Contents

Chapter 1
Objective-First Nanophotonic Design

Abstract The abstract for the book.

In this chapter, we introduce an "objective-first" strategy for designing nanophotonic devices.

General description:

- Only uses desired fields.
- Can attempt to design any linear device.
- Allows for non-physical fields.

This strategy is unique in that it asks the user only for what electromagnetic field the device should produce, and then attempts to generate the design without further user intervention.

1.1 The electromagnetic wave equation

1.1.1 Physics formulation

First, let's derive our wave equation, starting with the differential form of Maxwell's equations,

$$\nabla \times E = -\mu_0 \frac{\partial H}{\partial t} \tag{1.1}$$

$$\nabla \times H = J + \varepsilon \frac{\partial E}{\partial t}, \tag{1.2}$$

where E, H, and J are the electric, magnetic and electric current vector fields, respectively, ε is the permittivity and μ_0 is the permeability, which we assume to be that of vacuum everywhere.

For time dependence $\exp(-i\omega t)$, where ω is the angular frequency, these become

$$\nabla \times E = -i\mu_0 \omega H \tag{1.3}$$

$$\nabla \times H = J + i\varepsilon \omega E, \tag{1.4}$$

which we can combine to form our wave equation,

$$\nabla \times \varepsilon^{-1} \nabla \times H - \mu_0 \omega^2 H = \nabla \times \varepsilon^{-1} J. \tag{1.5}$$

For further information, as well as simplifications to the wave equation in reduced dimensions, please see the Appendix.

1.1.2 Numerical formulation

Now, on top of the analytical formulation of the wave equation (1.5) we will now add a numerical, or discretized, formulation. This will be needed in order to solve for arbitrary structures.

First, we discretize our computational space according to the Yee grid, which allows us to easily define the curl ($\nabla\times$) operators in (1.5) as described in the Appendix. This allows us, with a change of variables to formulate (1.5) as

$$A(p)x = b(p), \tag{1.6}$$

where $H \rightarrow x$, $\varepsilon^{-1} \rightarrow p$; and where

$$A(p) = \nabla \times \varepsilon^{-1} \nabla \times -\mu_0 \omega^2 \tag{1.7}$$

and

$$b(p) = \nabla \times \varepsilon^{-1} J. \tag{1.8}$$

Note that our use of $A(p)$ and $b(p)$ instead of A and b simply serves to clarify the dependence of both A and b to p.

Additionally, we use periodic boundary conditions with stretched-coordinate perfectly matched layers where necessary for our examples.

1.1.3 Solving for H

With our numerical formulation, we can now solve for the H-field (the E-field can be computed from the H-field using (1.4)) by using general linear algebra solvers. Doing so is also simply known as a time-harmonic or a finite-difference frequency-domain (FDFD) simulation.

Now, while a full three-dimensional problem is computationally quite taxing; in one- and two-dimensions, (1.6) is easily solved using the standard sparse solver

included in Matlab, and this technique is regularly employed in the examples which follow.

1.1.4 Solving for ε^{-1}

The next step, after having built a field-solver or simulator (finds x given p) for our wave equation, is to build a structure-solver for it. In other words, we need to be able to solve for p given x.

To do so, we return to (1.5) and remark that $\varepsilon^{-1}(\nabla \times H) = (\nabla \times H)\varepsilon^{-1}$ and $\varepsilon^{-1}J = J\varepsilon^{-1}$ since scalar multiplication is communicative. This allows us to rearrange (1.5) as

$$\nabla \times (\nabla \times H)\varepsilon^{-1} - \nabla \times J\varepsilon^{-1} = \mu_0 \omega^2 H \tag{1.9}$$

which we now write as

$$B(x)p = d(x), \tag{1.10}$$

where

$$B(x) = \nabla \times (\nabla \times H) - \nabla \times J \tag{1.11}$$

and

$$d(x) = \mu_0 \omega^2 H. \tag{1.12}$$

Solving this system would now seem to allow us to choose an electromagnetic field and then find the structure to produce it; which strongly suggests that it will be useful in the design of nanophotonic devices.

In terms of computational complexity, as with (1.6), (1.10) in its current form can be solved using standard tools.

1.1.5 Bi-linearity of the wave equation

Although additional mathematical machinery must still be added in order to get a useful design tool, we have really shown so far is that the wave equation is separately linear or *bi-linear* in x and p. Namely that,

$$A(p)x - b(p) = B(x)p - d(x). \tag{1.13}$$

In other words, fixing p makes solving the wave equation for x a linear problem, and vice versa. Note that the joint problem, where both x and p are allowed to vary, is not linear.

The bi-linearity of the wave equation is fundamental in our objective-first strategy which relies on the fact that we already know how to solve linear systems well, and is the reason why we chose $\varepsilon^{-1} \to p$ instead of the more natual $\varepsilon \to p$. Indeed,

this property forms a natural division of labor in the objective-first scheme, which we outline below.

1.2 The objective-first design problem

We now build off of the field-solver and the structure-solver, as previously outlined, by formulating the design problem and outlining the objective-first strategy.

1.2.1 Design objectives

A design objective, $f(x)$, is simply defined as a function we wish to be minimal for the design to be produced.

For instance, in the design of a device which must transmit efficiently into a particular mode, we could choose $f(x)$ to be the negative power transmitted into that mode. Or, if the device was to be a low-loss resonator, we could choose $f(x)$ to be the amount of power leaking out of the device.

In general, there are multiple choices of $f(x)$ which can be used to describe the same objective. For example, $f(x)$ for a transmissive device may not only be the negative power transmitted into the desired output mode, but it could also be the amount of power lost to other modes, or even the error in the field values at the output port relative to the field values needed for perfect transmission. These design objectives are equivalent in the sense that, if minimized, all would produce structures with good performance. At the same time, we must consider that the computational cost and complexity of using one $f(x)$ over another may indeed vary greatly.

1.2.2 Convexity

Before formulating the design problem, we would like to inject a note regarding the complexity of various optimization problems.

