Nonlinear design in nanophotonics

by

David Liu

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

Signature redacted

Author ....
Department of Physics
September 21, 2015

Signature redacted

Certified by. ....
Steven G. Johnson
Professor of Applied Mathematics
Thesis Supervisor

Signature redacted

Certified by. ....
John D. Joannopoulos
Francis Wright Davis Professor of Physics
Thesis Supervisor

Signature redacted

Accepted by ....
Nergis Mavalvala
Associate Department Head

Nonlinear design in nanophotonics

by

David Liu

Submitted to the Department of Physics
on September 21, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

In the first part of this thesis, we present a new technique for the design of transformation-optics devices based on large-scale optimization to achieve the optimal effective isotropic dielectric materials within prescribed index bounds. In addition to the optimization, a key point is the identification of the correct boundary conditions to ensure reflectionless coupling to untransformed regions while allowing maximum flexibility in the optimization. We apply our technique to the design of multimode waveguide bends and mode squeezers, in which all modes are transported equally without scattering. In the second part of this thesis, we introduce a direct, efficient, and flexible method for solving the non-linear lasing equations of the steady-state *ab initio* laser theory (SALT). We validate this approach in one-dimensional as well as in cylindrical systems, and demonstrate its scalability to full-vector three-dimensional calculations in photonic-crystal slabs. Our method paves the way for efficient and accurate simulations of microlasers which were previously inaccessible. In the third part of this thesis, we introduce a theory of degenerate lasing modes based on SALT. We present an analytical method to determine the stable superposition of lasing modes, and also a numerical method for cases in which the degeneracy is unphysically broken by the discretization. We demonstrate these ideas in examples such as a uniform dielectric cylinder, a metallic rectangular cavity, and a hexagonal cavity made of air holes.

Thesis Supervisor: Steven G. Johnson
Title: Professor of Applied Mathematics

Thesis Supervisor: John D. Joannopoulos
Title: Francis Wright Davis Professor of Physics

Acknowledgments

First, I want to thank my advisor Steven G. Johnson for teaching me how to solve problems, debug code, write papers, and give talks. Steven's approach is very practical and detail-oriented, and I am very fortunate to be able to learn the right way to do all of these things from the best. I also want to thank my thesis committee: Professors John D. Joannopoulos, Marin Soljačić, and Liang Fu for taking their time to read my thesis and provide suggestions. I would also like to thank my undergraduate advisor, Prof. Roscoe White at Princeton Plasma Physics Lab who introduced me to theoretical research.

My time spent in the Ab-Initio and Nanostructures and Computation groups have been full of learning. I would like to thank fellow group members Xiangdong Liang, Fan Wang, Adi Pick, Bo Zhen, Wade Hsu, and Ognjen Ilic for all their insight and help. The work we have done also would not have been possible without the expertise of our collaborators: I would like to thank Prof. Lucas Gabrielli, Prof. Michal Lipson, Alex Cerjan, Prof. Doug Stone, Prof. Li Ge, Matthias Liertzer, Sofi Esterhazy, and Prof. Stefan Rotter.

These past few years in Boston have been a blast, and would not have been the same without some of the great people I've met. I especially enjoyed taking classes and quals with Jimmy Liu those first two years. I want to thank Andrew Lai for fun times spent lifting weights in the gym, discussing science and philosophy in the student center. Additionally, I want to thank Vincent Liu and Leon Li for their friendship and the academic and career advice they provided.

Finally, I want to thank my parents Yu Liu and Ling Shen, as well as my sister Amanda for all their support.

Contents

List of Figures

Chapter 1

Introduction

In physics and applied mathematics, a very common recurring theme is the idea of a *linear* approximation of something that is *nonlinear*. In classical mechanics, one describes a mass on a spring with a linear differential equation, Hooke's Law, when in reality the behavior is nonlinear and contains terms to all powers in the displacement of the mass. In electrodynamics, one describes electromagnetic fields with Maxwell's equations in linear media, when in reality all media have some degree of nonlinearity and there are second and higher-order terms (it just takes very large fields for the higher-order terms to become noticeable). In quantum mechanics, one finds energy shifts due to minor modifications to a Hamiltonian by using perturbation theory to find the linear part of a response that has terms of all orders. In numerical analysis, one optimizes nonlinear *functions* by taking linear approximations of the function at each iteration to find the best step size and direction. One also finds solutions to nonlinear *equations* by approximating them by a sequence of *linear* equations (this is Newton's method [1–3]).

This thesis is about the numerical solution of complex problems in electromagnetism, but the most interesting part of any numerical algorithm is typically the analytical work to formulate the problem effectively and derive the algorithm, and in our case the interplay between linear and nonlinear behaviors is at the heart of this analysis. For many design problems in photonics, Maxwell's equations can be treated as approximately linear in the electromagnetic fields, but the solutions are highly nonlinear as a function of the geometric arrangement of materials. Optimization-based design, as in chapter 2 and in many other works [4–15], typically revolves around a sequence of linearizations of this geometric nonlinearity, but in chapter 2 we go further and eliminate the solution of Maxwell's equations entirely by combining large-scale optimization and a design technique called "transformation optics" [16–22] for the first time. The resulting theoretical design for a multi-mode integrated bend with unprecedented low intermode crosstalk was experimentally fabricated and validated by our collaborators in the Lipson group at Cornell [23]. In chapters 3 and 4, we turn to a problem in which Maxwell's equations are nonlinear in the electric field as well—the SALT (steady-state ab-initio lasing theory) equations of steady-state lasing [24–28], in which the laser gain saturates for a strong field. SALT provides the most tractable formulation of lasing theory, but even so it had only been previously solved in 1d and relatively simple 2d systems [25–27, 29–31]. To solve the nonlinear SALT equations tractably in 3d for the first time, the key (in chapter 3) was to combine SALT, modern computational-electromagnetism techniques, and the right choice of linearization so that we could solve the exact nonlinear problem (in 10^5 variables or more) by a tractable sequence of sparse linear problems. However, SALT theory itself required modification (in chapter 4) to generally handle the common case of lasing modes with degeneracies (pairs of equal-frequency solutions), such as the left and right-circulating modes in a ring laser [26,29,32,33]. Obtaining a correct, general SALT model capable of handling arbitrary degeneracies numerically required the technique of linearization in multiple guises. To solve the lasing problem near threshold (at the lowest powers where lasing occurs), we employed a variant of perturbation theory in the laser "pump" strength, and we were able to analytically derive a number of key properties of the degenerate lasing modes. This near-threshold solution is also the key to numerical solvers, because it forms the starting "guess" for the abovementioned Newton solvers at higher laser powers where the nonlinearity is strong. But at these powers, we had to invent a second linearization technique to correct for numerical degeneracy-splitting that arises from the computational discretization—these splittings are a familiar minor annoyance in the solution of linear wave resonances [34], but turn into a major obstacle in the nonlinear context of SALT for reasons described in chapter 4. In the remaining introduction, we introduce each of these topics from the later chapters at a less-technical level, and we will try to provide some of the background necessary to understand the problems that we solved and the state of the art before our contributions.

1.1 Steering light with coordinate transformations

Optical devices are everywhere in modern technology, and a key requirement is to steer light where you want to go: bending, warping, focusing, expanding, and otherwise distorting electromagnetic waves in prescribed ways. To do this, one uses the geometry: one has a menu of available materials (glass, Silicon, etc.) with different indices of refractions $n = \sqrt{\varepsilon\mu}$ (where ε is the dielectric permittivity and μ is the magnetic permeability) that one can put in different places with different shapes. Modern fabrication technologies such as the lithography used to make computer chips [35] give us remarkable freedom in the arrangement and shaping of materials, even near nanometer scales. The design challenge is to come up with the arrangement of materials that makes light do what you want. For any given arrangement, powerful computational techniques are available to solve for the electromagnetic wave behaviors [36–39], so one approach is to simply try different geometries until the solution is what you want. Of course, there are too many possibilities to try all of them, but one can use a combination of intuition, exact or approximate analytical results, symmetry, and computational search to hone in on a good design. Thousands of papers in optics have been published with designs based on these ideas. Most recently, many authors have been beginning to employ computational methods that can vary thousands or even millions of parameters to "discover" a structure that optimizes some optical behavior [4, 12, 13, 15, 40–44], but typical "inverse design" methods of this sort are extremely expensive (requiring a complete solution of Maxwell's equations at each parameter step) and therefore can only optimize relatively small regions of space compared to the wavelength of light (e.g. regions that are at most tens of wavelengths in diameter, and usually only a few wavelengths in 3d). We wanted to employ computational design for large multimode structures that could be 100 wavelengths in diameter, and so we came up with a new approach based on an idea called "transformation optics" (TO) which had not previously been employed for inverse design.

TO is a relatively new area of photonics that deals with how to design materials that warp light geometrically. It relies on an elegant mathematical equivalence in Maxwell's equations between coordinate transformations and material transformations [16–18, 21, 22, 45]. Maxwell's equations in the frequency (ω) domain (for linear time-invariant materials) in the electric $\mathbf{E}$ and magnetic $\mathbf{H}$ fields from the current $\mathbf{J}$ and charge $\rho = -\nabla \cdot \mathbf{J}$ densities, and for electric permittivity $\varepsilon(\mathbf{x}, \omega)$ and magnetic permeability $\mu(\mathbf{x}, \omega)$ are given by

$$\begin{aligned} \nabla \times \mathbf{H} &= -i\omega\varepsilon\mathbf{E} + \mathbf{J} \\ \nabla \times \mathbf{E} &= i\omega\mu\mathbf{H} \end{aligned} \tag{1.1}$$

These equations, which are written in the Cartesian coordinates $\mathbf{x}$, can be rewritten in arbitrary coordinates $\mathbf{x}'$ (with Jacobian matrix $\mathcal{J}$), without changing their form, by making the substitutions $\nabla = \mathcal{J}^{-1}\nabla$, $\mathbf{E}' = \mathcal{J}^{-1}\mathbf{E}$, $\mathbf{H}' = \mathcal{J}^{-1}\mathbf{H}$, and $\mathbf{J}' = \mathcal{J}^{T}\mathbf{J}/\det\mathcal{J}$. The equations are then rearranged in the form

$$\begin{aligned} \nabla' \times \mathbf{H}' &= -i\omega\left(\frac{\mathcal{J}^{T}\varepsilon\mathcal{J}}{\det\mathcal{J}}\right)\mathbf{E}' + \mathbf{J}' \\ \nabla' \times \mathbf{E}' &= i\omega\left(\frac{\mathcal{J}^{T}\mu\mathcal{J}}{\det\mathcal{J}}\right)\mathbf{H}' \end{aligned} \tag{1.2}$$

Amazingly, Eqs. 2.1 and 1.2 have the same form, with the only difference being that the effective permittivity ε and permeability μ become complex tensors involving the Jacobian in Eq. 1.2. Initially, TO began as a computational tool to solve problems more easily [16, 46]. For example, by applying a numerical solver in a Cartesian box to domains with other shapes (e.g. bend/cylindrical domains and other distortions), one can solve the distorted problem using the equivalent dielectric and magnetic material rather than reconstructing the solver for the new coordinate system, which requires considerably more work. Another example is the perfectly matched layer (PML) technique [38,47,48], which is a computational tool for numerical discretizations designed to accurately model infinitely large computational cells with outgoing radiation. The PML formulation is given by a complex coordinate transformation at the boundaries that is designed to "absorb" outgoing radiation; using TO, this transformation can be mapped to an artificial absorbing boundary material layer that effectively does the same thing. Consequently, a frequency-domain solver that handles arbitrary anisotropic ε and μ can implement a PML with no underlying changes.

More recently, TO has emerged as a computational tool to design real-life devices with useful photonic properties. Examples of devices designed using TO include spherical cloaks [17, 49, 50], ground-plane "carpet" cloaks [51–55], lenses [56–60], waveguide bends [41, 61–65], splitters, and many others . A useful and desired transformation of light, such as bending a light beam in a semicircle or spatially squeezing a guided wave, can be attained by simply filling the space where we would like the transformation effect to happen with the appropriate dielectric and magnetic material provided by the TO recipe. This recipe can be used without having to worry about the nature of the light; that is, all solutions of Maxwell's equations are transformed in the same way, as schematically shown in Fig 1-1. This quality makes TO an especially attractive design option for *multimode* systems, which involve many channels of data being carried in a single device, such as a multimode fiber [66]. Today's telecommunications systems are largely driven by integrated photonics [67], which is the optical analogy of electronic circuits, with the signals carried by on-chip waveguides rather than electronic interconnects. One major issue with on-chip photonic components is that they are almost exclusively single-mode, because they can only be routed effectively for a very narrow range of bandwidths: the turns must be designed for a small range of frequencies in mind; any other frequency would require a different construction. However, high-bandwidth applications have recently become increasingly important, due to the ability to carry much more data in multiple channels [66]. A crucial requirement for effective handling of multimode propagation is the prevention of *crosstalk*, or mixing between the different modes. Because TO-based devices are

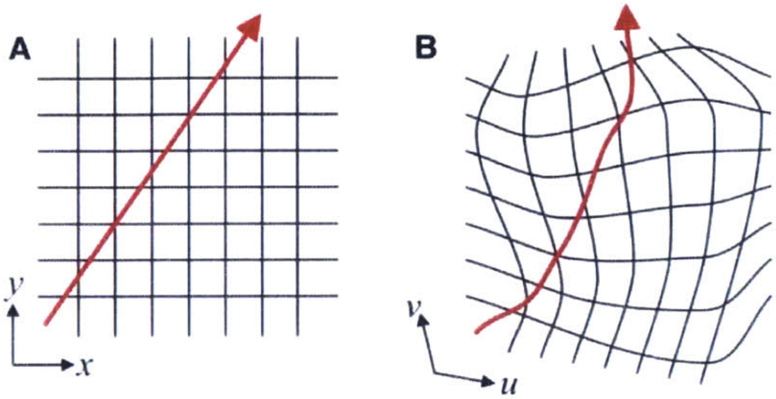


Figure 1-1: Schematic of a ray of light in original cartesian coordinates (left) and transformed coordinates (right), from Ref. [17].

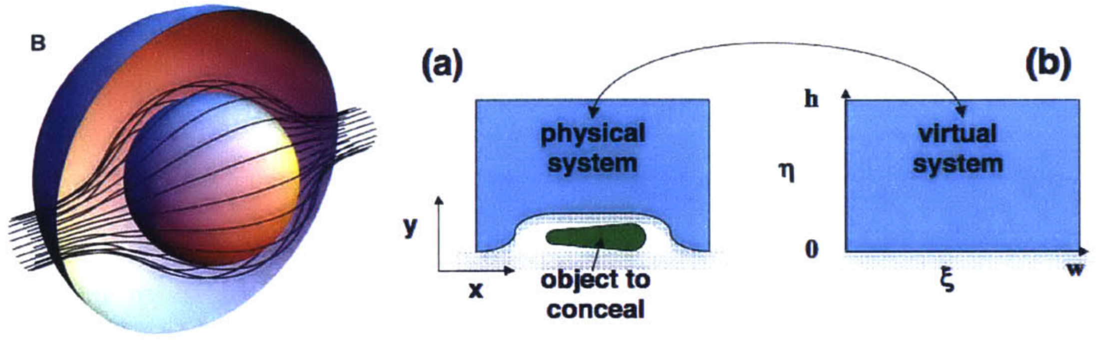


Figure 1-2: Two cloak designs based on transformation optics: spherical cloak (left, from Ref. [17]) and ground-plane carpet cloak (right, from Ref. [51]).

designed to transform all solutions of Maxwell's equations in the same geometric way, regardless of frequency, TO provides an especially attractive design option for dealing with multimode applications. However, previous work such as bends [62, 65, 68] and carpet cloaking [51, 54, 69] turned out to have both serious unaddressed problems with interface reflections (they only transform a certain region of space, but when light hits the boundary of that region it can have large reflections) and was highly suboptimal (the design was overconstrained in some ways, and underconstrained in others), as explained in chapter 2.

