discretized and solved in a novel way, but all kinds of electromagnetic methods can be used. The formulation is able to find structures for ZOD responses as well as wide-angle responses. For surface profiles it is observed that not all colors are equally easy to

obtain, and explanations for this have been given. First, colors that can be created by the interference of two different heights/ levels seem to be easier to obtain; second, since the width of the design domain was fixed and it diffracts to wider angles for lower wavelengths, blue colors seem easier to obtain for wide-angle responses; and third, colors that can be obtained by high-/low-pass filtering instead of bandpass/bandstop filtering seem easier to obtain—most likely because these types of filters in general are easier to synthesize. It was also observed that all designs contain sharp edges as a part of the final design, which is a consequence of the fact that strong constructive/destructive interference is needed to filter the colors of the visible spectrum.

As a part of the study, the colors produceable in the zero-order mode by two-level interference were also derived analytically and mapped in Fig. 7(a). To avoid mesh-dependent features in the design, a regularization scheme was introduced that exploits the fact that the color was imposed as a constraint to the optimization problem. This made it possible to obtain design with only the essential features, retaining a good response.

ACKNOWLEDGMENTS

REFERENCES

1. P. Vukusic and J. R. Sambles, "Photonic structures in biology," Nature **424**, 852–855 (2003).
2. S. Kinoshita and S. Yoshioka, "Structural colors in nature: the role of regularity and irregularity in the structure," ChemPhysChem **6**, 1442–1459 (2005).
3. A. Saito, "Material design and structural color inspired by biomimetic approach," Sci. Tech. Adv. Mater. **12**, 064709 (2011).
4. S. Kinoshita, *Structural Colors in the Realm of Nature* (World Scientific, 2008).
5. R. Hooke, Micrographia, 1665, http://www.gutenberg.org.
6. S. Kinoshita, S. Yoshioka, Y. Fujii, and N. Okamoto, "Photophysics of structural color in the Morpho butterflies," Forma **17**, 103–121 (2002).
7. R. T. Lee and G. S. Smith, "Detailed electromagnetic simulation for the structural color of butterfly wings," Appl. Opt. **48**, 4177–4190 (2009).
8. A. Saito, M. Yonezawa, J. Murase, S. Juodkazis, V. Mizeikis, M. Akai-Kasaya, and Y. Kuwahara, "Numerical analysis on the optical role of nano-randomness on the Morpho butterfly's scale," J. Nanosci. Nanotechnol. **11**, 2785–2792 (2011).
9. A. Saito, Y. Miyamura, Y. Ishikawa, J. Murase, M. Akai-Kasaya, and Y. Kuwahara, "Reproduction, mass-production and control of the Morpho-butterfly's blue," Proc. SPIE **7205**, 720506 (2009).
10. Front Page—Nanoplast, http://www.nanoplast.dk/, 2013.
11. O. Karthaus, *Biomimetics in Photonics* (CRC Press, 2012).
12. K. Yu, T. Fan, S. Lou, and D. Zhang, "Biomimetic optical materials: integration of natures design for manipulation of light," Prog. Mater. Sci. **58**, 825–873 (2013).
13. M. J. Uddin and R. Magnusson, "Highly efficient color filter array using resonant Si_3N_4 gratings," Opt. Express **21**, 12495–12506 (2013).
14. X. Sheng, S. G. Johnson, J. Michel, and L. C. Kimerling, "Optimization-based design of surface textures for thin-film Si solar cells," Opt. Express **19**, A841–A850 (2011).
15. K. Kumar, H. Duan, R. S. Hegde, S. C. W. Koh, J. N. Wei, and J. K. W. Yang, "Printing colour at the optical diffraction limit," Nat. Nanotechnol. **7**, 557–561 (2012).
16. Y. R. Wu, A. E. Hollowell, C. Zhang, and L. J. Guo, "Angle-insensitive structural colours based on metallic nanocavities and coloured pixels beyond the diffraction limit," Sci. Rep. **3**, 1194 (2013).
17. G. Park, C. Lee, D. Seo, and H. Song, "Full-color tuning of surface plasmon resonance by compositional variation of Au@Ag core-shell nanocubes with sulfides," Langmuir **28**, 9003–9009 (2012).
18. T. Xu, H. Shi, Y. Wu, A. F. Kaplan, J. G. Ok, and L. J. Guo, "Structural colors: from plasmonic to carbon nanostructures," Small **7**, 3128–3136 (2011).
19. Z. Shi and Y. Gao, "Design of Dammann gratings by particle swarm optimization," in *Symposium on Photonics and Optoelectronics*, June 2010.
20. X. Li, Q. Tan, and G. Jin, "Surface profile optimization of antireflection gratings for solar cells," Optik **122**, 2078–2082 (2011).
21. K. S. Friis and O. Sigmund, "Robust topology design of periodic grating surfaces," J. Opt. Soc. Am. B **29**, 2935–2943 (2012).
22. J. Andkjær, V. E. Johansen, and O. Sigmund, "Inverse design of nano-structured surfaces for color effects," J. Opt. Soc. Am. B **31**, 164–174 (2014).
23. J. S. Jensen and O. Sigmund, "Topology optimization for nanophotonics," Laser Photon. Rev. **5**, 308–321 (2011).
24. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed., McGraw-Hill Physical and Quantum Electronics Series (McGraw-Hill, 1996).
25. J. E. Harvey, "Radiometry rocks," Proc. SPIE **8483**, 848304 (2012).
26. D. C. O'Shea, "Scalar diffraction theory," in *Diffractive Optics: Design, Fabrication, and Test* (SPIE, 2003), pp. 17–35.
27. P. Dutré, K. Bala, and P. Bekaert, *Advanced Global Illumination* (A K Peters/CRC Press, 2006).
28. J. E. Harvey, C. L. Vernold, A. Krywonos, and P. L. Thompson, "Diffracted radiance: a fundamental quantity in nonparaxial scalar diffraction theory," Appl. Opt. **38**, 6469–6481 (1999).
29. J. E. Harvey, C. L. Vernold, A. Krywonos, and P. L. Thompson, "Diffracted radiance: a fundamental quantity in nonparaxial scalar diffraction theory: errata," Appl. Opt. **39**, 6374–6375 (2000).
30. A. W. Lohmann, "About the philosophies of diffraction," in *International Trends in Optics*, J. W. Goodman, ed. (Academic, 1991), Chap. 11, pp. 155–164.
31. R. S. Berns, F. W. Billmeyer, and M. Saltzman, *Billmeyer and Saltzman's Principles of Color Technology* (Wiley-Interscience, 2000).
32. CIE: International Commission on Illumination, http://www.cie.co.at/.
33. International Electrotechnical Commission, "Colour management—default RGB colour space—sRGB," IEC 61966-2-1, 1999.
34. K. Svanberg, "The method of moving asymptotes—a new method for structural optimization," Internat. J. Numer. Methods Engrg. **24**, 359–373 (1987).
35. O. Sigmund, "Morphology-based black and white filters for topology optimization," Struct. Multidiscip. Optim. **33**, 401–424 (2007).
36. T. Borrvall, "Topology optimization of elastic continua using restriction," Arch. Comput. Methods Eng. **8**, 351–385 (2001).