Specifically, we want to introduce the notion of *convexity* and to simply note the difference between problems that are convex and those which are not. The difference is simply this: convex problems have a single optimum point (only one local optimum, which is therefore the global optimum) which we can reliably find using existing numerical software, whereas non-convex problems typically have multiple optima and are thus much more difficult to reliably solve.

That a convex problem can be reliably solved, in this case, means that regardless of the starting guess, convex optimization software will always arrive at the globally optimal solution and will be able to numerically prove global optimality as

well. Thus, formulating a design problem in terms of convex optimization problems virtually eliminates any ideas of chance or randomness.

For the examples presented in this chapter, we use CVX, a convex optimization software written for Matlab.

1.2.3 Typical design formulation

The typical, and most straightforward formulation of the design problem is

$$\underset{x,p}{\text{minimize}} \quad f(x) \tag{1.14}$$

$$\text{subject to} \quad A(p)x - b(p) = 0, \tag{1.15}$$

which states that we would like to vary x and p simultaneously in order to decrease $f(x)$ while always satisfying physics (the electromagnetic wave equation). Such a formulation, if solved using a steepest-descent method, is the well-known adjoint optimization method.

1.2.4 Objective-first design formulation

In contrast, the objective-first formulation switches the roles of the wave equation and the design objective with one another,

$$\underset{x,p}{\text{minimize}} \quad \|A(p)x - b(p)\|^2 \tag{1.16}$$

$$\text{subject to} \quad f(x) = f_{\text{ideal}}. \tag{1.17}$$

This first means that, as seen from (1.16), we allow for non-zero residual in the electromagnetic wave equation. This literally means that we allow for *non-physical* x and p, since $A(p)x - b(p) \neq 0$ is now allowed. For this reason, we denote $A(p)x - b(p)$ the *physics residual*.

Secondly, we see from (1.17) that we always force the device to exhibit ideal performance, even if doing so means breaking the laws of physics. As such, our strategy will be to vary x and p in order to decrease the physics residual (1.16) to zero, while always maintaining ideal performance.

Our initial motivation for doing this was that it allowed us to solve for x and p separately, as will be outlined below, and that always forcing ideal performance might provide a mechanism to "override" local optima in the optimization process.

To this end we have found that such a strategy actually allows us to design very unintuitive devices which exhibit very good performance, even when starting from completely non-functional initial guesses. Furthermore, we have found this to be true even true when the physics residual is never brought to exactly zero.

From a numerical standpoint, although the objective-first formulation is still non-convex in its original form, the bi-linearity of the physics residual term allows us to naturally break the original problem into two sub-problems which we outline below.

Lastly, we add an additional constraint to the original formulation, which is to set hard-limits on the allowable values of p, namely $p_0 \leq p \leq p_1$. This is actually a relaxation of the ideal constraint, which would be to allow p to only have discrete values, $p \in p_0, p_1$, but such a constraint would be essentially force us to only be able to perform brute force trial-and-error.

Our objective-first formulation is thus,

$$\begin{aligned} \underset{x,p}{\text{minimize}} \quad & \|A(p)x - b(p)\|^2 \\ \text{subject to} \quad & f(x) = f_{\text{ideal}} \\ & p_0 \leq p \leq p_1. \end{aligned} \tag{1.18}$$

1.2.5 Field sub-problem

Since the objective-first problem in its original form is still non-convex, we break it down into two convex sub-problems. The first of these is the field sub-problem, which simply involves fixing p and independently optimizing x,

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|A(p)x - b(p)\|^2 \\ \text{subject to} \quad & f(x) = f_{\text{ideal}}. \end{aligned} \tag{1.19}$$

This problem is convex, and actually quadratic, which means that it can even be solved in the same way as a simple least-squares problem.

1.2.6 Structure sub-problem

The second sub-problem is formulated by fixing x and independently optimizing p. At the same time, we use the bi-linearity property of the physics residual from (1.13) to rewrite the problem in a way that makes its convexity explicit,

$$\begin{aligned} \underset{p}{\text{minimize}} \quad & \|B(x)p - d(x)\|^2 \\ \text{subject to} \quad & p_0 \leq p \leq p_1. \end{aligned} \tag{1.20}$$

The structure sub-problem is also convex, but not quadratic. However, use of the CVX package still allows us to obtain the result quickly and reliably.

1.2.7 Alternating directions

We use a simple alternating directions scheme to peice together (1.19) and (1.20), which is to say that we simply alternately solve each and continue until we reach some stopping point, normally measured by how much the physics residual has decreased.

The advantage of such the alternating directions method is that the physics residual is guaranteed to monotonically decrease with every iteration, which is useful in that no safeguards are needed to guard against "rogue" steps in the optimization procedure. Note that this robustness stems from the fact that, among other things, each sub-problem does not rely on previous values of the variable which is being optimized, but only on the variable which is held constant.

The disadvantage of such a simple scheme is that the convergence is quite slow, although we have found it to be sufficient in our cases. See ref for related methods that exhibit far better convergence.

1.3 Waveguide coupler design

We first apply the objective-first formulation with the alternating directions algorithm to the design of nanophotonic waveguide couplers in two dimensions, where our goal is to couple light from a single input waveguide mode to a single output waveguide mode with as close to unity efficiency as possible. We would also like to allow the user to choose arbitrary input and output waveguides, as well as to select arbitrary modes within those waveguides (as opposed to allowing only the fundamental mode, for example).

This problem is very general and, in essence, encompasses the design of all linear nanophotonic components, because the function or performance of all such components is simply to convert a defined set of input modes into a defined set of output modes. Such a broad, general problem is ideally suited for an objective-first strategy, since no approximations or simplifications of the electromagnetic fields are required; we only make the simplification of working in two dimensions and dealing only with a single input and output mode.

Say something about TE mode in 2D.

1.3.1 Choice of design objective

As mentioned in section 1.2.1 multiple equivalent choices of design objective exist which should allow one to achieve the same device performance; however, we will choose, for generality, the following design objective,

$$f(x) = \begin{cases} x & \text{at boundary,} \\ 0 & \text{elsewhere,} \end{cases} \tag{1.21}$$

That is, $f(x)$ simply selects the outermost values of the field in the design space.