When one applies TO as a computational tool (e.g. for PML [48, 70]), whether the transformed ε and μ are physical or not are irrelevant; the computer does not care. However, if we want a design for a structure, like a cloak or a bend, that we actually

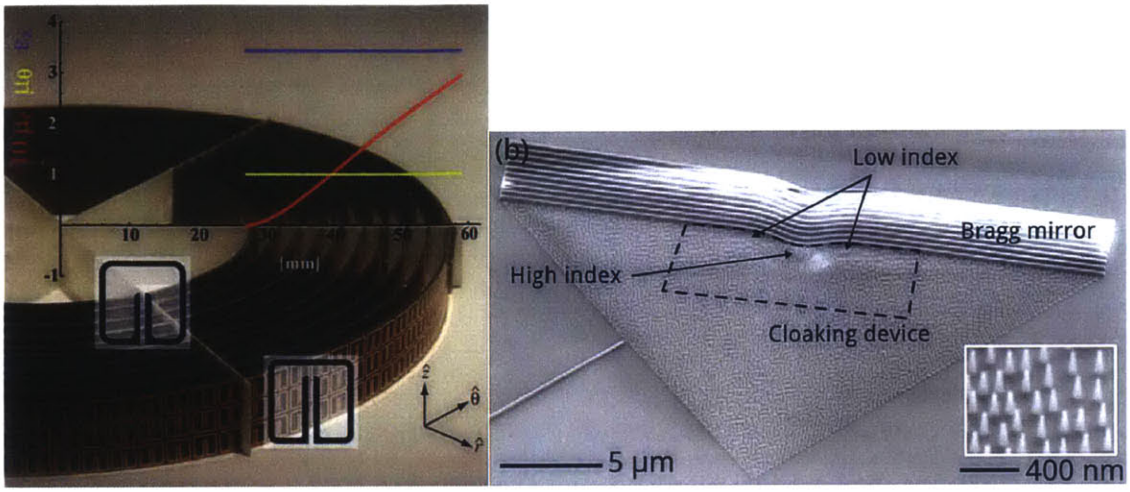


Figure 1-3: Examples of metamaterials. Left: resonant anisotropic magnetic structures, from Ref. [74]; Right: effective dielectric-only material using silicon pillars, from Ref. [75].

want to fabricate in the real world, then we have to *obtain* these materials somehow. In practice, the materials required for most of the arbitrary transformations we can think of are highly anisotropic, with all nine elements of both dielectric and magnetic tensors varying continuously in space. Needless to say, such control over material parameters is a very difficult engineering problem, though progress has been made on this point using resonant metamaterials [71–74], which are formed from subwavelength structures that can be tailored to create an effective anisotropic material with desired properties. While metamaterials have many exciting effective properties, they suffer from some notable shortcomings as well. First, for very anisotropic materials with large spatial variations, the nanostructures needed can be prohibitively difficult to fabricate accurately, and the avoidable imperfections and disorder that comes from limitations of the fabrication process can completely disrupt the intended effect [18]. Additionally, if we wanted truly universal geometric manipulation (that works for light of many frequencies), we would need dispersion-less (frequency-independent ε and μ) materials, which is almost impossible to realize in the resonant magnetic nanostructures [73] and lossy metals (for optical scales) that are typically needed.

One way around this significant roadblock is to try to find a transformation with an approximately *scalar* $\varepsilon > 1$, which can be realized by using methods such as grayscale lithography [76–79] or variable-radius air holes in a dielectric slab [80] to create an effective scalar index that varies continuously in some range. For some arbitrary desired distortion of light, e.g. a cloaked object or a bend, some numerical computation [81,82] is required to find a near-isotropic transformation that can be approximated by a scalar ε. However, in order to obtain the simplest possible computation, previous authors over-constrained the problem in some ways (they required the entire boundary shape of the transformation to be specified), and under-constrained it in others (they completely ignored the interface discontinuity), resulting in designs that had suboptimal performance. In particular, a multimode bend [62,63,65,68,83,84] designed in this way has unacceptably large interface reflections, and a carpet cloak [51] designed in this way requires an enormous "cloak" region compared to the size of the object being cloaked.

So instead, we start by examining the actual conditions for this scalarization to work: first, the transformation should be confined to a plane (i.e. with the z-coordinate unchanged) and with the in-plane transformation $x'(x, y), y'(x, y)$ satisfying (or satisfying as closely as possible) the Cauchy–Riemann equations of complex analysis (which give conformal maps) [85, 86]. This is because the Cauchy–Riemann equations happen to be mathematically equivalent to the condition that TO gives a nonmagnetic material with scalar ε for 2d transformations [87]. Second, the transformation should be continuous at the interfaces; that is, couple smoothly to untransformed regions at the boundaries. If this condition is not satisfied, i.e. there are jumps in the transformation or the Jacobian, then there will be singularities in the required materials which will lead to large scattering. Most work in TO dealing with designing isotropic dielectric materials [51, 62, 87–90] has focused on satisfying the first condition, without a serious treatment of the second condition. As explained in chapter 2, the second condition is at least as important as the first for properly designing TO devices with minimal scattering, and these two conditions cannot both be satisfied perfectly [86]. Thus, we have devised a powerful procedure, based on large-scale nonlinear optimization [91–94], to find the transformations that satisfy these two conditions as closely as possible, while at the same time parameterizing [95] and constraining the problem in just the right way. Using this procedure, we present a framework for design of isotropic dielectric TO devices that satisfy the essential conditions of having a realizable TO material, which has not been done in previous work.

1.2 Using a sequence of *linear* equations to solve a *nonlinear* laser equation

A laser [96,97] consists of three essential components, as shown in Fig. 1-4: a resonant cavity that traps light, a *gain* medium that *amplifies* the light trapped by the cavity through stimulated emission [33], and an external power source that pumps the gain medium to induce *population inversion* (having a large number of atoms in the excited state versus the ground state). To get a good sense of this process, we examine a simplified picture of a laser. For a resonant cavity without gain, the source-free solutions of Maxwell's equations are known as *resonances*, and their frequencies ω lie in the complex plane below the real axis (technically, the resonances are poles in the Green's function) as schematically shown in Fig. 1-5. These modes have a finite lifetime, defined in the positive quantity

$$Q \equiv -\frac{\mathrm{Re}\omega}{\mathrm{Im}\omega},$$

because their energy leaks out of the cavity over time; that is, the confinement is not perfect. As one increases the gain by pumping the cavity using an external power source, gain cancels loss and the resonances approach the real-ω axis (their lifetime increases). At a certain pump strength, the resonant mode actually reaches the real axis, and at this pump strength, called the "threshold", is when lasing starts happening. Instead of continuing up the complex plane past the real axis, the mode acquires a finite lasing *amplitude*, which *saturates* the gain and causes the system to reach a steady-state, with the gain (from the external pump and stimulated emission) balancing the loss (from energy radiating away from the imperfect confinement of the cavity).

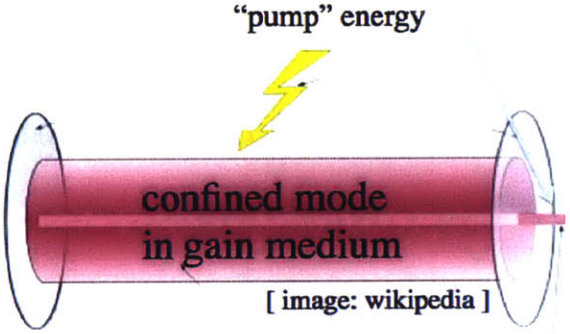


Figure 1-4: Schematic of the essential components of a laser. A laser consists of 1) a resonant cavity that traps light, 2) a gain medium embedded in that cavity, and 3) an external source that pumps the atoms of the gain to population inversion. When we have these three things, as in this setup consisting of two mirrors, we have a cavity mode in a steady lasing state, resulting in stimulated emission of coherent light.

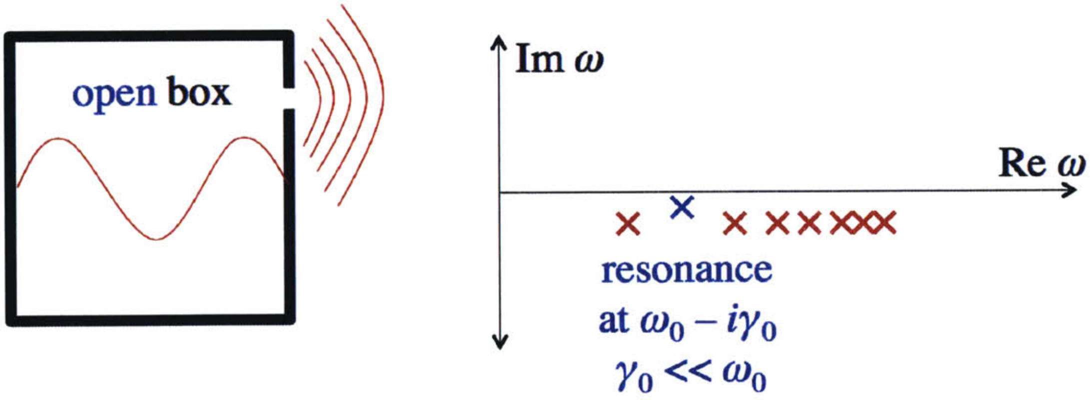


Figure 1-5: Simplified picture of resonant cavity. An oscillating dipole source, strongly peaked around ω_0, is placed inside the box, and energy "leaks" out through the hole. The resonances (technically poles of the Greens function) are shown on the complex plane below the real axis, with the lifetime of this leakage related to the imaginary part γ_0.

The simplest model of a gain medium is an ensemble of electrons in two-level atoms, and a crucial component of the stimulated emission leading to lasing is the population inversion of this ensemble of atoms. The semiclassical theory of lasers, which treats the electromagnetic fields classically using Maxwell's equations, and treats the dynamics of the atomic electrons quantum mechanically, was developed in the 1960's by Haken [33, 98] and also independently by Lamb [99]. The central equations of this theory are known as the Maxwell–Bloch (MB) equations (the "Bloch" part allegedly coming from the fact that the two-level atom part of the model mathematically resembled the quantum problem of spins in a magnetic field studied by the physicist Felix Bloch). It was the first truly *ab-initio* theory of lasers, describing a wide range of phenomena not captured by simpler models previously used (such as utilizing rate equations to describe photon number and population inversion, which is essentially a mean-field treatment).

While the MB equations were intended to describe real-world complex lasers with as few approximations as possible, somewhat ironically the first people to use the equations always began by greatly simplifying them and making drastic approximations to get the equations into a tractable form [100], often with the end goal of having an exact solution of a much simpler set of equations. This was because the MB equations, in their original form, are very difficult to solve even with the help of computers. First, there are multiple fields that one has to keep track of: in addition to the electric field $\mathbf{E}(\mathbf{x}, t)$, there is the gain polarization $\mathbf{P}(\mathbf{x}, t)$ and the population inversion $D(\mathbf{x}, t)$ (the density of atoms that are in the excited state). More importantly, all three fields are both space and time-dependent, so a very large amount of computational work is needed in order to solve the three coupled partial-differential equations that describe their time evolution. Finally, the dynamics of the laser contain multiple time scales that happen to be very separated in magnitude: for example, the optical frequency is much faster than the relaxation of polarization, which in turn is much faster than the relaxation of the population inversion. Because full time-domain simulation [37, 101, 102] must capture all three of these time scales, it may take an extremely long period of real-world computing time for the simulation to reach the

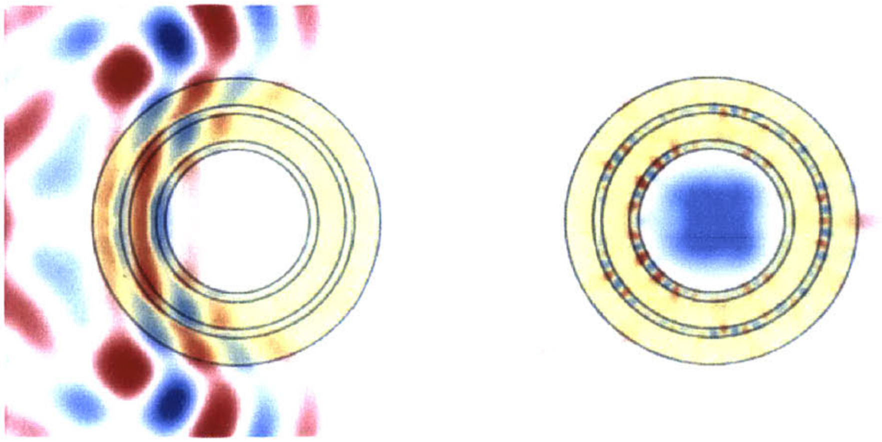


Figure 1-6: Time-domain simulation of cylindrical photonic crystal laser from Ref. [103]: An incident field (as seen in the left panel) excites a lasing mode which eventually reaches steady state (as seen in the right panel). Due to the presence of three different physical time-scales, it takes a very long time for the simulation to reach the steady state.

steady-state which gives us the most useful information about a laser (an example is shown and described in Fig. 1-6). Even with today's hardware, a full MB simulation of a 3d microlaser geometry, such as the examples in Fig. 1-7, is prohibitively expensive and would require a very large amount of computing power that is often not accessible to typical practitioners of laser theory.

To address these issues, Doug Stone's group at Yale University introduced the steady-state ab-initio lasing theory (SALT) in 2006 [28]. In a nutshell, SALT converts the three time-dependent partial-differential equations in three fields of MB into a single time-*independent* equation in a single field, **E**, greatly reducing the computational complexity for the solution. It does so by making a series of well-founded approximations (with minimal loss of generality) and an ansatz that the total electric field **E** is composed of a sum of steady-state *lasing modes*. The lasing modes are obtained by tracking the "passive" resonances (as explained above) to the real axis as the pump strength is increased, and then finding the steady-state lasing mode once the pole reaches the real axis. (Technically, the familiar "leaky mode" picture of resonances as slowly decaying "eigenfunctions" is only a local approximation for $\text{Im}\omega \ll \text{Re}\omega$ solutions, and the SALT model handles more general resonances [30].)

While SALT is much easier to solve computationally than MB, it is still a nonlinear eigenproblem with a nonlinear dependence on both the eigenfrequency ω and the eigenfunction $\mathbf{E}$. In the numerical analysis literature, there are many algorithms that deal with how to solve the nonlinear eigenvalue problem [104]

$$\hat{L}(\omega)\mathbf{E} = 0,$$

where $\hat{L}$ is an operator (or a matrix) that depends in a nonlinear way on the eigenfrequency ω, and $\mathbf{E}$ is an *eigenfunction.* However, the SALT equation is of the form

$$\hat{L}(\omega, |\mathbf{E}|^2)\mathbf{E} = 0, \tag{1.3}$$

with the matrix depending also on the eigenfunction. The most common problem of this form that is solved are the equations of density function theory (DFT) [105–107], but the methods of that theory cannot be applied here because unlike DFT, the SALT solution $\mathbf{E}$ does not minimize a functional, and also $\hat{L}$ is not hermitian. This diffulty presented a challenge to the first people who tried to solve SALT: they devised a method using basis functions known as "constant-flux" states to expand $\mathbf{E}$ [26,28], and solved a reduced version of Eq. 1.3 for the coefficients in the expansion of $\mathbf{E}$. However, a major issue with this method was that the construction of a specialized basis for each geometry was often unwieldy and not scalable to 3d complex geometries that are of the most interest to the laser community.

The closest thing to a standard method to solve general nonlinear equations is Newton's method [1,3]. but this method requires an initial guess that is already very near the root. Without such a guess, convergence to a root is at best extremely slow and at worst not even guaranteed. In our case however, we can exploit an important fact about lasers: the lasing solution is part of a continuous family of solutions as the pump strength is varied, starting from a linear (in $\mathbf{E}$) eigenproblem at threshold. This means that the solution for one pump strength will be very close to that for a nearby pump strength. Hence, one solution provides the perfect initial guess for Newton's method when solving for the nearby solution. When the gain is completely turned off,

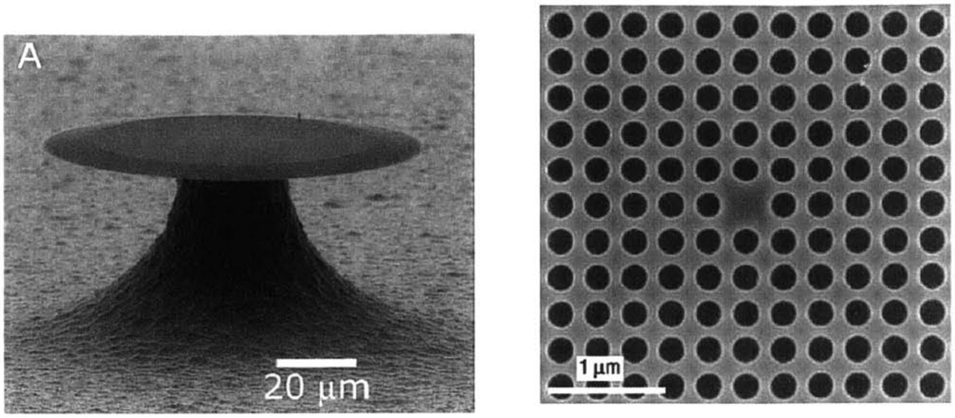


Figure 1-7: Two examples of experimentally fabricated lasers. Left: a microdisk laser from Ref. [108], and right: a photonic crystal laser from Ref. [109].

the problem is linear, so this provides a natural starting point, and the rest can be obtained by slowly increasing the pump strength, and successively using the previous solution as an initial guess. A second key point is that for a general discretization scheme such as finite-difference frequency-domain (FDFD) [38, 110] or finite-element method (FEM) [39, 111], there are often up to 10^7 unknowns, but the linear equations of each iteration of Newton's method are *sparse* [112]: that is, the matrix of coefficients contains mostly zeros. This is an advantage because there exists many fast algorithms for solving sparse linear problems [113, 114]. In Chapter 3, we put all of these points together and describe a framework for solving SALT directly (as opposed to indirectly using a specialized basis, as was done before), and we demonstrate the scalability of our method to the 3d complex geometries that were previously inaccessible.