When placed into the objective-first problem (1.18), this will result in fixing the boundary values of the field at the edge of the design space to those of an ideal device, as shown in figure 1.1. In this case, we choose such an ideal device to have perfect (unity) coupling efficiency,

$$f_{\text{ideal}} = \begin{cases} x_{\text{perfect}} & \text{at boundary,} \\ 0 & \text{elsewhere.} \end{cases} \tag{1.22}$$

These ideal fields are simply obtained by using the input and output mode profiles at the corresponding ports and using values of zero at the remaining ports.

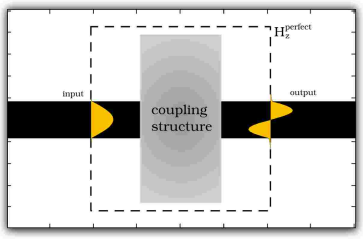

Fig. 1.1 Formulation of the design objective.

Such a design objective is general in the sense that the boundary values of the device contain all the information necessary to determine how the device will interact with its environment, when excited with the input mode in question. In other words, we only need to know the boundary field values, and not the interior field values to determine the performance of the device; and thus, it would be conceivable that such a scheme might be generally applied to linear nanophotonic devices beyond just waveguide mode couplers.

In our case, we only need to know the value of H_z and its derivative along the normal direction, $\partial H_z/\partial n$, along the design boundary in order to completely characterize its performance. Alternatively, one can, of course, use the outermost two layers of the H_z instead of calculating a spatial derivative.

1.3.2 Application of the objective-first strategy

Having chosen our design objective we apply alternating directions to (1.18) which results in solving the following two sub-problems iteratively.

$$\begin{aligned} &\underset{x}{\text{minimize}} && \|A(p)x - b(p)\|^2 \\ &\text{subject to} && x = x_{\text{perfect}}, \text{ at boundary} \end{aligned} \tag{1.23}$$

$$\begin{aligned} &\underset{p}{\text{minimize}} && \|B(x)p - d(x)\|^2 \\ &\text{subject to} && p_0 \leq p \leq p_1 \end{aligned} \tag{1.24}$$

For the results throughout this chapter, we uniformly choose $p_0 = 1/12.25$ and $p_0 = 1$, corresponding to ε^{-1} of silicon and air respectively. Additionally, since a starting value for p is initially required, we always choose to use a uniform value of $p = 1/9$ across the entire design space. There is nothing really unique about such a choice, although we have noticed that initial value of p near 1 often result in poor designs. Note, that, unlike p, we do not require an initial guess for x.

The only other significant value that needs to be set initially is the frequency, or wavelength of light. We use wavelengths in the range of 25 to 42 grid points for the results in this chapter.

Lastly, for all the examples presented in the chapter, we run the alternating directions algorithm for 400 iterations. Although we do not present the convergence results here, such information can be obtained by inspecting the source.

Also, small footprint and computation time need to be mentioned.

1.3.3 Coupling to a wide, low-index waveguide

As a first example, we design a coupler from the fundamental mode of a narrow, high-index waveguide to the fundamental mode of a wide, low-index waveguide. Such a coupler would be useful for coupling from an on-chip nanophotonic waveguide to an off-chip fiber for example.

The input and output mode profiles used as the ideal fields are shown in the upper-left corner of figure 1.2. The final structure is shown in the upper right plot, and the simulated H_z fields, under excitation of the input mode, are shown in the bottom plots.

Figure 1.2 then shows that the design structure has nearly unity efficiency and converts between the input and output modes within a very small footprint.

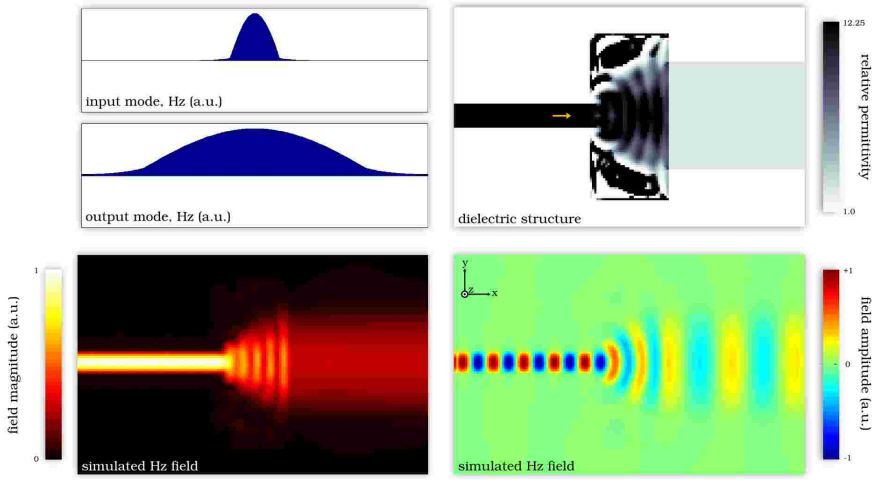

Fig. 1.2 Coupler to a wide low-index waveguide. Efficiency: 99.8%, footprint: 36×76 grid points, wavelength: 42 grid points.

1.3.4 Mode converter

In addition to coupling to a low-index waveguide, we show that we can successfully apply the objective-first method to convert between modes of a waveguide. We do this by simply selecting the output mode in the design objective to be the second-order waveguide mode, as seen in figure 1.3.

Note that the design of this coupler is made challenging because of the opposite symmetries of the input and output modes. Moreover, because our initial structure is symmetric, we initially have exactly 0% efficiency to begin with. Fortunately, the objective-first method can still design an efficient coupler in this case as well.

1.3.5 Coupling to an air-core waveguide mode

We can then continue to elucidate the generality of our method by coupling between waveguides which confine light in completely different ways.

Figure 1.4 shows a high-efficiency coupling device between an index-guided input waveguide and a "air-core" output waveguide, in which the waveguiding effect is achieved using distributed Bragg reflection.

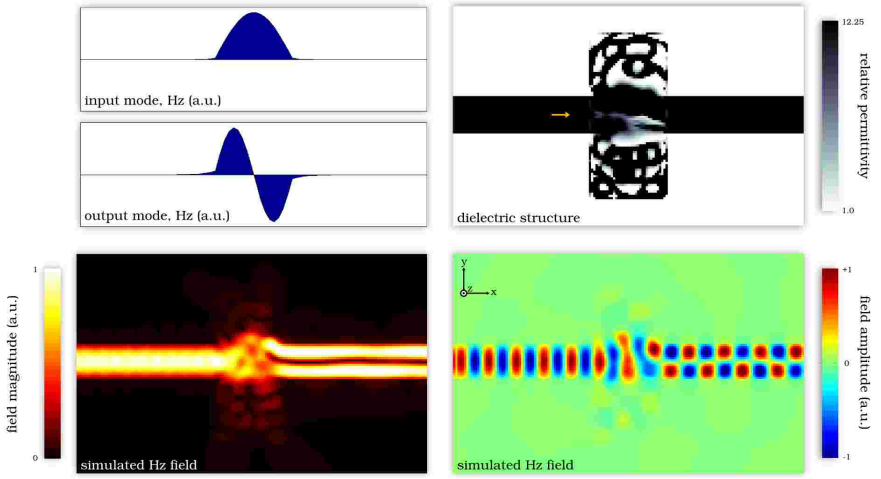

Fig. 1.3 Mode converter. Efficiency: 98.0%, footprint: 36×76 grid points, wavelength: 42 grid points.

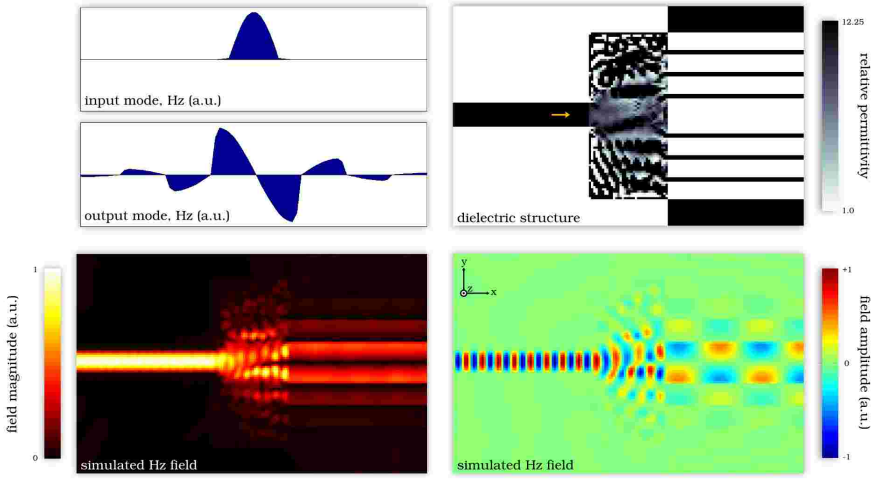

Fig. 1.4 Coupler to a wide low-index waveguide. Efficiency: 98.9%, footprint: 36×76 grid points, wavelength: 25 grid points.

1.3.6 Coupling to a metal-insulator-metal waveguide

Additionally, our design method can also generate couplers between different material systems such as between dielectric and metallic (plasmonic) waveguides, as shown in figure 1.5.

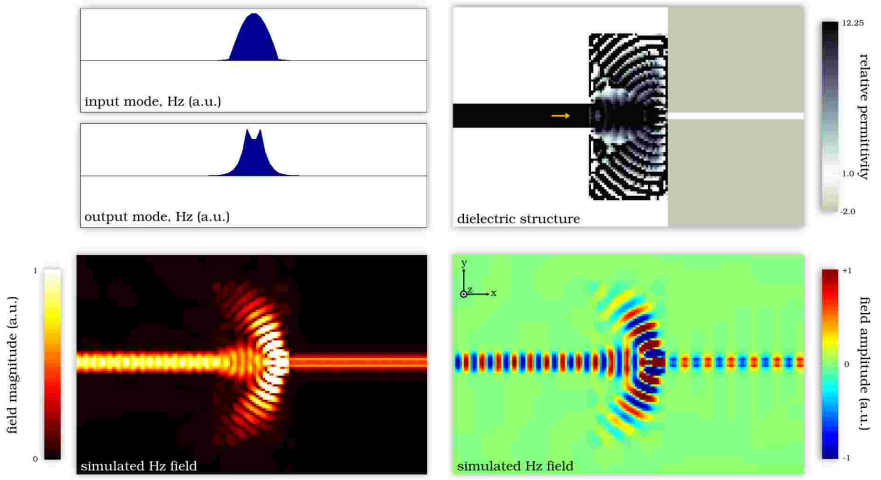

Fig. 1.5 Coupler to a wide low-index waveguide. Efficiency: 97.5%, footprint: 36×76 grid points, wavelength: 25 grid points.

In this case, the permittivity of the metal ($\varepsilon = -2$) is chosen to be near the plasmonic resonance ($\varepsilon = -1$).

1.3.7 Coupling to a metal wire plasmonic waveguide mode

Lastly, figure 1.6 shows that efficiently coupling to a plasmonic wire is achievable as well.

1.4 Optical cloak design

In the previous section, we showed that couplers between virtually any two waveguide modes could be constructed using the objective-first design method, and based on the generality of the method one can guess that it may also be able to generate designs for any linear nanophotonic device.

Now, we extend the applicability of our method to the design of metamaterial devices which operate in free-space. In particular, we adapt the waveguide coupler algorithm to the to the design of optical cloaks.