1.3 Degenerate modes in the presence of a lasing nonlinearity

The third part of our thesis, which is an extension of the second area described above, deals with the fascinating problem of degeneracies in lasers with symmetric geometries. For equations in which the operator does not depend on the eigenfunction, whenever there there is a degeneracy and two eigenfunctions solve an equation with the same eigenfrequency:

$$\hat{L}(\omega)\mathbf{E}_1 = 0$$
$$\hat{L}(\omega)\mathbf{E}_2 = 0,$$

then it is easy to see that any superposition of the two modes also solves the equation with that frequency:

$$\hat{L}(\omega)(a_1\mathbf{E}_1 + a_2\mathbf{E}_2) = 0. \tag{1.4}$$

When the equations *are* nonlinear in the eigenfunctions (as in Eq. 1.3, and as they are in SALT), then it is no longer true that any superposition also solves the equation: that is, for

$$\hat{L}(\omega, |a_1\mathbf{E}_1 + a_2\mathbf{E}_2|^2)(a_1\mathbf{E}_1 + a_2\mathbf{E}_2) = 0, \tag{1.5}$$

the choice of a_1 and a_2 is *not* arbitrary, as it is in Eq. 1.4. Finding the correct coefficients in the two-mode superposition then becomes an important problem in obtaining the field-pattern of the mode that actually lases. The classical example of a degenerate pair of lasing modes is the ring resonator [32, 115]. The whispering-gallery modes come in pairs: standing-wave $\cos m\phi$ and $\sin m\phi$ modes (or alternatively, $e^{\pm im\phi}$ circulating modes). It is well-established that the steady-state superposition observed both experimentally and in time-domain simulations are the circulating $e^{\pm im\phi}$ modes [116]. This result also makes sense physically, because the circulating modes are the ones that utilize the gain medium most uniformly and efficiently: $|\mathbf{E}|^2$ is ϕ-invariant for $e^{\pm im\phi}$ modes, whereas standing-wave modes have zeros in ϕ where $\mathbf{E}$ is not using the gain.

However, whispering-gallery modes are only one of many possible degeneracies that can occur in the resonant cavities of lasers. There has been relatively little work in the laser literature on degenerate lasing modes other than whispering gallery modes for cylinders and rings. In general, degeneracies arise for two reasons. First, a degeneracy, known as an "accidental" degeneracy [34], can arise from a delicate balance in some parameters that has been carefully tuned. For example, a dielectric rectangle's aspect ratio can be adjusted such that two non-degenerate modes coalesce. However, any disruption of this careful tuning will also break the degeneracy. Since a consequence of lasing is that the gain (modelled in SALT by a complex dielectric) changes in a spatially complex way due to hole-burning, we expect such degeneracies to split above threshold. Second, the degeneracy may come from a geometric symmetry of the system [117, 118]. Examples include dipole modes in a cavity shaped like an equilateral triangle, or quadrupole modes in a hexagonal cavity. If there is a way for the superposition of degenerate modes to respect the original symmetry above threshold, then the degeneracy is maintained. It turns out that this is not the full story: as we will see in chapter 4, the intensity pattern for the correct superposition actually breaks mirror symmetry, but not rotational symmetry, so that the gain actually acquires chirality. In principle, this chirality should break the degeneracy (which can be seen with representation theory arguments). However, due to an elegant and somewhat unexpected consequence of a property known as Lorentz reciprocity of Maxwell's equations [119], the degeneracy actually remains.

Using this fact, we obtain an analytic form for the stable circulating solution in *all* geometries with the same symmetry group as regular polygons, known as the $C_{n\mathrm{v}}$ symmetry group [117, 118] (or the group with n-fold rotational symmetry along with mirror planes through each of the edges and faces of the regular polygon), and the same methods are extensible to other symmetry groups. We also show the validity of this solution (as well as the instability of others) by developing a threshold perturbation theory that casts the near-threshold SALT problem into a much smaller 2×2 nonlinear eigenproblem for the complex coefficients a_1 and a_2 in the superposition of degenerate threshold modes (Eq. 1.5). This eigenproblem is simple enough that it can be solved analytically to give both the lasing coefficients and the "passive" poles in the presence of the lasing nonlinearity, providing a powerful and general way to test the stability of any given lasing mode candidate. By using the symmetry arguments and perturbation theory described above in conjunction with the numerical solver we presented in chapter 3, we give a comprehensive procedure to obtain the stable lasing mode for two-fold degeneracies in any geometry with the $C_{n\mathrm{v}}$ symmetry of a regular polygon, opening up the SALT formalism to areas previously only minimally explored.

Finally, in the last part of chapter 4, we address an important point in the solution of degenerate systems using generic numerical discretization schemes such as finite-difference frequency-domain (FDFD) [38,110], which was used for our numerical solver in chapter 3. While regular polygon can have n-fold symmetry for any n, the Cartesian x-y grid used for FDFD is restricted to $C_{4\mathrm{v}}$ symmetry. As a result, degenerate modes can split when projected onto the grid, because the grid no longer has the symmetry responsible for the degeneracy in the first place. This presents a major problem for our procedure, because no superposition of two *nearly* degenerate modes solves Eq. 1.5, even if we take care to construct the stable circulating solution using symmetry and perturbation theory. To address this issue, we developed a method to artificially restore the degeneracy using a perturbation $\delta\varepsilon(\mathbf{x})$ to the dielectric function. Using a method known as quadratic programming [91], we find the smallest-normed perturbation that restores the degeneracy, which guarantees that our method converges to $\delta\varepsilon(\mathbf{x}) = 0$ in the infinite-resolution limit.

Chapter 2

Transformation Inverse Design

2.1 Overview

In this chapter, which was published in Ref. [120], we introduce the technique of transformation inverse design, which combines the elegance of transformation optics [16–19] (TO) with the power of large-scale optimization (inverse design), enabling automatic discovery of the best possible transformation for given design criteria and material constraints. We illustrate our technique by designing multimode waveguide bends [62, 63, 65, 68, 83, 84, 88, 121–126] and mode squeezers [83, 84, 127–129], then measuring their performance with finite element method (FEM) simulations. Most designs in transformation optics use either hand-chosen transformations [17, 20, 63, 68, 74, 88, 126, 127, 130–135] (which often require nearly unattainable anisotropic materials), or quasiconformal and conformal maps [22, 51–54, 57, 58, 69, 75, 89, 121–125, 127, 136–144] which can automatically generate nearly-isotropic transformations (either by solving partial differential equations or by using grid generation techniques) but still require *a priori* specification of the entire boundary shape of the transformation. Further, neither technique can directly incorporate refractive-index bounds. On the other hand, most inverse design in photonics involves repeatedly solving computationally expensive Maxwell equations for different designs [4–15, 40–44, 145–147]. Transformation inverse design combines elements of both transformation optics and inverse design while overcoming their limitations. First, the use of optimization allows us to incorporate

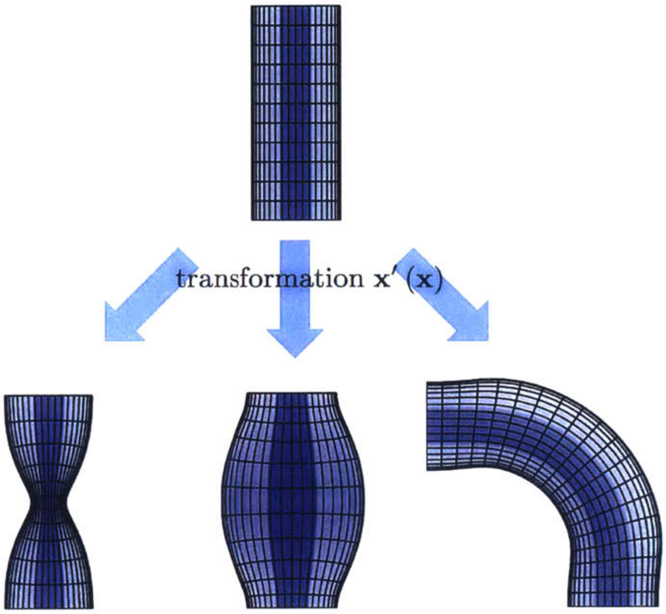


Figure 2-1: Three possible applications of transformation optics for multimode waveguides: squeezer, expander, and bend. Dark areas indicate higher refractive index.

arbitrary fabrication constraints while at the same time searching the correct space of transformations without unnecessarily underconstraining or overconstraining the problem. Second, instead of solving Maxwell's equations, we require only simple derivatives to be computed at each optimization step. This is because transformation optics works by using a coordinate transformation $\mathbf{x}'(\mathbf{x})$ that warps light in a desired way (e.g. mapping a straight waveguide to a bend, or mapping an object to a point or the ground for cloaking applications [17, 50–55, 139]) and then employing transformed materials which are given in terms of the Jacobian $\boldsymbol{\mathcal{J}}_{ij} = \partial x'_j / \partial x_i$ to mathematically mimic the effect of the coordinate transformation. This transforms all solutions of Maxwell's equations in the same way (as opposed to non-TO multimode devices which often have limited bandwidth and/or do not preserve relative phase between modes [11, 41, 61, 64, 148–154]), and is therefore particularly attractive for designing multimode optical devices [22, 45, 83, 87, 155] (such as mode squeezers, expanders, splitters, couplers, and multimode bends) with no intermodal scattering. Examples of such transformations are shown in Fig. 2-1.

One major difficulty with transformation optics is that most functions $\mathbf{x}'(\mathbf{x})$ yield highly anisotropic and magnetic materials. In principle, these transformed designs can be fabricated with anisotropic microstructures [71–74] or naturally birefringent materials [130, 134]. However, in the infrared regime (where metals are lossy) it is far easier to instead fabricate effectively isotropic dielectric materials, provided that the refractive index falls within the given bounds n_{min} and n_{max} of the fabrication process (for example, subwavelength nanostructures [53–55, 58, 71, 75, 80, 139, 156, 157] or waveguides with variable thickness [59, 60, 79, 158–160]). This requirement means that we would prefer to consider the subset of transformations that can be mapped to approximately isotropic dielectric materials.

The theory of transformation optics with nearly isotropic materials is intimately connected to the subjects of conformal maps (which are isotropic by definition [85, 86]) and *quasiconformal maps* [which in mathematical analysis are defined as *any* orientation-preserving transformation with bounded anisotropy (as quantified in Sec. 2.2.4)]. However, in transformation optics the term "quasiconformal" has become confusingly associated with only a single choice of quasiconformal map suggested by Li and Pendry [51]. In that work, Li and Pendry proposed minimizing a mean anisotropy with "slipping" boundary conditions (defined in Sec.2.2.4), which turns out to yield a transformation that is essentially conformal up to a *constant stretching* (and thus anisotropy) everywhere. This map, which also happens to minimize the peak anisotropy given the slipping boundary conditions [51, 161], is sometimes confusingly called "the quasiconformal map" [53, 54, 57, 121, 137]. However, we point out in Sec. 2.2.4 that slipping boundary conditions are not the correct choice if one wishes to ensure a reflectionless interface between transformed and untransformed regions. Instead, for interfaces to be reflectionless requires at least continuity of the transformation $\mathbf{x}'$ at the interface [90, 155, 162, 163] and, as we show in Sec. 2.2.3 for the case of isotropic dielectric media, continuity of the Jacobian $\mathcal{J}$ as well. If one fixes the transformation on part or all of the boundary (instead of just the corners) and minimizes the peak anisotropy, the result is called (in analysis) an *extremal* quasiconformal map [164–169]. We point out in Sec. 2.2.2 that this extremal quasiconformal map can never be conformal except in trivial cases. Additionally, previous work in quasiconformal transformation optics underconstrained the space of transformations in one way but overconstrained it in another. Li and Pendry's method, along with other work on extremal quasiconformal maps in mathematical analysis, assumed that the entire boundary *shape* of the transformed domain is specified *a priori* (even if the *value* of the transformation at the boundary is not specified). In contrast, transformation inverse design allows parts of the boundary shape to be freely chosen by the optimization, only fixing aspects of the boundary that are determined by the underlying problem (e.g. the input/output facets of the boundary in Fig. 2-1) as explained in Secs. 2.3.2, allowing a much larger space of transformations to be searched. Also, for such stricter boundary conditions, minimizing the mean anisotropy is *not* equivalent to minimizing the peak anisotropy [165, 170–172], and we argue below that the peak anisotropy is a better figure of merit for transformation optics in general.

We solve all of these problems by using large-scale numerical optimization to find the transformation with minimal peak anisotropy that exactly obeys continuity conditions at the boundary with untransformed regions. This allows the input/output interfaces to transition smoothly and continuously into untransformed devices while also satisfying fabrication constraints (e.g. bounds on the attainable refractive indices and bend radiii). A large space of arbitrary smoothly varying transformations (that satisfy the continuity conditions and fabrication constraints) is explored quickly and efficiently by parametrizing in a "spectral" basis [91, 95] of Fourier harmonics and Chebyshev polynomials. The optimized transformation is then scalarized (as in the case of previous work on quasiconformal transformation optics) into an isotropic dielectric material that guides modes with minimal intermodal scattering and loss. In the case of a multimode bend, for which our design was recently fabricated and characterized [23], we achieve intermodal scattering at least an order of magnitude smaller than a conventional non-TO bend.

In Sec. 2.2.1, we review the equations of transformation optics. In Sec. 2.2.2, we describe situations where the transformation-designed material can be mapped to isotropic media. In Sec. 2.2.3, we point out that such isotropic transformations, due to their analyticity, always have undesirable interface discontinuities when coupled into untransformed regions. In Secs. 2.2.4 and 2.2.5, we review the techniques of quasiconformal mapping (as used in both the transformation optics and mathematical analysis literature) and scalarization of nearly isotropic transformations. We show that the inherent restrictions of quasiconformal mapping can be circumvented by directly optimizing the map using transformation inverse design. In Secs. 2.3.1 and 2.3.2, we design a nearly isotropic transformation for a 90°-bend by perturbing from the highly anisotropic circular bend transformation. In Secs. 2.3.3 and 2.3.4, we set up the bend optimization problem and the spectral parameterization. In Sec. 2.4, we present the optimized structure, which reduces anisotropy by several orders of magnitude compared to the circular TO bend. In Sec. 2.4.1, we present finite element simulation results comparing our optimized design to the conventional non-TO bend and the circular TO bend. In Secs. 2.4.2 we show that minimizing the mean anisotropy can lead to pockets of high anisotropy (which in turn leads to greater intermodal scattering) while minimizing the peak does not. In Secs. 2.4.3, we discuss the tradeoff between the bend radius and the optimized anisotropy. In Sec. 2.6 we briefly present methods and results for applying transformation inverse design to optimize mode squeezers.

2.2 Mathematical preliminaries

2.2.1 Transformation optics

The frequency domain Maxwell equations (fields $\sim e^{-i\omega t}$), without sources or currents, in linear isotropic dielectric media $[\boldsymbol{\varepsilon} = \varepsilon(\mathbf{x}),\ \boldsymbol{\mu} = \mu_0]$ are

$$\begin{aligned} \nabla \times \mathbf{H} &= -i\omega\varepsilon(\mathbf{x})\mathbf{E} \\ \nabla \times \mathbf{E} &= i\omega\mu_0\mathbf{H}. \end{aligned} \tag{2.1}$$

Consider a coordinate transformation $\mathbf{x}'(\mathbf{x})$ with Jacobian $\mathcal{J}_{ij} = \frac{\partial x'_j}{\partial x_i}$. We define the *primed* gradient vector as $\nabla' \equiv \left(\frac{\partial}{\partial x'}, \frac{\partial}{\partial y'}, \frac{\partial}{\partial z'}\right) = \boldsymbol{\mathcal{J}}^{-1}\nabla$ and the primed fields as

$\mathbf{E}' \equiv \boldsymbol{\mathcal{J}}^{-1}\mathbf{E}$ and $\mathbf{H}' \equiv \boldsymbol{\mathcal{J}}^{-1}\mathbf{H}$. One can then rewrite Eq. 2.1, after some rearrangement [16, 21], as

$$\begin{aligned} \nabla' \times \mathbf{H}' &= -i\omega\boldsymbol{\varepsilon}'\mathbf{E}' \\ \nabla' \times \mathbf{E}' &= i\omega\boldsymbol{\mu}'\mathbf{H}', \end{aligned} \tag{2.2}$$

where the effects of the coordinate transformation have been mapped to the equivalent tensor materials

$$\boldsymbol{\mu}' = \mu_0 \frac{\mathcal{J}^T\mathcal{J}}{\det \boldsymbol{\mathcal{J}}} \qquad \boldsymbol{\varepsilon}' = \varepsilon(\mathbf{x}) \frac{\mathcal{J}^T\mathcal{J}}{\det \boldsymbol{\mathcal{J}}}. \tag{2.3}$$

This equivalence has become known as *transformation optics* (TO).

Most useful applications of TO require that the transformation be coupled to untransformed regions (e.g. the input and output straight waveguides in the case of a bend transformation, or the surrounding air region for the case of a ground-plane cloaking transformation). However, in order for TO to guarantee that the interface between transformed and untransformed regions be reflectionless, the transformation must be equivalent to a *continuous* transformation of all space that is the identity $\mathbf{x}'(\mathbf{x}) = \mathbf{x}$ in the "untransformed" regions, as depicted in Fig. 2-2. More generally, the untransformed regions can be simple rotations or translations, but when examining a particular interface, we can always choose the coordinates to be $\mathbf{x}' = \mathbf{x}$ at that interface. It is clear by construction that continuous $\mathbf{x}'$ is sufficient for reflectionless interfaces [155, 162, 163], and this is in fact a necessary condition as well [90]. Although a general anisotropic transformation need only have $\mathbf{x}'(\mathbf{x})$ continuous at the interface , we show below that an *isotropic* transformation will also have a continuous $\boldsymbol{\mathcal{J}}$ at the interface. These boundary conditions are essential for designing useful transformations without interface reflections.

2.2.2 Transformations to isotropic dielectric materials

For the vast majority of transformations, the materials in Eq. 2.3 are anisotropic tensors. However, for certain transformations, the tensors are effectively scalar. Suppose that

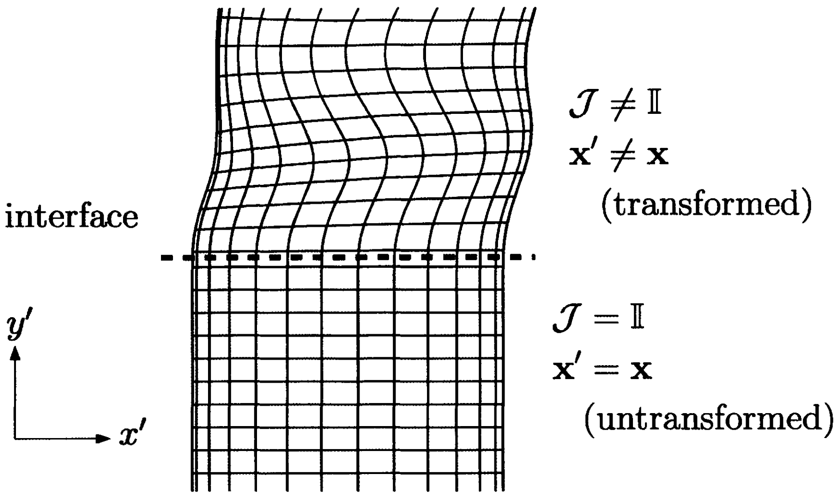


Figure 2-2: The interface between the transformed and untransformed region must have $\mathbf{x}'$ continuous in order for there not to be any interface reflections.

the transformation $\mathbf{x}'(\mathbf{x})$ is 2D ($z' = z$ and $\frac{\partial \mathbf{x}'}{\partial z} = 0$), making $\mathcal{J}$ block-diagonal (with the zz element independent of the xy block). Then, the xy block of $\mathcal{J}^T\mathcal{J}$ is isotropic if and only if the diagonal elements are equal and the off-diagonal elements vanish:

$$\begin{aligned} |\nabla x'|^2 - |\nabla y'|^2 &= 0 \\ \nabla x' \cdot \nabla y' &= 0. \end{aligned} \tag{2.4}$$

In this case, the $\mathcal{J}$ part of Eq. 2.3 becomes

$$\frac{\mathcal{J}^T\mathcal{J}}{\det \mathcal{J}} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \frac{1}{\det \mathcal{J}} \end{pmatrix}. \tag{2.5}$$

This isotropy has different implications for transverse-magnetic (TM) polarized modes in 2D (which have $\mathbf{E} = E\hat{z}$ and $\mathbf{H} \cdot \hat{z} = 0$) versus transverse-electric (TE) polarized modes (which have $\mathbf{E} \cdot \hat{z} = 0$ and $\mathbf{H} = H\hat{z}$). For TM-polarized modes, the fields $\mathbf{E}'$, $\mathbf{H}'$ in the primed coordinate system are also TM-polarized and Eq. 2.2

becomes

$$\nabla' \times \mathbf{H}' = -i\omega \frac{\varepsilon(\mathbf{x})}{\det \mathcal{J}} \mathbf{E}'$$
$$\nabla' \times \mathbf{E}' = i\omega\mu_0 \mathbf{H}'.$$

Hence, for TM modes, an isotropic transformation can be exactly mapped to an isotropic dielectric material. Similarly, for TE-polarized modes the equivalent material is isotropic and *magnetic*. However, if $\det \mathcal{J}$ varies slowly compared to the wavelengths of the fields, then the transformation can *approximately* still be mapped to an isotropic dielectric material by making an eikonal approximation (as in Ref. 173 Ch. 8.10) and commuting $1/\mu'$ with one of the curls in the Maxwell equations. In particular, Eq. 2.2 can be written:

$$\nabla' \times \nabla' \times \mathbf{E}' = \omega^2 \varepsilon_0 \mu_0 (\det \mathcal{J}) \mathbf{E}' + \mathcal{O}(\nabla \det \mathcal{J}),$$

where the last term can be neglected for slowly varying transformations. Because the TM case is conceptually simpler and does not require this extra approximation, we work with it exclusively for the rest of this paper. Also, because the non-trivial aspects of the transformation occur in the xy plane, we hereafter use $\mathcal{J}$ to denote the xy block.

2.2.3 Conformal maps and uniqueness

If a transformation has an isotropic $\mathcal{J}$, then the transformation preserves angles in the xy plane. Additionally, if $\det \mathcal{J} > 0$, then the transformation also preserves handness and orientation. The combination of these two properties is called a conformal map [85], and is the only case where the situation in Sec. 2.2.2 can be realized. We only consider transformations with $\det \mathcal{J} > 0$ in order to restrict ourselves to dielectric materials. Also, a $\det \mathcal{J} > 0$ transformation coupled continuously to an untransformed ($\det \mathcal{J} = 1$) region would require singularities ($\det \mathcal{J} = 0$) at some points. Conformal maps are described by analytic functions, which are of the form $x' + iy' = w'(w)$

(where $w \equiv x + iy$ is the untransformed complex coordinate) and whose real and imaginary parts satisfy the Cauchy–Riemann equations of complex analysis [85,86].

However, true conformal maps cannot directly be used for transformation optics in typical applications, because of the impossibility of coupling them to untransformed regions with the boundary conditions discussed in Sec. 2.2.1. In particular, the uniqueness theorem of analytic functions [86, Thm. 10.39] tells us that if $w'(w) = w$ in some region, then $w'(w) = w$ everywhere (similarly for a simple rotation or translation in some regions).

As a corollary, in the limit where a transformation becomes more and more isotropic in the neighborhood of an interface, it must have a continuous $\mathcal{J}$, not just a continuous $\mathbf{x}'(\mathbf{x})$. It is easy to see this explicitly in the example of Fig. 2-2: continuity of $\mathbf{x}'(\mathbf{x})$ at the interface requires that $\frac{\partial \mathbf{x}'}{\partial x} = (1,\, 0)$ on both sides of the interface, which determines the first row of $\mathcal{J}$. The isotropy of $\mathcal{J}^T\mathcal{J}$ then forces $\mathcal{J} = \mathbb{I}$. Therefore, in the sections that follow (where we search for *approximately* isotropic maps), we will impose the condition of continuous $\mathcal{J}$ as a boundary condition on our transformations. The resulting transformations are nearly isotropic in the interior and exactly isotropic on the interfaces. This condition, discussed at the end of Sec. 2.2.5, also has the useful consequence of producing a continuous refractive index $n' = \sqrt{\varepsilon'\mu'}$.

2.2.4 Quasiconformal maps and measures of anisotropy

Because true conformal maps cannot be used, one widely used alternative is to search for a nearly isotropic transformation, which can be *approximated* by an isotropic material at the cost of some scattering corrections to the exactly transformed modes of the nearly isotropic material. To do this, one must first quantify the measure of anisotropy that is to be minimized. The isotropy condition (Eq. 2.4) is equivalent to $\lambda_1 = \lambda_2$, where $\lambda_1(x,\, y) \geq \lambda_2(x,\, y)$ are the two eigenvalues of $\mathcal{J}^T\mathcal{J}$. While $\lambda_1 - \lambda_2$ works as a measure of anisotropy, it is convenient for optimization purposes to define *differentiable* measures that can be expressed directly in terms of the trace and determinant of $\mathcal{J}$, and precisely such quantities have been developed in the literature on quasiconformal maps [164,165,169,174,175].

A general transformation is an arbitrary function of x and y or, equivalently, an arbitrary function $w'(w,\, \bar{w})$, of w and $\bar{w} = x - iy$, which may not be analytic in w. The anisotropy can be related to the *Beltrami coefficient* [165, 169]

$$\mu_{\mathrm{B}}\,(w,\, \bar{w}) \equiv \left(\frac{\partial w'}{\partial \bar{w}}\right)\left(\frac{\partial w'}{\partial w}\right)^{-1}.$$

The term *quasiconformal map* refers to *any* map that has bounded $|\mu_{\mathrm{B}}| < 1$, which includes all non-singular sense preserving ($\det \mathcal{J} > 0$) transformations. It can be shown that the *linear distortion* K [164, 165] satisfies

$$K \equiv \frac{1 + |\mu_{\mathrm{B}}|^2}{1 - |\mu_{\mathrm{B}}|^2} = \sqrt{\frac{\lambda_1}{\lambda_2}} \geq 1.$$

Various other measures of anisotropy have been defined in the grid generation literature [81, 82], including the *Winslow* and *Modified Liao* functionals, which are given by $\Phi \equiv \int \mathrm{d}^2x\,(K + \frac{1}{K})$ and $\Phi \equiv \int \mathrm{d}^2x\,(K^2 + \frac{1}{K^2})$, respectively. However, for the rest of this work we refer to the quantity $\mathbb{K} - 1 \geq 0$ as the "anisotropy", where $\mathbb{K}$ is the *distortion function* [164, 165], defined as

$$\mathbb{K}\,(x,\, y) \equiv \frac{1}{2}\left(K + \frac{1}{K}\right) = \frac{\mathrm{tr}\mathcal{J}^T\mathcal{J}}{2\det\mathcal{J}} \geq 1. \tag{2.6}$$

The tensor $\frac{\mathcal{J}^T\mathcal{J}}{\det\mathcal{J}}$ is known as the *distortion tensor* [165].

As mentioned in the introduction, an *extremal* quasiconformal map is one that minimizes the peak anisotropy, given the shape of the transformed region and the values of the transformation on some or all of the boundary [164, 165, 169, 174]. Because $\mathbb{K}$, $\mathbb{K} - 1$, K, and $|\mu_{\mathrm{B}}|$ are all monotonic functions of one another, they are equivalent for the purpose of finding an extremal quasiconformal map. However, $\mathbb{K}$ is numerically convenient because it is a differentiable function of the entries of $\mathcal{J}$. These quantities are *not* generally equivalent for minimizing the *mean* anisotropy [165, 170–172], and we argue in Sec. 2.4.2 that the peak anisotropy is a better figure of merit. However, in the special case where the value of the transformation is only fixed at the *corners* of the domain and is allowed to vary freely in between (a "slipping"' boundary condition), Li and Pendry showed that it is equivalent to minimize the mean (either Winslow or modified-Liao) or the peak anisotropy, and that these yield a constant-anisotropy map (a uniform scaling of a conformal map) [161] that they and other authors have used for transformation optics [22, 52–54, 57, 58, 69, 75, 89, 121, 127, 136–141]. However, the slipping boundary will generally lead to reflections at the interface between the transformed and untransformed regions because of the resulting discontinuity in the transformation, which can only be reduced by making the transformation domain very large in cases (e.g. cloaking) with localized deformations. In order to design compact transformation-optics devices, especially for applications such as bends where the deformation is nonlocalized, we will instead impose continuity of the transformation and/or its Jacobian on the input/output facets of the domain, while at the same time allowing the shape of some or all of the boundary to vary (unlike all previous work on quasiconformal maps, to our knowledge).

2.2.5 Scalarization errors for nearly isotropic materials

The minimum-anisotropy quasiconformal map is then *scalarized* (as in Ref. 51) by approximating it with an isotropic dielectric material. As shown in Sec. 2.2.2, a perfectly isotropic 2D transformation of a geometry with an isotropic dielectric material that guides TM modes $\mathbf{E}_0$, $\mathbf{H}_0$ can be mapped to a transformed material and geometry that is also isotropic dielectric and guides TM modes $\mathbf{E}'_0$, $\mathbf{H}'_0$. This is exact for $\mathbb{K} = 1$, but for a *nearly* isotropic transformation with $\mathbb{K} > 1$, the equivalent permeability is $\boldsymbol{\mu}' = \mathbb{I} + \Delta\boldsymbol{\mu}$, where the anisotropic part $\Delta\boldsymbol{\mu}$ is proportional to $\mathbb{K} - 1$ to lowest order. While $\Delta\boldsymbol{\mu}' \neq 0$ cannot be fabricated using dielectric gradient index processes, one can neglect this small correction so that the actual fabricated material has permeability $\boldsymbol{\mu}'_{\text{approx}} = \mathbb{I}$. In practice, we absorb any $\Delta\boldsymbol{\mu}'$ into ε' by multiplying ε' by the average eigenvalue of $\boldsymbol{\mu}'$

$$\langle \boldsymbol{\mu}' \rangle = \frac{\lambda_1 + \lambda_2}{2\sqrt{\lambda_1 \lambda_2}} = \frac{\mathrm{tr}\mathcal{J}^T\mathcal{J}}{2\det\mathcal{J}} \tag{2.7}$$

but this does not change the $\mathcal{O}(\mathbb{K} - 1)$ error.

A Born approximation [176–178] tells us that, given an exact transformation with no scattering, any small error of $\Delta\boldsymbol{\varepsilon}$ and $\Delta\boldsymbol{\mu}$ will generically lead to scattered fields with magnitudes of $\mathcal{O}(|\Delta\boldsymbol{\varepsilon}| + |\Delta\boldsymbol{\mu}|)$ and scattered power of $\mathcal{O}(|\Delta\boldsymbol{\varepsilon}|^2 + |\Delta\boldsymbol{\mu}|^2)$. The modes of the approximate scalarized material $\boldsymbol{\mu}'_{\text{approx}}$, $\varepsilon'_{\text{approx}}$ are then the exact guided modes plus scattered power corrections of $\mathcal{O}(|\Delta\boldsymbol{\mu}|^2) = \mathcal{O}[(\mathbb{K} - 1)^2]$.

A similar analysis explains why we must explicitly impose continuity of $\mathcal{J}$ at the input/output facets of the domain. As explained in Sec. 2.2.3, a purely isotropic transformation in the neighborhood of the interface, along with a continuity of $\mathbf{x}'$, would automatically yield continuous $\mathcal{J}$, so one might hope that minimizing anisotropy would suffice to obtain a nearly continuous $\mathcal{J}$. Unfortunately, as we show below, the resulting discontinuity in $\det \mathcal{J}$ (and hence the discontinuity in the refractive index) is of order $\mathcal{O}(\sqrt{\mathbb{K} - 1})$, which would lead to $\mathcal{O}(\mathbb{K} - 1)$ power loss due to reflections, much larger than the $\mathcal{O}[(\mathbb{K} - 1)^2]$ power scattering from anisotropy in the interior. This would make it pointless to minimize the anisotropy in the interior, since the boundary reflections would dominate. In fact, our initial implementation of the bend optimization in Sec. 2.3.4 did not enforce continuity of $\mathcal{J}$, and we obtained a large 2% index discontinuity at the endfacets for $\max_{\mathbf{x}} \mathbb{K} - 1 \approx 0.0005$. Therefore, in Sec. 2.3.4 we impose continuity of $\mathcal{J}$ explicitly.

Here, we briefly derive the fact that the endfacet discontinuity scales much worse with anisotropy than the scalarization errors in the transformation interior, which leads us to impose an explicit continuity constraint on the Jacobian $\mathcal{J}$. In particular, we examine the Jacobian $\mathcal{J}$ for nearly isotropic transformations ($\mathbb{K} \approx 1$) that also have $\mathbf{x}' = \mathbf{x}$ explicitly constrained at the interfaces. (The following analysis can also be straightfowardly extended to situations where $\mathbf{x}'$ is a simple rotation of $\mathbf{x}$ on the interface, or where the interface has an arbitrary shape.) In this case, the Jacobian is

$$\mathcal{J} = \begin{pmatrix} 1 & 0 \\ \delta & 1 + \Delta \end{pmatrix},$$

where $\delta \equiv \frac{\partial x'}{\partial y}$ and $\Delta \equiv \frac{\partial y'}{\partial y} - 1$ are small quantities ($\ll 1$) if $\mathcal{J}^T\mathcal{J}$ is nearly isotropic.