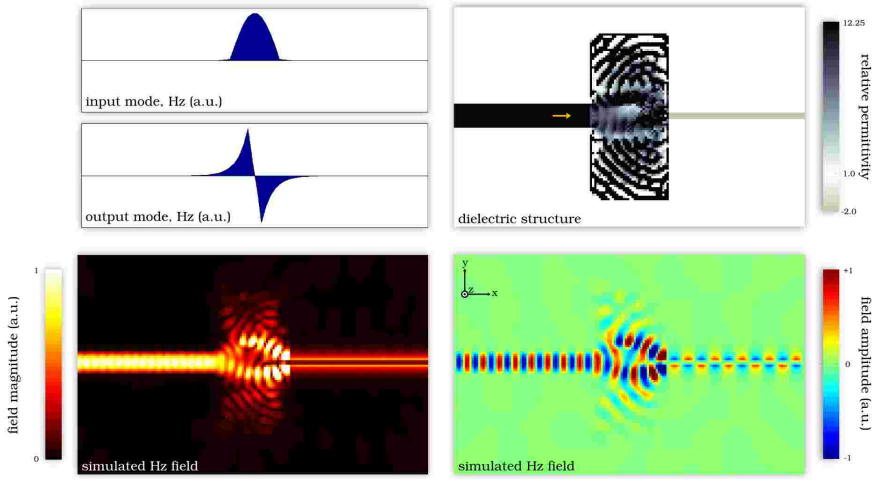

Fig. 1.6 Coupler to a wide low-index waveguide. Efficiency: 99.1%, footprint: 36×76 grid points, wavelength: 25 grid points.

1.4.1 Application of the objective-first strategy

Adapting the method used in section 1.3 to the design of optical cloaks really only requires one to change the simulation environment to allow for free-space modes. This is accomplished by modifying the upper and lower boundaries of the simulation domain from absorbing boundary conditions to periodic boundary conditions, which allows for plane-wave modes to propagate without loss until reaching the left or right boundaries, where absorbing boundary conditions are still maintained.

In terms of the design objective, we allow the device to span the entire height of the simulation domain, and thus consider only the leftmost and rightmost planes as boundary values. Specifically, for this section the input and output modes are plane waves with normal incidence, as can be expected for good cloaking devices. The achieved results all yield high efficiency, although we note that the cloaking effect is only measured for a specific input mode. That is to say, just as the waveguide couplers previously designed were single-mode devices, so the cloaks designed in this section are also "single-mode" cloaks.

An additional modification, as compared to section 1.3, is that we now disallow the structure to be modified in certain areas which, naturally, contain the object to be cloaked.

With these simple changes we continue to solve (1.18) with the alternating directions method in order to now design optical cloaks instead of waveguide couplers. Once again, as in section 1.3, each design is run for 400 iterations with a uniform initial value of $p = 1/9$ for the structure (where the structure is allowed to vary), and the range of p is limited to $1/12.25 \leq p \leq 1$.

1.4.2 Anti-reflection coating

As a first example, we attempt to design the simplest and most elementary "cloaking" device available, which, we argue, is a simple anti-reflection coating; in which case the object to be cloaked is nothing more than the interface between two dielectric materials. In this case we use the interface between air and silicon, as shown in figure 1.7

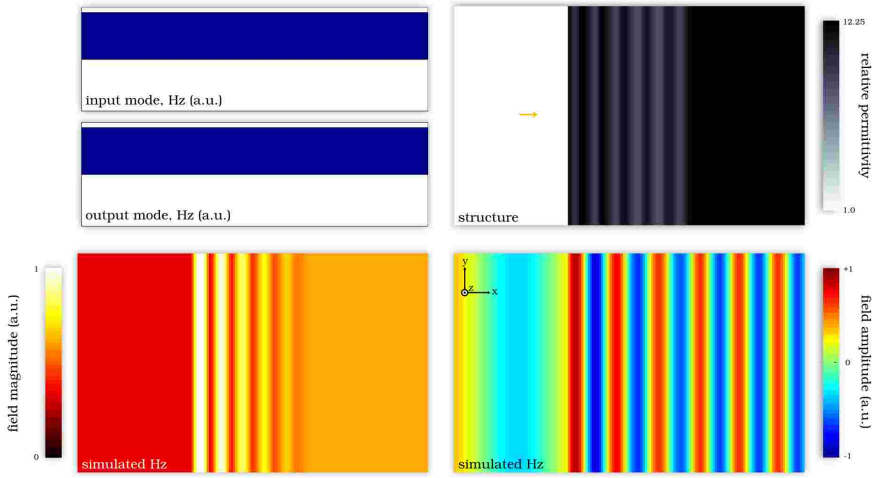

Fig. 1.7 Anti-reflection coating. Efficiency: 99.99%, footprint: 60×100 grid points, wavelength: 63 grid points.

Unsurprisingly for such a simple case, we achieve a very high efficiency device. Note also that the efficiency of the device can be deduced by eye, based on the absence of reflections or standing waves in bottom two plots of figure 1.7.

1.4.3 Wrap-around cloak

Next, we design a cloak for a plasmonic cylinder, which is quite effective at scattering light as can be seen from figure 1.8.

In designing the wrap-around cloak, we allow the structure to vary at all points within the design area except in the immediate vicinity of the plasmonic cylinder. Application of the objective-first strategy results in an efficient device as seen in figure 1.9.

Note that our cloak employs only isotropic, non-magnetic materials, and at the same time it is specific to a particular input and to a particular object.

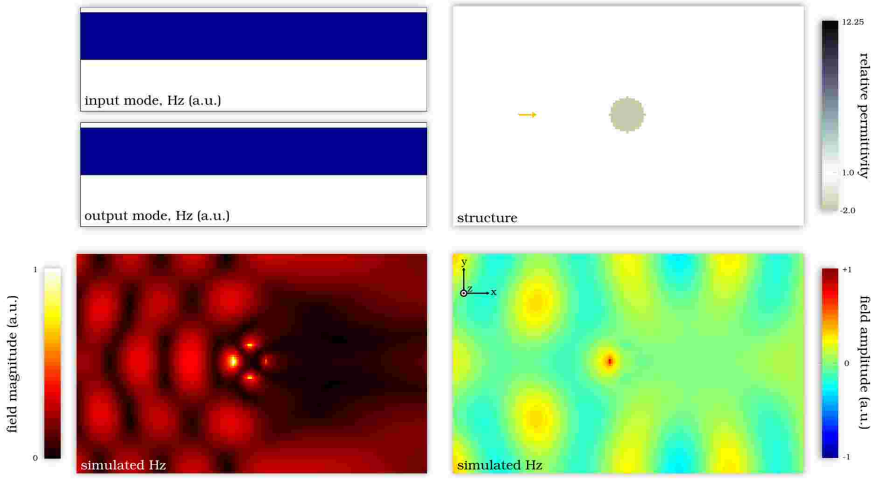

Fig. 1.8 Plasmonic cylinder to be cloaked. 68.5% of light is diverted away from the desired output mode.