The anisotropy (Eq. 2.6) is then:

$$\begin{aligned}\mathbb{K} - 1 &= \frac{1 + \delta^2 + (1 + \Delta)^2}{2\,(1 + \Delta)} - 1 \\ &= \frac{1}{2}\left(\delta^2 + \Delta^2\right) + \mathcal{O}\left(\delta^2 \Delta + \Delta^3\right). \end{aligned} \tag{2.8}$$

The determinant then satisfies

$$\begin{aligned}\det \mathcal{J} - 1 &= \Delta \\ &= \sqrt{2\,(\mathbb{K} - 1) - \delta^2} + \mathcal{O}(\delta^2\Delta + \Delta^3) \\ &= \mathcal{O}(\sqrt{\mathbb{K} - 1}).\end{aligned}$$

This square-root dependence is also reflected in the refractive index $n' = \sqrt{\varepsilon' \mu'}$ and leads to $\mathcal{O}(\mathbb{K} - 1)$ power loss due to interface reflections that overwhelm the $\mathcal{O}[(\mathbb{K} - 1)^2]$ corrections to scattered power due to the scalarization of nearly isotropic transformations (as explained in Sec. 2.2.5). Hence, it becomes necessary to explicitly constrain $\mathcal{J} = \mathbb{I}$ *in addition* to $\mathbf{x}' = \mathbf{x}$.

2.2.6 General optimization of anisotropy

In this paper, we directly minimize $\mathbb{K}$ using large-scale numerical optimization while keeping track of constraints on the transformation $\mathbf{x}'$ and its Jacobian $\mathcal{J}$, as well as the engineering fabrication bounds $n_{\min}$ and $n_{\max}$. By using numerical optimization, we can in principle achieve both a lower mean anisotropy and a lower peak anisotropy than by traditional quasiconformal mapping, since the optimization is also free to vary the boundary shape (with at most the input/output interfaces fixed, although in some cases their locations and shapes are allowed to vary as well). The minimization problem can be written, for example, as

$$\min_{\mathbf{x}'(\mathbf{x})} \|\mathbb{K}(\mathbf{x})\| \text{ subject to } \begin{cases} \mathbf{x}',\ \mathcal{J} \text{ continuous at input/output interfaces} \\ n_{\min} \leq n'(\mathbf{x}) \leq n_{\max} \end{cases}, \tag{2.9}$$

where $||\mathbb{K}(\mathbf{x})||$ is a *functional* norm taken over the domain of $\mathbf{x}'(\mathbf{x})$. We consider two possible norms: the L_1 norm (the mean $\langle\mathbb{K}\rangle_{\mathbf{x}}$), and the L_∞ norm ($\max_{\mathbf{x}} \mathbb{K}$). We show in Sec. 2.4.2 that minimizing the mean can lead to pockets of high anisotropy which can cause increased scattering. Directly optimizing the peak anisotropy on the other hand, avoids such pockets while simultaneously keeping the mean nearly as low. The continuity of $\mathbf{x}'$ and $\mathcal{J}$ at the input/output interfaces, as well as other constraints on the interface locations, are imposed implicitly by the parametrization of $\mathbf{x}'(\mathbf{x})$ (as explained in Sec. 2.3.4).

2.3 Multimode Bend design

In this section, we design a bend transformation (depicted in Fig. 2-3) using general methods to (locally) solve the optimization problem of Eq. 2.9. In contrast, previous work on TO bend design either utilized materials that were either anisotropic or consisted of multiple stacked isotropic layers [63, 65, 68, 83, 84, 88, 126] or employed slipping boundary conditions [121–125] (which result in endfacet reflections when coupled to untransformed waveguide).

2.3.1 Simple circular bends

First, we consider a simple circular bend transformation (which we refer to hereafter as the circular TO bend) that maps a rectangular segment of length L and width unity (in arbitrary distance units to be determined later) into a bend with inner radius R and outer radius $R+1$ (as shown in Fig. 2-3). For convenience, we choose the untransformed coordinates to be $R \leq x \leq R+1$ and $-\frac{L}{2} \leq y \leq \frac{L}{2}$, with the untransformed segment length $L = \frac{\pi R}{2}$ equal to the inner arclength of the bend. The transformation $\mathbf{x}'(\mathbf{x})$ can be written as

$$\begin{aligned} x' &= r\,\cos\theta \\ y' &= r\,\sin\theta \\ z' &= z, \end{aligned} \tag{2.10}$$

where $r = x$ and $\theta = \frac{y}{R}$. While $\mathbf{x}'$ is continuous at the input/output interfaces $y = \pm\frac{L}{2}$, one issue is that $\mathcal{J}$ is not continuous there, which can be seen from $\det\mathcal{J} = \frac{x}{R} \neq 1$. Another issue is that $\boldsymbol{\mu}' \neq \mu_0\mathbb{I}$ is highly anisotropic. The anisotropy for this transformation is $\mathbb{K}(x,\, y) - 1 = \frac{x}{2R} + \frac{R}{2x} - 1$, which has a peak value $\max_{\mathbf{x}} \mathbb{K} - 1 \approx \frac{1}{2R^2}$ for $R \gg 1$ at the outer radius $x = R{+}1$. Note that one can instead choose $r = \exp(\frac{\pi x}{2L})$, which gives the *conformal* bend $x' + iy' = \exp[\frac{\pi}{2L}(x + iy)]$. As explained in Sec. 2.2.3, this map has zero anisotropy, but neither $\mathbf{x}'$ nor $\mathcal{J}$ are continuous at the input/output interfaces, leading to large reflections there.

2.3.2 Generalized bend transformations

In order to address the problems of the circular TO bend, we look for minimum anisotropy *and* continuous-interface transformations of the form of Eq. 2.10, where the intermediate polar coordinates are now arbitrary functions $r(x,\, y)$ and $\theta(x,\, y)$. The ratio L/R is now an optimization parameter. The Jacobian then satisfies

$$\begin{aligned} \mathrm{tr}\mathcal{J}^T\mathcal{J} &= |\nabla r|^2 + |r\nabla\theta|^2 \\ \det\mathcal{J} &= |\nabla r \times r\nabla\theta|\,. \end{aligned}$$

We find that the optimization always seems to prefer a symmetric bend (and if the optimum is unique, it must be symmetric), so we impose a mirror symmetry in order to halve our search space:

$$\begin{aligned} r(x,\, y) &= r(x,\, -y) \\ \theta(x,\, y) &= -\theta(x,\, -y). \end{aligned} \tag{2.11}$$

We also require interface continuity of $\mathbf{x}'$ and $\mathcal{J}$ (as discussed in Sec. 2.2.3, which give the conditions at $y = \pm\frac{L}{2}$:

$$\begin{aligned} r &= x \\ \theta &= \pm\frac{\pi}{4} \\ \frac{\partial r}{\partial y} &= 0 \\ \frac{\partial \theta}{\partial y} &= \frac{1}{x}. \end{aligned} \tag{2.12}$$

2.3.3 Numerical optimization problem

Besides minimizing the objective function $\mathbb{K}$, the optimization must keep track of several constraints. First, any fabrication method will bound the overall refractive index n' to lie between some values n_{min} and n_{max}. We choose units so that the width of the transformed region is unity ($R \leq x \leq R+1$), and consider transforming a straight waveguide of width $\Delta_{\text{w}} < 1$. Δ_{w} should be small enough so that the exponential tails of the waveguide modes are negligible outside the transformed region. In the straight waveguide segment to be transformed (as well as the straight waveguides to be coupled into the input and output interfaces of the bend), $n(\mathbf{x})$ is high in the core $\left|x - R - \frac{1}{2}\right| < \frac{\Delta_{\text{w}}}{2}$ and low in the cladding $\left|x - R - \frac{1}{2}\right| > \frac{\Delta_{\text{w}}}{2}$. For convenience, we write this refractive index as a product $n(\mathbf{x}) = n_0 p(x)$ of an overall refractive index n_0 and a normalized profile $p(x)$ that is unity in the cladding and some value greater than unity in the core (determined by the ratio of the high and low index regions of the straight waveguide). The transformed refractive index is given by

$$n'(\mathbf{x}) = \sqrt{\varepsilon'\mu'} = n_0 p(x) \sqrt{\frac{\mathrm{tr}\mathcal{J}^T\mathcal{J}}{2(\det\mathcal{J})^2}}$$

where the average eigenvalue μ' of the magnetic permeability (Eq. 2.7) has been absorbed into the dielectric index. The overall refractive-index scaling n_0 is then allowed to freely vary as a parameter in the optimization. Second, like the circular TO bend, the optimum TO bend is expected to have a tradeoff between the bend radius and anisotropy. Because of this expected tradeoff, we can choose to either minimize R while keeping $\mathbb{K}$ fixed, or minimize $\mathbb{K}$ while keeping R fixed. We focus on the latter choice, since the bend radius is the more intuitive target quantity to know beforehand. Also, we find empirically that optimizing $\mathbb{K}$ converges much faster than optimizing R while yielding the same local minima.

With these constraints, there are several ways to set up the optimization problem, depending on which norm we are minimizing. One method is to minimize the peak anisotropy $\max_{\mathbf{x}} \mathbb{K}$ with $\mathbf{x} \in G$ for some grid G of some points to be defined in Sec. 2.3.4. However, the peak (the L_∞ norm) is not a differentiable function of the design parameters, so it should not be directly used as the objective function. Instead, we perform a standard transformation [91]: we introduce a dummy variable t and indirectly minimize the peak $\mathbb{K}$ using a differentiable inequality constraint between t and $\mathbb{K}(\mathbf{x})$ at all $\mathbf{x} \in G$:

$$\min_{r(\mathbf{x}),\,\theta(\mathbf{x}),\,n_0,\,L,\,t} t \text{ subject to}: \begin{cases} \text{continuity conditions 2.11, 2.12} \\ n_{\min} \leq n_0 p\left(x\right) \sqrt{\frac{\mathrm{tr}\mathcal{J}^T\mathcal{J}}{2(\det\mathcal{J})^2}} \leq n_{\max} \,\text{for}\, \mathbf{x} \in G \\ R = R_0 \\ \mathbb{K}(\mathbf{x}) \leq t \,\text{for}\, \mathbf{x} \in G \end{cases} . \tag{2.13}$$

For comparison, we explain in Sec. 2.4.2 why the L_∞ norm is better to minimize than the L_1 norm (the mean anisotropy).

The minimization of the L_1 norm, $\langle\mathbb{K}\rangle_{\mathbf{x}} = \int \mathbb{K}\,\mathrm{d}x\,\mathrm{d}y/\text{area}$ [which *is* differentiable in terms of the parameters $r(\mathbf{x})$, $\theta(\mathbf{x})$, n_0, and L] is implemented as

$$\min_{r(\mathbf{x}),\,\theta(\mathbf{x}),\,n_0,\,L} \langle\mathbb{K}\rangle_{\mathbf{x}} \text{ subject to}: \begin{cases} \text{continuity conditions 2.11, 2.12} \\ n_{\min} \leq n_0 p\left(x\right) \sqrt{\frac{\mathrm{tr}\mathcal{J}^T\mathcal{J}}{2(\det\mathcal{J})^2}} \leq n_{\max} \text{ for } \mathbf{x} \in G \\ R = R_0 \end{cases} . \tag{2.14}$$

We use the circular bend $r = x$, $\theta = \frac{\pi y}{2L} = \frac{y}{R}$ as a starting guess, and search the space of general transformations $r(\mathbf{x})$, $\theta(\mathbf{x})$ by perturbing from this base case. (We

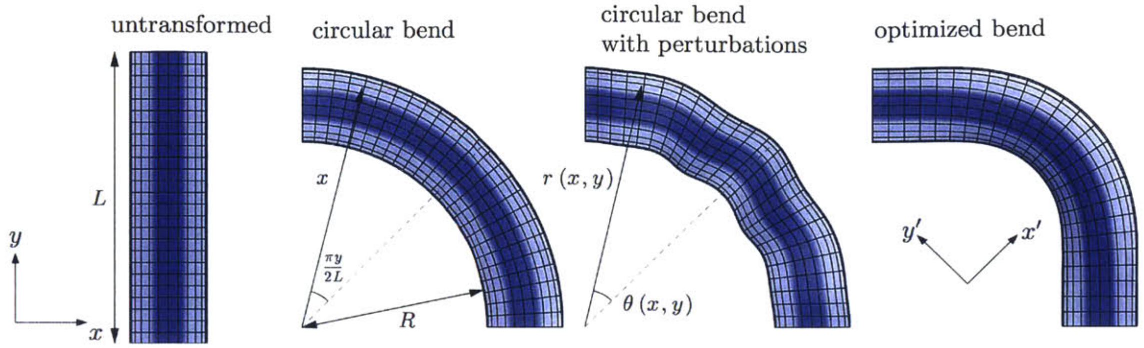


Figure 2-3: In the transformation process, the untransformed straight waveguide is bent, perturbed, and optimized. Darker regions indicate higher refractive index

only perform *local* optimization; not global optimization, but comment in Sec. 2.4.3 on a simple technique to avoid being trapped in poor local minima.) As explained in Sec. 2.3.4, the perturbations will be parameterized such that the symmetry and continuity constraints are satisfied automatically. Fig. 2-3 shows a schematic of the bend transformation optimization process. First, the straight region is mapped to a circular bend. Then, the intermediate polar coordinates r and θ for every point $\mathbf{x}$ are perturbed, using an optimization algorithm described at the end of Sec. 2.3.4, and the desired norm (either L_1 and L_∞) of the anisotropy is computed. This process is repeated at each optimization step until the structure converges to a local minimum in $\|\mathbb{K}\|$.

2.3.4 Spectral parameterization

To faciliate efficient computation of the objective and constraints, the functions r and θ can be written as the circular bend transformation plus perturbations parametrized in the spectral basis [91, 95]:

$$\begin{aligned} r\,(x,\,y) &= x + \sum_{\ell,\,m}^{N_\ell,\,N_m} C^{r}_{\ell m} T_\ell(2x - 2R - 1)\cos\frac{2m\pi y}{L} \\ \theta\,(x,\,y) &= \frac{\pi y}{2L} + \frac{1}{x}\sum_{\ell,\,m}^{N_\ell,\,N_m} C^{\theta}_{\ell m} T_\ell(2x - 2R - 1)\sin\frac{2m\pi y}{L}, \end{aligned} \tag{2.15}$$

where the coordinate $2x-2R-1$ has been centered appropriately for the domain $[-1,\ 1]$ of degree-ℓ Chebyshev polynomials T_ℓ. The sines and cosines have been chosen to satisfy the mirror-symmetry conditions of Eq. 2.11. The sine series also automatically satisfies the second continuity condition of Eq. 2.12. In order to satisfy the rest of the conditions, the following constraints are also imposed:

$$
\begin{aligned}
&\sum_m^{N_m} C^r_{\ell m} \left(-1\right)^m = 0 \\
&\sum_m^{N_m} C^\theta_{\ell m} \left(-1\right)^m m = \begin{cases} \frac{L}{8\pi} - \frac{R}{4} - \frac{1}{8}, & \ell = 0 \\ -\frac{1}{8}, & \ell = 1 \\ 0, & \ell \geq 2 \end{cases} .
\end{aligned}
\tag{2.16}
$$

These equations are solved to simply eliminate the $C^{r,\theta}_{\ell N_m}$ coefficients before optimization. Many nonlinear optimization algorithms allow both inequality constraints and equality constraints. In principle, an explicit equality constraint forcing the Jacobian to be continuous is also possible. In practice, this method does not work nearly as well as eliminating Fourier coefficients, as we have done here. The reason is that the region satisfying the equality constraint is usually a very small and disconnected region embedded in a much larger space of optimization parameters. Imposing an equality constraint does not strictly force the algorithm to stay within the subregion; there is often a penalty function associated with violating the constraint and consequently some straying away from the region is allowed. This makes convergence much slower than simply eliminating some of the degrees of freedom.

This spectral parametrization has several advantages over finite-element discretizations such as the piecewise-linear parameterization of Ref. 164. First, the spectral basis converges exponentially for smooth functions [95]. We found that only a small number ($N_\ell \times N_m < 100$) of spectral coefficients $C^{r,\theta}$ are needed to achieve very low-anisotropy ($\mathbb{K} - 1 \approx 10^{-4}$) transformations. Second, if the fabrication process favors slowly varying transformations (or if these are needed to make the eikonal approximation for the TE polarization, as in Sec. 2.2.2), this constraint may be imposed simply by using smaller N_ℓ and N_m.