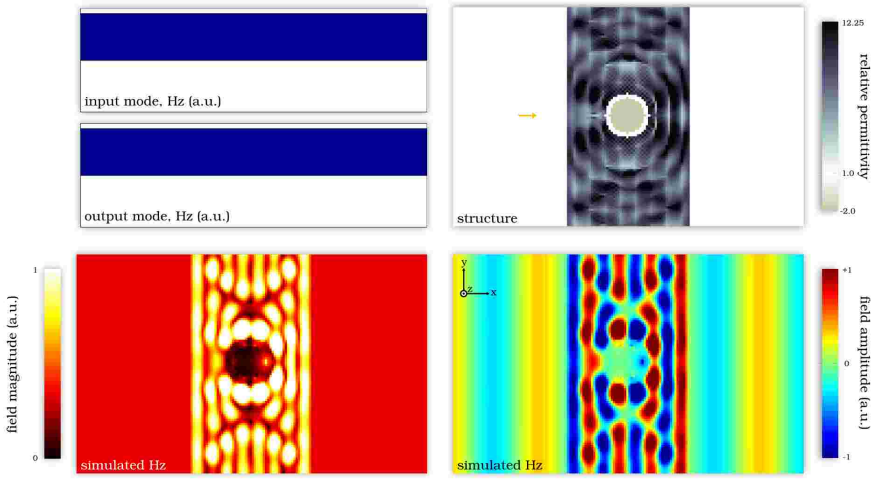

Fig. 1.9 Wrap-around cloak. Efficiency: 99.99%, footprint: 60×100 grid points, wavelength: 42 grid points.

1.4.4 Open-channel cloak

With a simple modification, from the previous section, we can design a cloak which features an open channel to the exterior electromagnetic environment. This simple

modification is forcing an air tunnel to be opened which connects the cylinder to the outside world both toward its front and back.

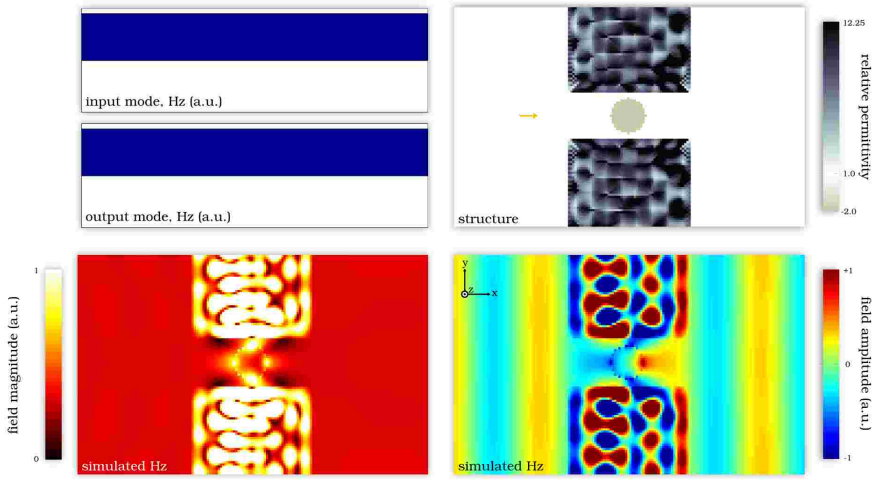

Fig. 1.10 Open-channel cloak. Efficiency: 99.8%, footprint: 60×100 grid points, wavelength: 42 grid points.

Such a design is still very efficient and exhibits the usefulness of the objective-first strategy in cases where other methods, such as transformation optics, may not be able to be applied.

1.4.5 Channeling cloak

Our last cloaking example replaces the plasmonic cylinder with a thin metallic wall in which is etched a sub-wavelength channel. Such a metallic wall is very effective at blocking incoming light (as can be seen from figure 1.11) because of its large negative permittivity ($\varepsilon = -20$), meaning that any cloaking device would be forced to channel all the input light into a very small aperture and then to flatten that light out into a plane wave again.

Once again, our method is still able to produce a very efficient design, as shown in figure 1.12.

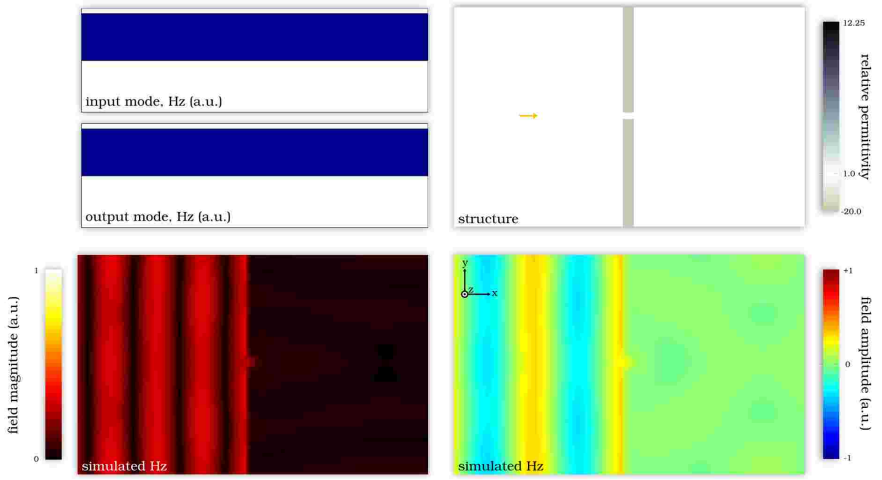

Fig. 1.11 Metallic wall with sub-wavelength channel to be cloaked. 99.9% of the light is blocked from the desired output plane-wave.