With this spectral parameterization, the formulation of the optimization problem (Eq. 2.13) becomes

$$\min_{\{C^{r,\theta}_{\ell m}\}, n_0, L, t} t \, \text{subject to}: \begin{cases} \text{constraint } 2.16 \\ n_{\min} \leq n_0 p\,(x) \sqrt{\frac{\text{tr}\mathcal{J}^T\mathcal{J}}{2(\det \mathcal{J})^2}} \leq n_{\max} \text{ for } \mathbf{x} \in G \\ R = R_0 \\ \mathbb{K}(\mathbf{x}) \leq t \text{ for } \mathbf{x} \in G \end{cases} . \quad (2.17)$$

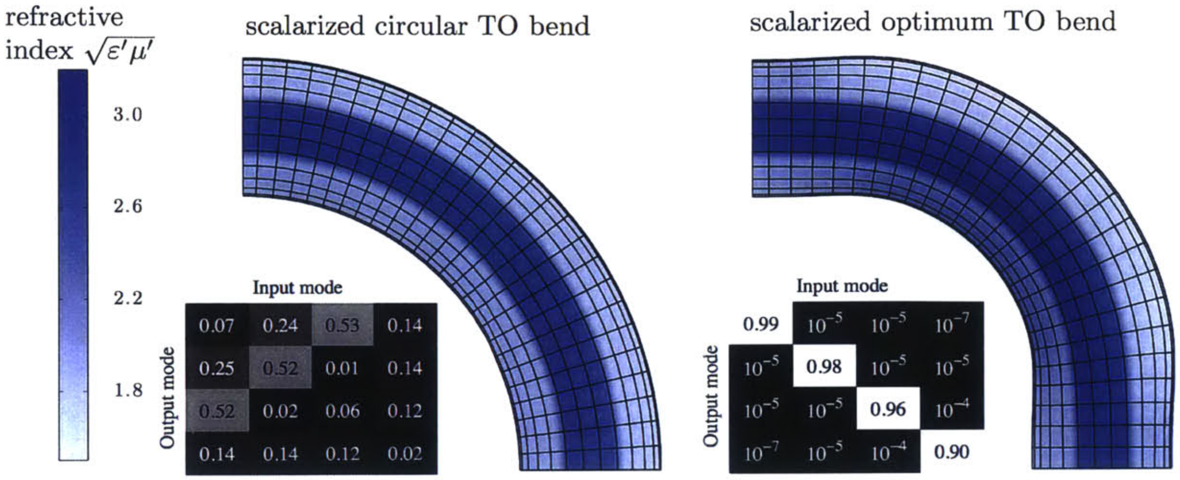
The local optimization was performed using the derivative-free COBYLA non-linear optimization algorithm [93, 179] in the NLopt package [94]. In principle, we can make the optimization faster by analytically computing the derivatives of the objective and constraints with respect to the design parameters and using a gradient-based optimization algorithm, but that is not necessary because both $\text{tr}\mathcal{J}^T\mathcal{J}$ and $\det \mathcal{J}$, which determine all the non-trivial objective and constraint functions in this optimization problem, are so computationally inexpensive to evaluate that the convergence rate is not a practical concern.

2.4 Optimization results

2.4.1 Minimal peak anisotropy

A min $||\mathbb{K}||_\infty$ design is shown in Fig. 2-4, along with the scalarized circular TO bend for comparison. The bend radius was $R = 2$ and the number of spectral coefficients was $N_\ell = 5$, $N_m = 8$. The objective and constraints were evaluated on a 100×140 grid G in $\mathbf{x}$ (Chebyshev points in the x direction and a uniform grid in the y direction). This design had $\max_\mathbf{x} \mathbb{K} - 1 \approx 5 \times 10^{-4}$ and mean $\langle \mathbb{K} \rangle - 1 \approx 10^{-4}$. In comparison, the circular TO bend of the same radius has $\max_\mathbf{x} \mathbb{K} - 1 \approx 0.1$ and $\langle \mathbb{K} \rangle - 1 \approx 10^{-2}$.

The $R = 2$ optimized design structure was compared in finite-element Maxwell simulations (using the FEniCS code [39]), with the conventional non-TO bend [simply



0.07	0.24	0.53	0.14
0.25	0.52	0.01	0.14
0.52	0.02	0.06	0.12
0.14	0.14	0.12	0.02

0.99	10^{-5}	10^{-5}	10^{-7}
10^{-5}	0.98	10^{-5}	10^{-5}
10^{-5}	10^{-5}	0.96	10^{-4}
10^{-7}	10^{-5}	10^{-4}	0.90

Figure 2-4: Optimization decreases anisotropy by a factor of 10^{-4}, while dramatically improving the scattered-power matrix.

bending the waveguide profile around a circular arc with $n'(\mathbf{x}') = n(\mathbf{x})$] and the scalarized circular TO bend. The four lowest-frequency modes of a multimode straight waveguide were injected at the input interface $y = \frac{L}{2}$, and the scattered-power matrix T was computed using the measured fields at the output interface $y = -\frac{L}{2}$. The scattered-power matrix is defined as

$$T_{ij} = \left| \int_{R}^{R+1} \mathrm{d}x\, \hat{\theta} \cdot \left(\mathbf{E}_j^0 \times \mathbf{H}_i\right) \right|^2_{y=-\frac{L}{2}},$$

where $-\hat{\theta}$ is the propagation direction of the guided modes, $\mathbf{E}_j^0$ is the normalized electric field of the jth exactly guided mode of the non-scalarized material ($\boldsymbol{\mu}'$, ε'), and $\mathbf{H}_i$ is the actual magnetic field of the approximate scalarized material at the interface after injecting a normalized mode $\mathbf{E}_i^0$ at the input interface. This makes T_{ij} equal to the power scattered into the jth output mode from the ith input mode. For a straight waveguide, which has no intermodal scattering, $T = \mathbb{I}$. Fig. 2-4 shows a dramatically improved T for the scalarized and optimized TO bend compared to the scalarized circular TO bend. [The rows and columns of T for the circular bend add up to less than one because some power has either been scattered out of the waveguide entirely, or some power has been scattered into fifth or higher-order modes. The rows

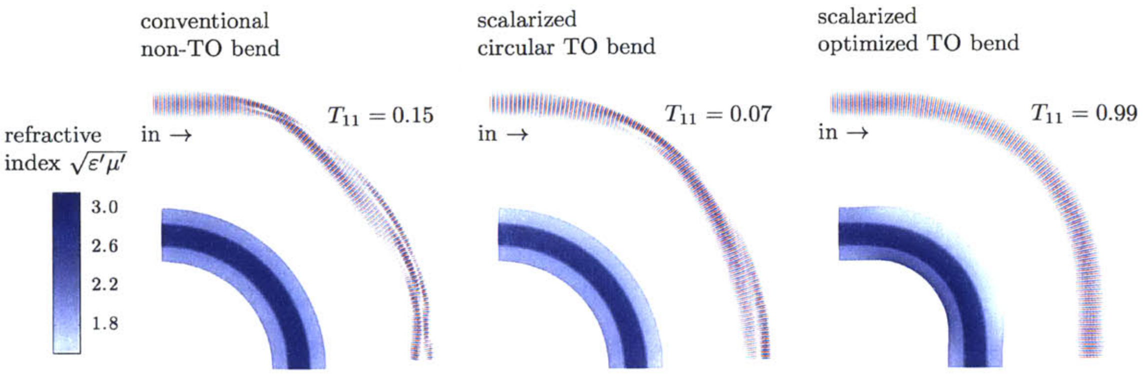


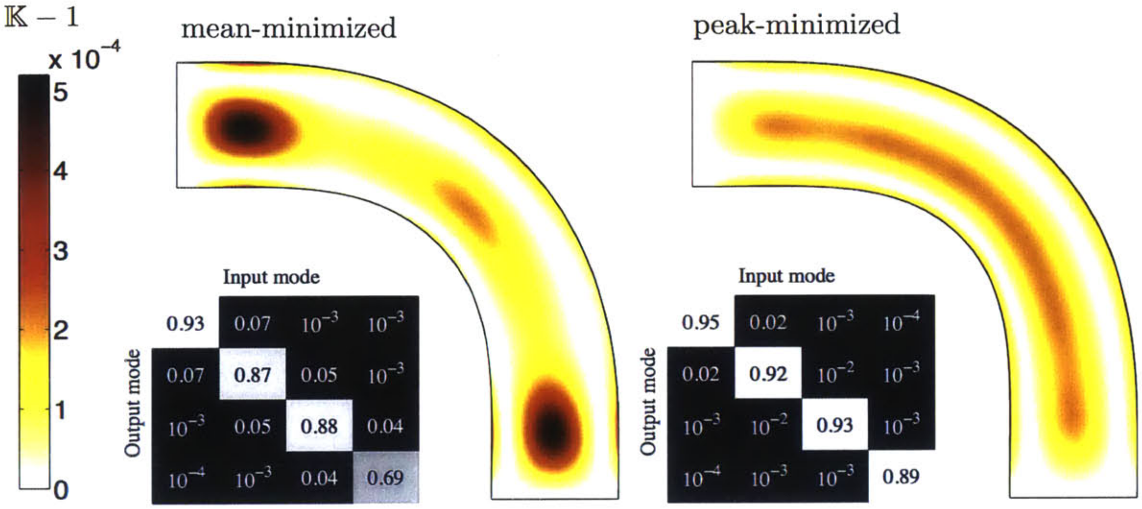
Figure 2-5: FEM field profiles show heavy scattering in the conventional non-TO and scalarized circular bends, but very little scattering in the optimized bend.

and columns of T for the optimized bend add up to nearly 1, with the small deficiency due to the $\mathcal{O}(\mathbb{K}-1)$ out-of-bend and higher-order intermodal scattering as well as mesh-descretization error.

The electric-field profiles for the fundamental mode, displayed in Fig. 2-5, show a dramatic difference in the performance of the optimized structure versus the other structures. Both the conventional and circular TO bend show heavy intermodal scattering in the bend region, while the optimized transformation displays very little scattering.

2.4.2 Minimizing max versus minimizing mean

We found a clear difference between minimizing the peak anisotropy versus minimizing the mean. The results of an optimization run with $R = 2.5$, $N_\ell = 3$, and $N_m = 6$ are shown in Fig. 2-6. Both structures had very low mean anisotropy $\langle\mathbb{K}\rangle_{\mathbf{x}} - 1$. The mean-minimized structure, at $\langle\mathbb{K}\rangle - 1 \approx 10^{-5}$, had a slightly lower mean than the peak-minimized structure which had $\langle\mathbb{K}\rangle - 1 \approx 1.5 \times 10^{-5}$. However, in terms of the peak anisotropy, the peak-minimized structure is the clear winner by a factor of 2.5, with $\max_{\mathbf{x}} \mathbb{K} - 1 \approx 2 \times 10^{-4}$ as opposed to $\max_{\mathbf{x}} \mathbb{K} - 1 \approx 5 \times 10^{-4}$ for the mean-optimized structure. Both structures were scalarized and tested in finite-element Maxwell simulations of the four lowest-frequency modes of the straight waveguide. The scattered-power matrix shows that the difference in $\max_{\mathbf{x}} \mathbb{K}$ resulted in an order



0.93	0.07	10^{-3}	10^{-3}
0.07	0.87	0.05	10^{-3}
10^{-3}	0.05	0.88	0.04
10^{-4}	10^{-3}	0.04	0.69

0.95	0.02	10^{-3}	10^{-4}
0.02	0.92	10^{-2}	10^{-3}
10^{-3}	10^{-2}	0.93	10^{-3}
10^{-4}	10^{-3}	10^{-3}	0.89

Figure 2-6: Anisotropy profile and scattered-power matrices for optimized designs that minimize the mean and the peak, with $R = 2.5$, $N_\ell = 3$, and $N_m = 6$.

of magnitude reduction in the intermodal scattering (as shown in the off-diagonal elements) and noticeably improved transmission, (especially in the element $T_{44} = 0.89$ for the fourth mode).

2.4.3 Tradeoff between anisotropy and radius

In optimized structures, we found that $\max_{\mathbf{x}} \mathbb{K}$ for the optimized bend, similar to the circular TO bend, decreases monotonically with R (as shown in Fig. 2-7). Unlike the circular bend, however, this tradeoff seems asymptotically *exponential* rather than $\mathcal{O}(R^{-2})$. In particular, there are two clearly different regimes for this tradeoff: a power law $\mathbb{K} - 1 \sim R^{-4}$ at small $R \lesssim 3$ and an exponential decay $\mathbb{K} - 1 \sim \exp(-0.34R)$, at larger R. The second regime was only attained after using successive optimization, because with only one independent optimization run the algorithm tended to get stuck in local minima. For successive optimization, the optimum structure is used as a starting guess for the next run, and the initial step size is set large enough so that the algorithm can reach better local minima than the previous one.

For $R \lesssim 3$, we found that there are multiple local minima and that independent optimizations for different R tend to be trapped in suboptimal local minima, as shown

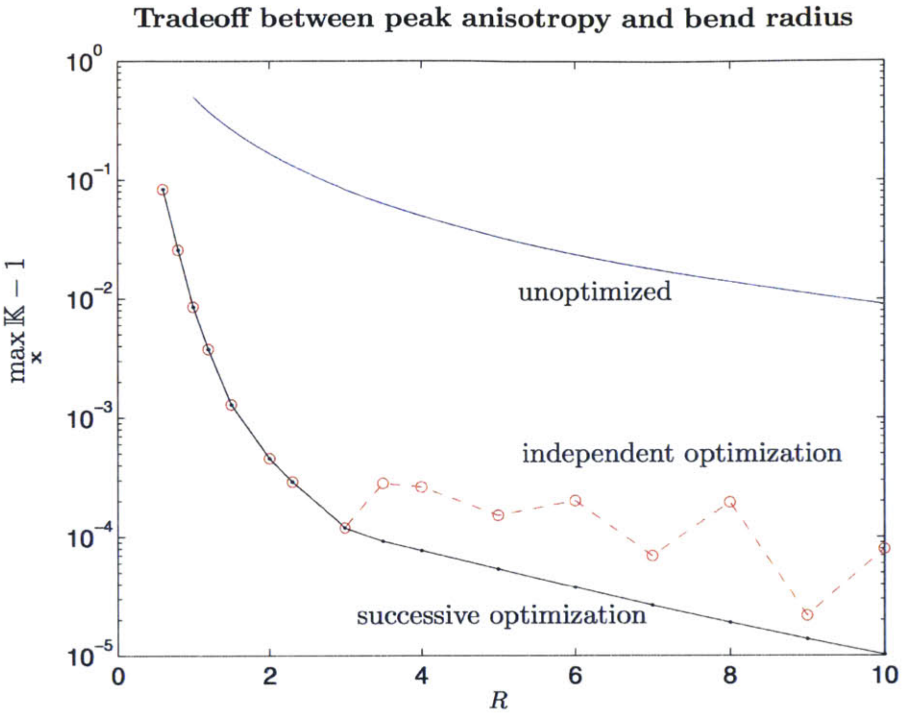


Figure 2-7: Successive optimization with $N_\ell = 5$, $N_m = 8$ results in a power law decaying tradeoff $\max_{\mathbf{x}} \mathbb{K} - 1 \sim R^{-4}$ at low R and an exponentially decaying tradeoff at higher R. For comparison, the unoptimized anisotropy for the circular TO bend is shown above.

by the open dots in Fig. 2-7. To avoid this problem, we used a "successive optimization" technique in which the optimal structure for smaller R is rescaled as the starting guess for local optima at a larger R, in order to stay along the exponential-tradeoff curve. (Another possible heuristic is "successive refinement" [180–183], in which optima for smaller $N_{\ell,m}$ are used as starting points for optimizing using larger $N_{\ell,m}$.)

2.5 Experimental realization

In this section, we present experimental results published in Ref. [23]. The design is based on the curving of a virtual space containing the straight multimode waveguide into a 90° bend, such that light travels along the curve as it would on the original straight waveguide, i.e., with minimal inter-mode coupling. It is important to observe that not only the modal distribution is preserved throughout the bend, but so is the phase relationship between modes, a key factor for any phase-sensitive application.

The modal superposition excited in multimode bends is exemplified in fig. 2-8, where we show the field distribution of a 4μm wide waveguide (supporting 16 modes) excited by the fundamental mode at $\lambda = 1.55\mu$m and bent with 78.8μm radius. One can see that the bending of the waveguide leads to a superposition of many higher modes at the output, which in turn introduce penalties to the bandwidth of this channel. In the 4μm waveguide shown in fig. 2-8, the limit in bandwidth-distance product that can be supported due only to difference in group velocities between the fundamental and second order mode is below 16 Gb·m/s (for complete symbol overlap), between the fundamental and third order mode is 6 Gb·m/s, and so on. Therefore, any device that couples a significant amount of power into these modes, such as the circular bend simulated here (71% of the power input in the fundamental mode is coupled to the second order mode and 23% to the third), will severely limit the communication data rate of the system. One way to minimize the mode mixing would be to use very large bending radii. For the 4μm waveguide, in order to ensure that 95% of power is coupled back to the fundamental mode of the straight waveguide, a bending radius of more than 1 mm is necessary, which is unacceptably large for photonics integration.

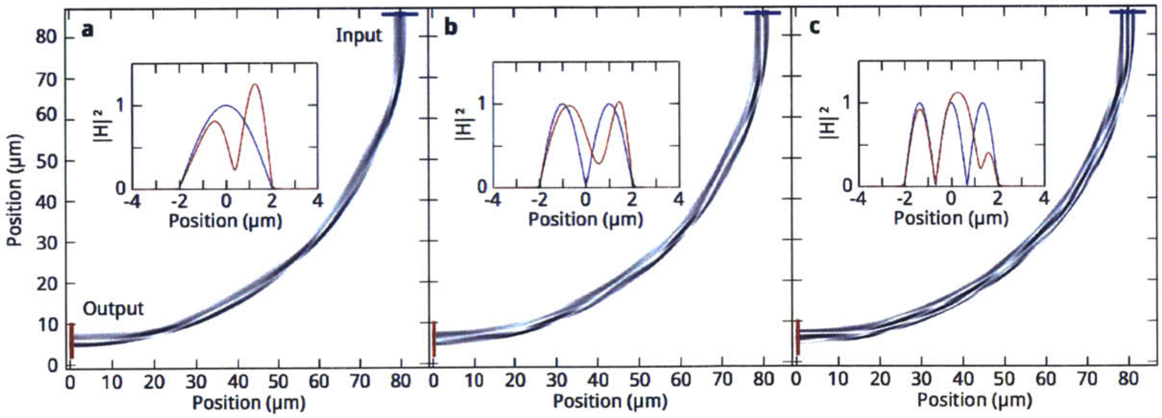


Figure 2-8: Finite element simulations of a multimode bend. The figure shows the magnetic field magnitude squared ($|H|^2$) for a conventional multimode bend when excited with the first three modes of the input multimode waveguide (a to c, respectively). The input modes (blue cross-sections, on the upper-right endpoints) are coupled to many other modes, as evidenced by the cross-section plots at the outputs (red, on the lower left endpoints). The waveguides are 4μm wide and the bends have 78.8μm radius. The simulations were performed using the FEniCS solver [39].

2.5.1 Design challenges

There are two main challenges in designing a waveguide bend via TO in a silicon photonics platform, in contrast to arbitrary metamaterials [62, 87, 123, 126]. The first design challenge is the need to employ only isotropic materials with refractive indices limited to values between the index of Si and the index of SiO_2 (roughly $1.5 < n < 3.5$). Although in principle anisotropic materials, such as SiC, can be used in silicon photonics, TO requires in general a gradient anisotropy, which is much more complex to achieve. Note that one could also use photonic crystals [56, 71, 80, 156] or grooves [73, 157] to tailor the index and anisotropy, however these discrete structures result in additional field scattering and inter-mode coupling due to the finite wavelength/period ratio (typically around 10:1). Instead, we use a grayscale-lithography technique described below that produces smooth gradients but is limited to mostly-isotropic effective indices. The second design challenge is to match the refractive index and the geometry of the device at its end facets to the multimode waveguides connected to it. If the connection is not matched, each mode from the straight waveguide will couple to many modes in the bend, similarly to the case of a conventional bend (fig. 2-8), immediately introducing inter-mode coupling, even if the bend itself doesn't. This means that the bend transformation must smoothly transition from the same shape of the input waveguides at its end facets to the required curve in its interior.

2.5.2 Final design and fabrication

The optimized multimode bend obtained has an effective radius of curvature of 19.7 times the width of the waveguide. Using a 4μm wide waveguide then results in a radius of 78.8μm, which we also used in the simulations of fig. 2-8 for comparison. In fig. 2-9 we show the first three propagating modes of the input waveguide traveling almost undisturbed through the optimized bend. The optimized bend's effective-index profile can be seen in fig. 2-10a.

The fabrication of this multimode bend is achieved using grayscale e-beam lithog-

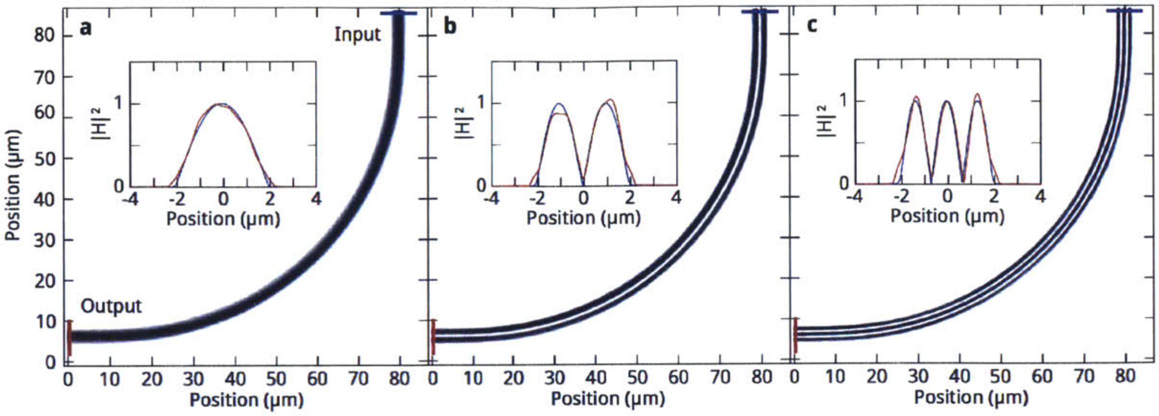


Figure 2-9: Finite element simulations of our optimized TO multimode multimode bend. The figure shows the magnetic field magnitude squared ($|H|^2$) for the cases when the bend is excited with the first three modes of the input multimode waveguide (a to c, respectively). The input modes (blue cross-sections, on the upper-right endpoints) are preserved throughout the bends, showing minimal inter-mode coupling at the outputs (red, on the lower left endpoints). The waveguides are 4μm wide and the bends have 78.8μm radius.

raphy, on a silicon-on-insulator (SOI) wafer with 3μm buried SiO_2 and 500 nm Si layer. We create the required non-uniform refractive index medium using the effective propagation index for our vertical slab structure, composed by the buried SiO_2 layer, the guiding Si layer and a cladding layer of SiO_2 deposited via plasma-enhanced chemical vapor deposition (PECVD). The effective propagation index of this structure is controlled by the thickness of the Si layer [59, 60, 159, 184, 185], such that the index map from the TO optimization (fig. 2-10a) is translated into a thickness map to be fabricated via grayscale lithography (fig. 2-10b). The grayscale lithography is achieved via dose modulation for patterning the photonic device with vertical resolution of approximately 10 nm. Note that while similar processes are employed in the fabrication of diffractive optical elements, micro-electro-mechanical structures, and lower contrast graded-index lenses [76–79, 186] with relatively weak height variations of 80 nm over distances of tens of microns, in our case the process enables strong height variations of 400 nm over less than 1μm while maintaining precise control of both the resist height profile on the nanometer scale. Figure 2-10 shows the grayscale patterned device with a smooth surface profile in Si.

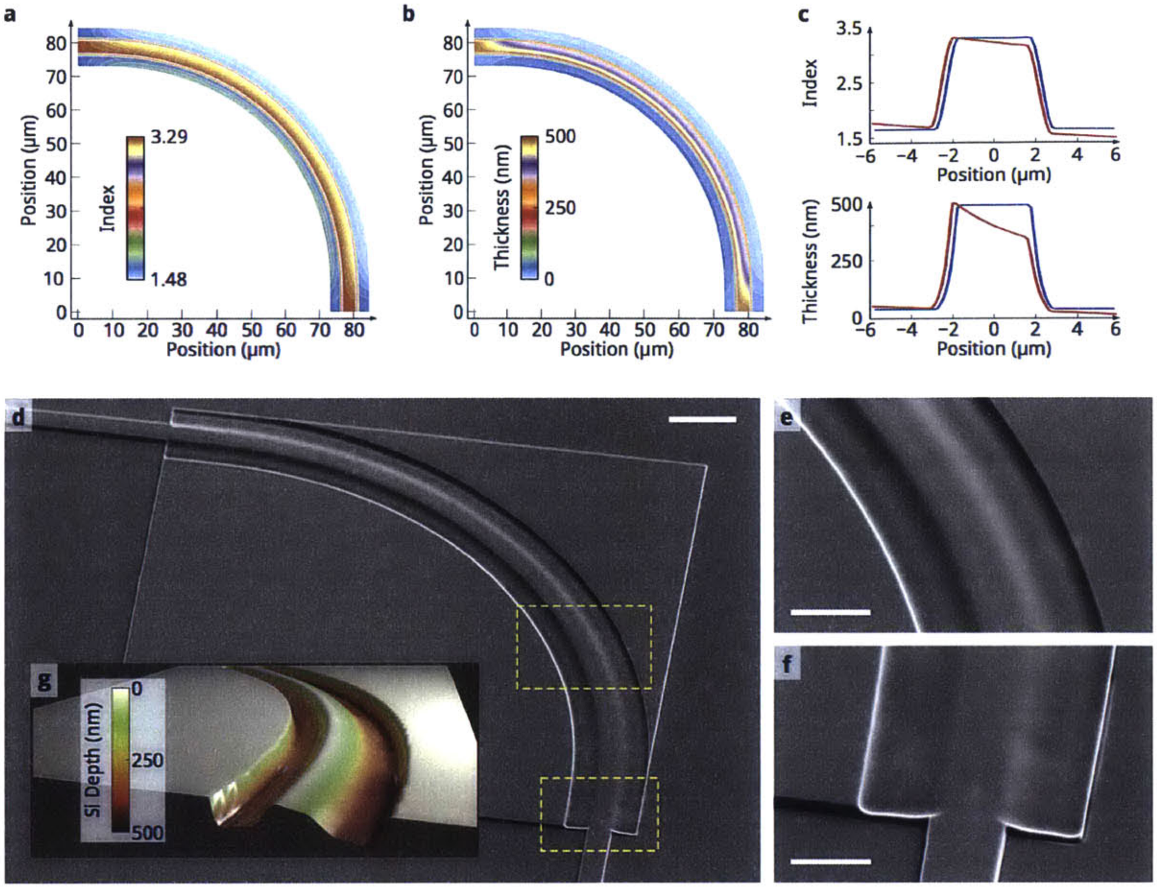


Figure 2-10: TO design and fabrication. Optimized refractive index profile (a) for the multimode bend and respective silicon layer thickness (b) to implement the bend. (c) Cross-sections of the refractive index and thickness of the profiles at the endpoints (blue) and at the center of the bend (red). (d) Scanning electron microscope images of the fabricated graded-index bend (10μm scale). The smoothness obtained by our grayscale process can be seen in (e) the close-ups of the bend interior (5μm scale), and (f) the connection with a conventional multimode waveguide at the output (4μm scale). (g) Atomic force microscope scan of a fabricated bend, showing the thickness profile in the silicon layer.

2.5.3 Characterization

To evaluate the performance of our mode-preserving multimode bend, we compare it to a circular multimode bend with rectangular profile and with same radius as our device (78.8μm). We measure the transmission coefficient of the fundamental mode through the bend, so that the presence of inter-mode coupling is evidenced by low transmitted power in the system. To ensure we are exciting and collecting only the fundamental waveguide mode, we couple light into our sample via a grating connected to a single mode waveguide, which, in turn, tapers very slowly (over $250\mu m$) to a $4\mu m$ wide multimode waveguide. This waveguide connects to the bend and the output is once again slowly tapered down to a single mode waveguide, radiating away any higher order modes that might have been excited along the bend. Light from the single mode waveguide is coupled via another grating to a fiber which in turn is connected to detector and a power meter. An optical microscope image of this system is shown in fig. 2-11a. The results are displayed in the plots of fig. 2-11b, where we show the histograms for measurements of 25 devices designed via TO and 11 conventional multimode bends.

The data presented in fig. 2-11b clearly show the vast improvement in transmission from our TO multimode bends with respect to the conventional ones, a direct result of the mode-preserving characteristic of our design. Moreover, the 2-dimesional simulations in figures 2-8 and 2-9 show a difference in transmission for the fundamental mode of 13.6 dB, closely agreeing with the experimental results.

It is also important to analyze the performance of our multimode bend against a conventional single mode one to evaluate how much the grayscale fabrication impacts in the total losses in the link. Measurements of 11 single mode waveguide bends on the same sample showed an averaged normalized transmission coefficient of -2.6 dB, very similar to our TO design (-2.5 dB). These numbers enforce our conclusions of minimal inter-mode coupling in the optimized bend and indicate that any additional losses introduced in the grayscale process are compensated by the naturally lower losses found in multimode waveguides (due to less interaction of the fields with the core interfaces). We note that the variations in the transmission seen in the measurements of the TO bend are observed in the single mode devices as well, indicating that the variations are due to fabrication steps common to all devices, which might have introduced impurities in the sample, and not from the grayscale process itself.

Thus we demonstrate an optimized design and fabrication process for a multimode photonic platform with very low inter-mode coupling. This platform can be used to enable multimode photonics while also pointing towards the possibility of developments in mode-multiplexing [187–191] for ultra-high bandwidth communications.

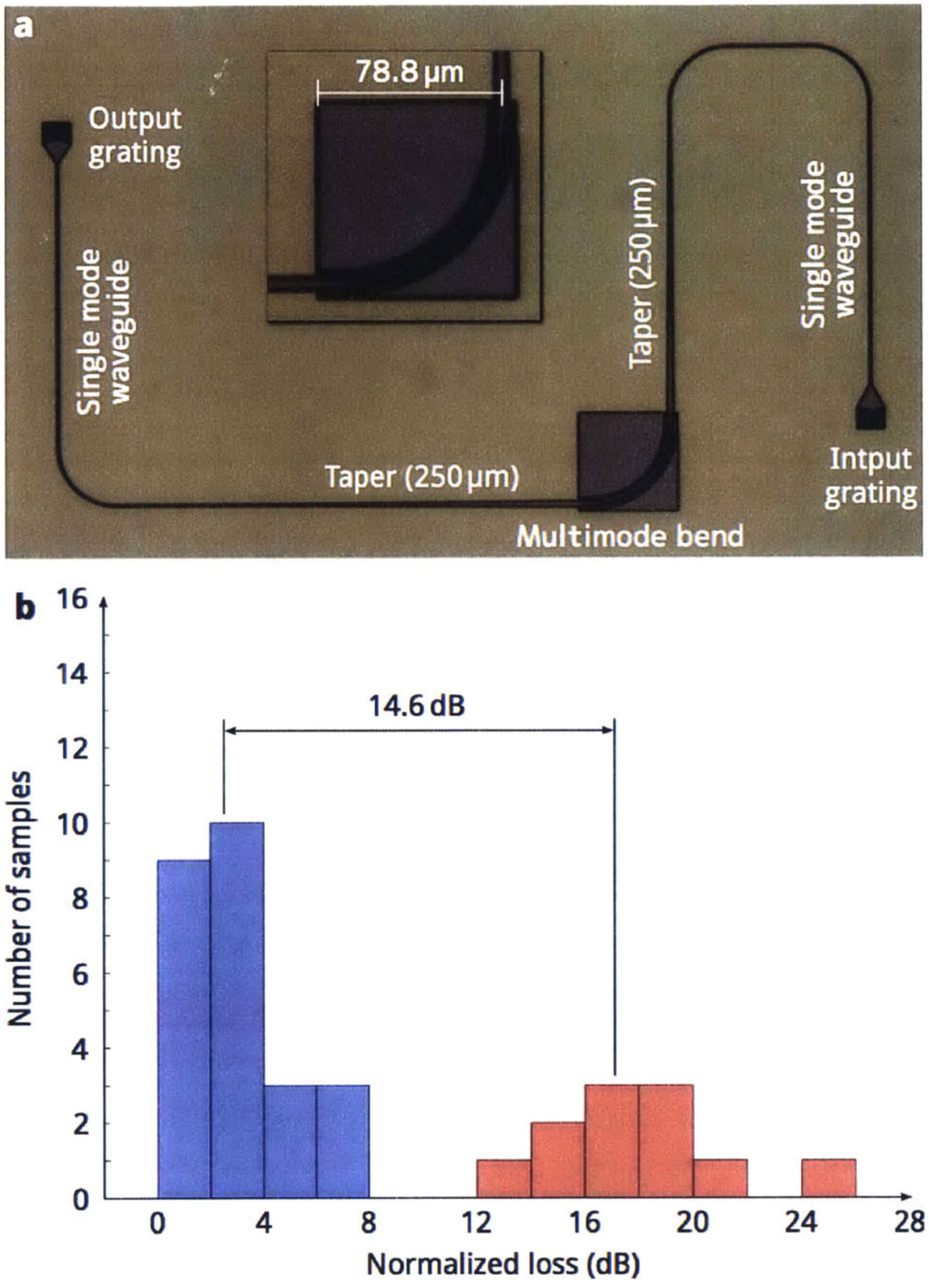


Figure 2-11: Multimode bend experiment. (a) Optical microscope image of a tested device. Due to the large length of the tapers, only the fundamental mode is excited at the multimode bend input. Conversely, higher order modes excited along the bend are radiated by the output taper, such that the power measured at the output grating reflects how well the fundamental mode is preserved by the bend. (b) Histograms of the measurements from our multimode bend design (blue) and a conventional multimode bend with rectangular cross-section (red) with same radius. There is a 14.6 dB improvement in the average transmission coefficient for the fundamental mode of the optimized bend with respect to the conventional one.