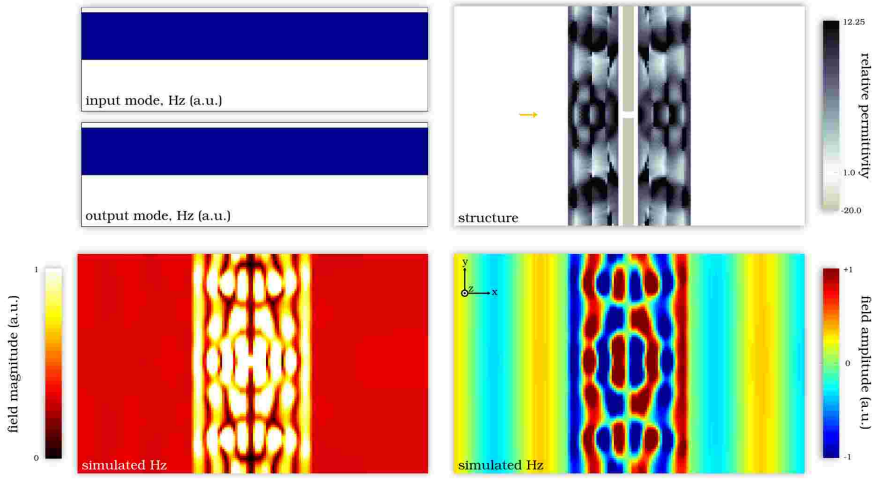

Fig. 1.12 Channeling cloak. Efficiency: 99.9%, footprint: 60×100 grid points, wavelength: 42 grid points.

1.5 Optical mimic design

1.6 Extending the method

1.6.1 3D

1.6.2 Multi-mode

1.6.3 Robustness

1.6.4 Binary structure

1.7 Appendix

1.7.1 Full 3D curl

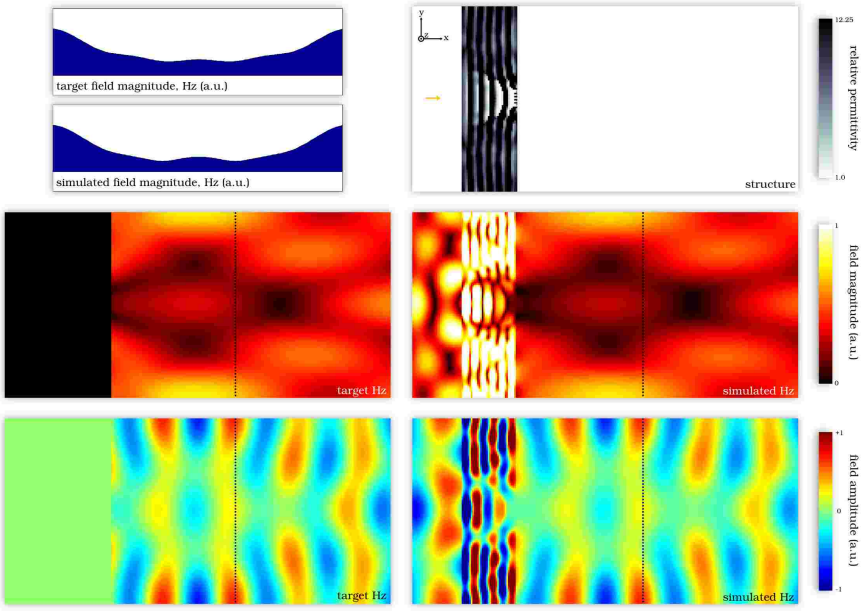

Fig. 1.13 test

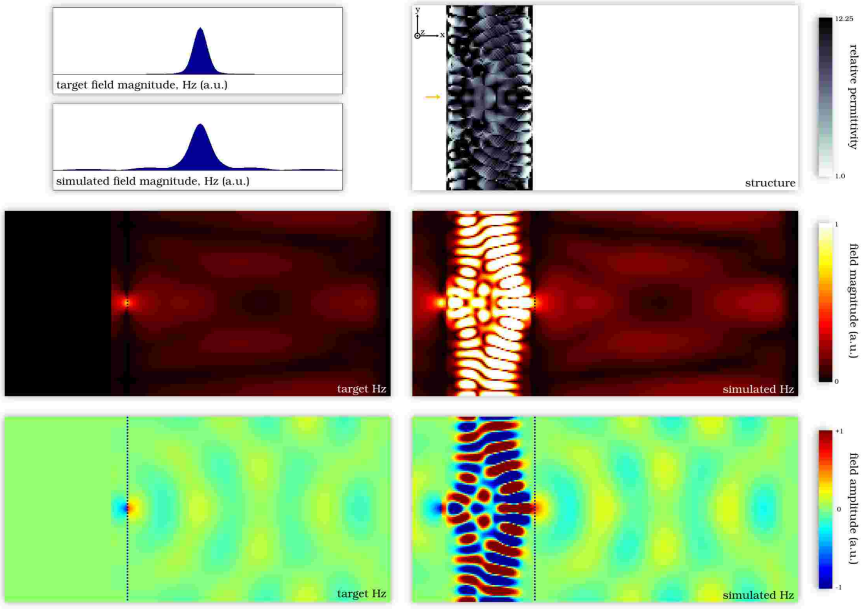

Fig. 1.14 test

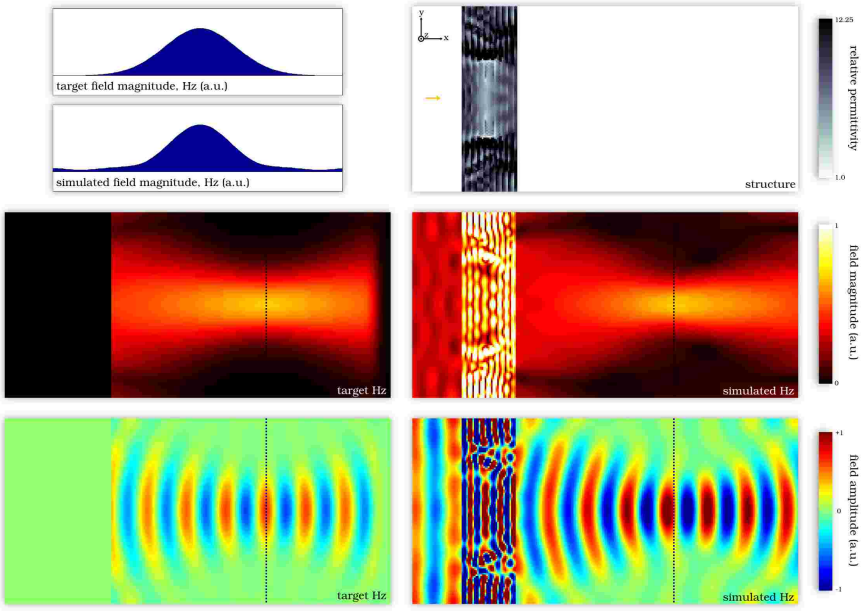

Fig. 1.15 test

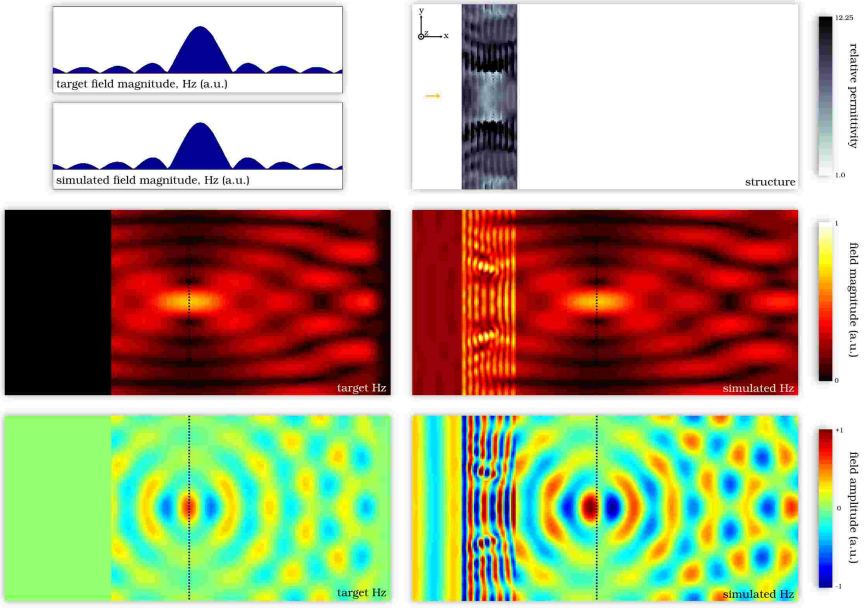

Fig. 1.16 test

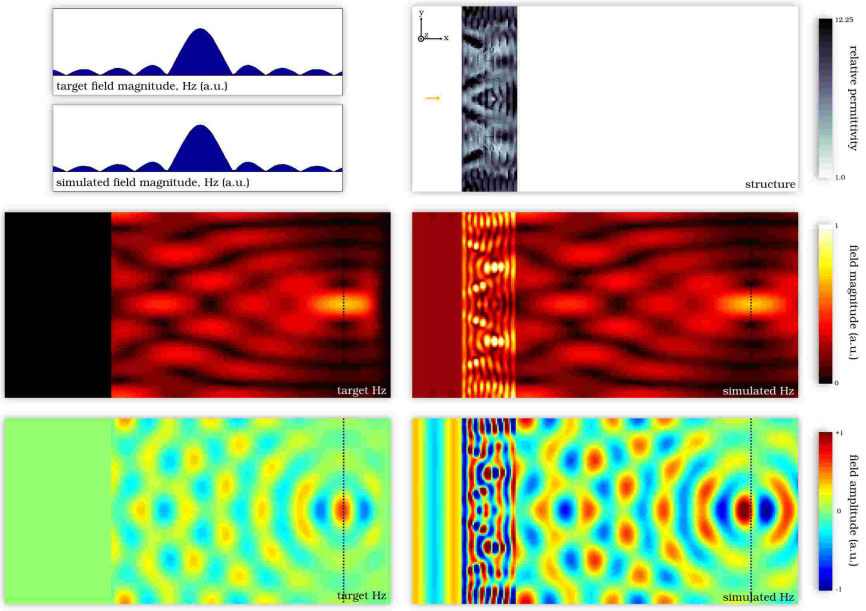

Fig. 1.17 test

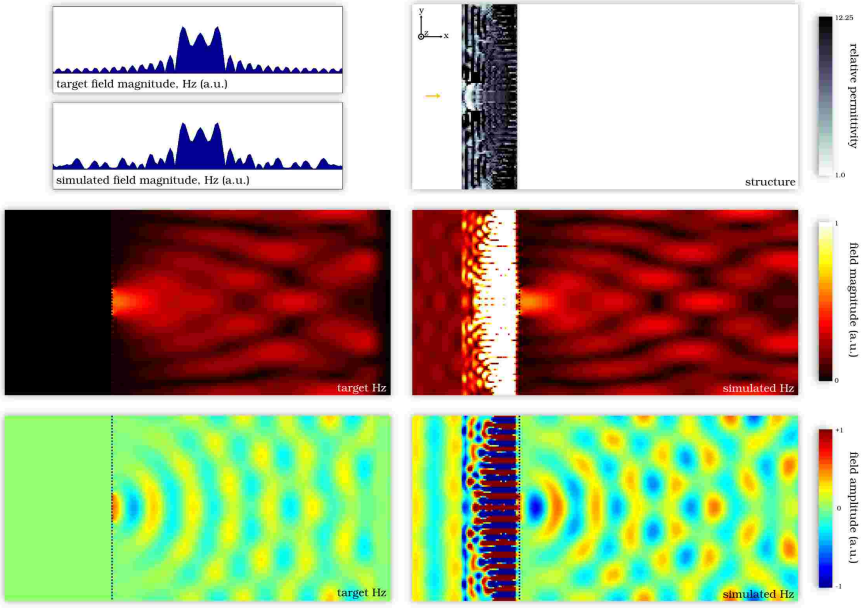

Fig. 1.18 test