2.6 Mode squeezer

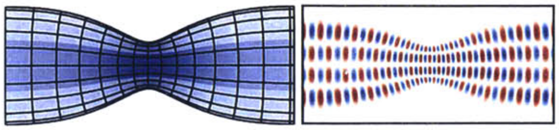
We also applied transformation inverse design to another interesting geometry: a mode squeezer that concentrates modes and their power in a small region in space, again with minimal intermodal scattering (quite unlike a conventional lens, which is intrinsically angle/mode-dependent), similar to the problem considered in Ref. 127 (which did not construct isotropic designs). We choose the untransformed region to be $-1 \leq x \leq 1$ and $0 \leq y \leq L$. The goal of this transformation $\mathbf{x}'(\mathbf{x})$ is to focus the beam by minimizing the *mid-beam width*

$$W \equiv \int_{-1}^{1} \mathrm{d}x \left. \sqrt{\left(\frac{\partial x'}{\partial x}\right)^2 + \left(\frac{\partial y'}{\partial x}\right)^2} \right|_{y=\frac{L}{2}} .$$

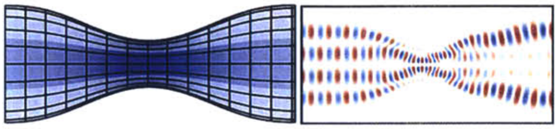
As in Sec. 2.3.4, the transformation is written as a perturbation from the identity transformation (which was used as the starting guess) and parameterized in the spectral basis

$$\begin{aligned} x'(x,\, y) &= x + \sum_{\ell,\, m}^{N_\ell,\, N_m} C^x_{\ell m} T_\ell(x) \, \sin \frac{(2m+1)\, \pi y}{L} \\ y'(x,\, y) &= y + \sum_{\ell,\, m}^{N_\ell,\, N_m} C^y_{\ell m} T_\ell(x) \, \sin \frac{(2m+1)\, \pi y}{L}, \end{aligned}$$

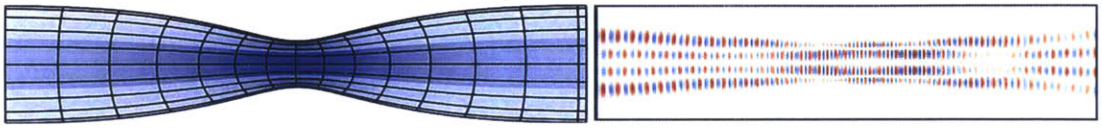
The sine series automatically satisfies mirror symmetry about $y = \frac{L}{2}$ and continuity of $\mathbf{x}'$ at the input/output interfaces $y = 0,\, L$. However, we found that constraining the coefficients $C^{x,y}$ to enforce continuity of $\mathcal{J}$ (as in Sec. 2.3.4) was not necessary (although it might give a better result) since the optimization algorithm only squeezed the center region while leaving the interfaces and the regions around them relatively untouched. In this problem, we could either minimize $\mathbb{K}$ for a fixed W or minimize W for a fixed $\mathbb{K}$, and we happened to choose the latter.



(a) optimized transformation, $T_{44} = 0.99$



(b) gaussian taper, $T_{44} = 0.63$



(c) optimized, then stretched transformation, $T_{44} = 0.92$

Figure 2-12: Optimized squeezer outperforms gaussian taper and stretched optimized squeezers in finite element simulations.

Finite-element Maxwell simulations, shown in Fig. 2-12, demonstrate that the optimized design is greatly superior to a simple Gaussian taper transformation designed by hand. The Gaussian transformation was given by $x'(\mathbf{x}) = x - x\alpha \exp[-\beta(y - \frac{L}{2})^2]$, where $\beta > 0$ and $0 < \alpha < 1$. Superficially, the design seems similar to an "adiabatic" taper between a wide low-index waveguide and a narrow high-index waveguide, and it is known that any sufficiently gradual taper of this form would have low scattering due to the adiabatic theorem [192]. However, the optimized TO design is much too short to be in this adiabatic regime. If it were in the adiabatic regime, then taking the same design and simply stretching the index profile to be more gradual (a taper twice as long) would reduce the scattering, but in Fig. 2-12 we perform precisely this experiment and find that the stretched design *increases* the scattering.

2.7 Concluding remarks

The analytical simplicity of TO design—no Maxwell equations need be solved in order to warp light in a prescribed way—paradoxically makes the application of computational techniques more attractive in order to discover the best transformation by rapidly searching a large space of possibilities. Previous work on TO design used optimization to some extent, but overconstrained the transformation by fixing the boundary shape while underconstraining the boundary conditions required for reflectionless interfaces. In fact, even our present work imposes more constraints than are strictly necessary—as long as we require continuous $\mathcal{J}$ at the input/output interfaces, there is no conceptual reason why those interfaces need be flat. A better bend, for example, might be designed by constraining the location of only two corners (to fix the bend radius) and constraining only $\mathcal{J}$ on other parts of the endfacets. However, we already achieve an exponential tradeoff between radius and anisotropy, so we suspect that further relaxing the constraints would only gain a small constant factor rather than yielding an asymptotically faster tradeoff. In the case of the mode squeezer, one could certainly achieve better results by imposing the proper $\mathcal{J}$ constraints at the endfacets. It would also be interesting to apply similar techniques to ground-plane cloaking.

All TO techniques suffer from some limitations that should be kept in mind. First, TO seems poorly suited for optical devices in which one *wants* to discriminate between modes (e.g. a modal filter) or to scatter light between modes (e.g. a mode transformer). TO is ideal for devices in which it is desirable that all modes be transported equally, with no scattering. Even for the latter case (such as our multimode bend), however, TO designs almost certainly trade off computational convenience for optimality, because they impose a stronger constraint than is strictly required: TO is restricted to designs where the solutions at *all points* in the design are coordinate transformations of the original system, whereas most devices are only concerned with the solutions at the endfacets. For example, it is conceivable that a more compact multimode bend could be designed by allowing intermodal scattering *within* the bend as long as the modes scatter back to their original configurations by the endfacet; the interior of the bend might not even be a waveguide, and instead might be a resonant cavity of some sort [9, 148, 149, 193]. However, optimizing over such structures seems to require solving Maxwell's equations in some form at each optimization step, which is far more computationally expensive than the TO design and, unlike the TO design, must be repeated for different wavelengths and waveguide designs.

Chapter 3

Scalable numerical approach for the steady-state *ab initio* laser theory

3.1 Overview

As lasers become increasingly complicated, especially in nanophotonic systems with wavelength-scale features [109, 194–196], there has been a corresponding increase in the computational difficulty of solving for their nonlinear behavior, as described by the Maxwell–Bloch (MB) equations [33]. To address this key challenge in the design and understanding of lasers, a highly efficient approach to finding the non-linear steady-state properties of complex laser systems has recently been introduced, known by the acronym SALT (steady-state *ab initio* laser theory). In this chapter, which was published in Ref. [197], we present a technique to *directly* solve the SALT formulation [26, 28, 30] of the steady-state MB equations (using finite-difference frequency-domain (FDFD) [38, 110] or finite-element methods (FEM) [36]), and we demonstrate that, unlike previous approaches to the SALT equations [26, 30], our technique scales to full three-dimensional (3D) low-symmetry geometries (such as photonic-crystal slabs [34]).

The SALT equations (reviewed in Sec. 3.2) simplify the general MB equations by removing the time dependence for steady-state modes, which allows SALT solvers to be potentially far more efficient than previous time-domain approaches [101, 102], while providing comparable accuracy [24, 25]. However, all earlier approaches to SALT required the intermediate construction of a specialized constant-flux (CF) basis for the laser modes. While efficient and yielding numerous insights in highly symmetric geometries where it can be constructed semi-analytically, the CF basis becomes unwieldy and numerically expensive for complex low-symmetry laser geometries, especially in three dimensions. In our approach, we solve the SALT equations *directly* as a set of coupled nonlinear partial differential equations (PDEs), using a combination of Newton-Raphson [1], sparse-matrix solver [113], and nonlinear eigenproblem [104] algorithms in standard FDFD or FEM discretizations. In Sec. 3.4, we validate our solver against previous CF solutions for one-dimensional (1D) and cylindrical systems, while demonstrating that even in one dimension the CF basis rapidly becomes large and expensive as the system is brought farther above threshold. Furthermore, we show in Sec. 3.3.4 that analytical outgoing-radiation boundary conditions, which are difficult to generalize to three dimensions [37], can be substituted by the standard PML (perfectly matched layer) method [37, 38, 48] which is equally effective at modeling open systems. We also demonstrate multi-mode laser solutions in Sec. 3.4.2).

We conclude in Sec. 3.4.2 with full 3D vectorial laser-mode solutions for a photonic-crystal slab microcavity [34]. The appendixes provide further details on the computational techniques we use in this chapter, but in general any standard computational method in electromagnetism could be combined with our nonlinear solver algorithms. We believe that this computational approach provides a powerful tool to design and explore laser phenomena in the complex geometries accessible to modern nanofabrication, which were previously intractable for accurate modeling.

The Maxwell-Bloch (MB) equations provide the most basic formulation of semiclassical laser theory. The propagation of the electromagnetic field is given by the classical Maxwell equations and only the interaction of the field with the gain medium, represented by an ensemble of two-level atoms embedded in a cavity or background linear medium, is treated quantum mechanically. The MB equations are a set of time-dependent coupled nonlinear equations that are typically hard to solve analytically, except by using many approximations and idealizations. In the generic case of laser systems where such approximations are not valid, the MB equations have typically been solved using numerically expensive time-domain simulations [101, 102]. For the case of steady-state lasing, as noted above, a much more efficient theory for calculating the multi-periodic solutions of the MB equations is the *steady-state ab-initio lasing theory* (SALT) [26–28]. This theory has proven to be a viable tool for describing laser systems ranging from random lasers [30, 198, 199] to coupled laser systems [200] and photonic-crystal lasers [201]. It makes no *a priori* assumptions about the geometry of the laser system, treats the open (non-Hermitian) character of the laser system exactly, and the non-linear hole-burning interactions between the laser modes to infinite order. More realistic and quantitative laser modeling typically requires treating a gain medium with three, four or more relevant atomic levels, but it has been shown that for the steady-state properties, under the same assumptions as SALT, the semiclassical equations can be reduced to an effective two-level (MB) system with renormalized parameters and solved with essentially the same efficiency as two-level SALT [24, 29]. SALT can also be used to describe quantum properties of lasers by combining the non-linear scattering matrix of SALT with input-output theory, leading specifically to a general formula for the linewidth of each mode in the non-linear steady-state [202].

For readers familiar with linear resonant cavities in photonics, which essentially trap light for a long time in a small volume, a laser can be semiclassically understood via the introduction of nonlinear gain (amplification) whose strength is determined by an input-energy "pump" [203]. As the pump strength is increased, one eventually reaches a "threshold" at which the gain balances the cavity loss and a steady-state real-frequency lasing ("active") mode comes into existence. A key element is that the gain is nonlinear: increasing the laser-mode amplitude depletes the excited states of the gain medium (via a "hole-burning" term in the gain), and so at a given pump strength above threshold there is a self-consistent stable laser amplitude. At higher pump strengths, however, this picture is complicated by the introduction of additional lasing modes, which interact nonlinearly and whose individual gains and losses are balanced simultaneously by the SALT equations. Also, while a linear "resonant mode" technically refers to a pole in the Green's function (or scattering matrix) at a complex frequency lying slightly below the real axis, a lasing mode can arise from *any* pole that is pushed up to the real axis by the gain, even poles that start out far from the real axis and do not resemble traditional resonant-cavity modes (for example, in random lasers [30]).

The three relevant physical quantities for a laser, which enter the MB equations are the electric field as well as the polarization and the inversion of the gain medium, all of which can depend on space and time. By making a multi-mode steady-state ansatz and further well-motivated approximations, this system is reduced to the non-linear SALT equations for the electric field of the lasing modes. Lasing is conceived of as the limit of an amplifying scattering process in which the input goes to zero, hence it will correspond to purely outgoing solutions with real frequency or, equivalently, to a pole in the relevant scattering matrix on the real axis. Until the external pump is strong enough for the gain to balance the loss there will be no solution of this type. However, when increasing the pump strength, non-trivial solutions appear at a sequence of thresholds and at different frequencies. The non-linear interaction between these solutions is through the spatial hole-burning and depletion of the gain medium: each lasing mode extracts energy from the pump in a space-dependent manner which in general makes it more difficult for subsequent modes to reach threshold, and also effectively changes the index of refraction of the gain medium.

A strategy for efficiently solving the SALT equations was introduced in [27, 30] and significantly extended in [26]. These existing methods can be viewed as a spectral integral-equation method [95]: they solve the nonlinear problem by first parametrizing each laser mode in terms of a specialized "spectral" basis, called the "constant-flux (CF) states", that solve a *linear* non-Hermitian Maxwell eigenproblem parametrized by its (unknown) real lasing frequency. Because the frequency is required to be real outside the cavity, the photon flux outside the laser cavity is conserved, unlike the well-known quasi-bound states of the system, which are also purely outgoing, but do not conserve flux. This basis is defined so that at the lasing threshold for each mode, where the non-linear hole-burning interaction term is zero, one member of the basis set *is* the lasing solution. Hence, by construction, the basis expansion for the SALT solution above but near threshold converges rapidly even when the non-linear terms are taken into account, and the SALT equations reduce to finding a relatively small number of expansion coefficients for each mode. In highly symmetric geometries such as 1D or cylindrical systems with uniform pumping, the CF states can be found semi-analytically in terms of known solutions of the Helmholtz equation in each homogeneous region (e.g., in terms of sinusoid or Bessel functions), and such a basis will typically converge exponentially quickly [95] to the SALT solutions. Furthermore, the CF basis can be used as a starting point for other analyses of laser systems, such as to identify the cause of mode suppression due to modal interactions [26, 30] and exceptional points [200, 204]. However, the CF basis also has some disadvantages for complex geometries or for lasers operating far above threshold where the nonlinearities are strong and the convergence is not so rapid. In complex geometries where Helmholtz solutions are not known analytically, the CF basis itself must be found numerically by a generic discretization (e.g., FDFD or FEM) for many real frequencies (since the lasing frequency is not known *a priori* above threshold) and for multiple CF eigenvalues at each frequency in order to ensure convergence. The lack of separable solutions in low-symmetry two-dimensional (2D) and 3D geometries also increases the number of basis functions that are required (in contrast to cylindrical systems, for example, where the solutions $\sim e^{im\phi}$ can be solved one m at a time). In three dimensions, where the discretization might have millions of points (e.g., on a $100 \times 100 \times 100$ grid), even storing a CF basis consisting of hundreds or thousands of modes becomes a challenge, not to mention the expense of computing this many 3D eigenfunctions numerically or of computing the resulting SALT equation terms. As a consequence, our approach in this chapter is to abandon the construction of the intermediate CF basis and instead to directly discretize and solve the nonlinear SALT PDEs. This approach enables us to solve even low-symmetry 3D systems, and greatly enhances the power of the SALT approach for modeling and for the design of realistic laser structures.

3.2 Review of SALT

The origin of the SALT equations are the MB equations, which nonlinearly couple an ensemble of two-level atoms with transition frequency k_a ($c = 1$) to the electric field [33, 96]:

$$-\nabla \times \nabla \times (\mathbf{E}^+) - \varepsilon_c \ddot{\mathbf{E}}^+ = \tfrac{1}{\varepsilon_0} \ddot{\mathbf{P}}^+, \tag{3.1}$$

$$\dot{\mathbf{P}}^+ = -i(k_a - i\gamma_\perp)\mathbf{P}^+ + \tfrac{g^2}{i\hbar}\mathbf{E}^+ D, \tag{3.2}$$

$$\dot{D} = \gamma_\parallel (D_0 - D) - \tfrac{2}{i\hbar}[\mathbf{E}^+ \cdot (\mathbf{P}^+)^* - \mathbf{P}^+ \cdot (\mathbf{E}^+)^*], \tag{3.3}$$

Here, $\mathbf{E}^+(\mathbf{x}, t)$ and $\mathbf{P}^+(\mathbf{x}, t)$ are the positive-frequency components of the electric field and polarization, respectively. The coupling to the negative-frequency components is neglected in terms of a rotating wave approximation (RWA) which is both very useful for simplifying the equations and very accurate under general conditions. Note that at no point did we or will we assume the standard slowly-varying envelope approximation, which, if used, reduces the accuracy of the MB solutions. The population inversion of the medium $D(\mathbf{x}, t)$ is given by $D_0(\mathbf{x}, d)$ in the absence of lasing, which is roughly proportional to the external pumping rate and thus generally referred to as the pump strength. One of the useful features of SALT is that this pump strength can have an arbitrary spatial profile in addition to a varying global amplitude, such that one can represent different experimental pumping protocols by evolving along a "pump trajectory" which we parametrize here by d, following Ref. [200]. Note that if there are gain atoms in unpumped regions of the laser, then the pump strength D_0 will be negative in these regions and thus the SALT equations will automatically take into account absorption due to unexcited gain atoms. $\gamma_\perp$ and $\gamma_\parallel$ are the relaxation rates of the polarization and inversion, respectively. The linear cavity dielectric function $\varepsilon_c(\mathbf{x})$ is homogeneous outside the cavity region, and consequently a finite spatial domain can be used for the laser system with an outgoing boundary condition. We have assumed a scalar $\varepsilon_c(\mathbf{x})$ and dipole matrix element g, although in anisotropic gain media they can be generalized to tensors.

The attractive feature of SALT is that it provides access to the spatial profiles of the lasing modes as well as to the lasing frequencies of a multi-mode microlaser at very low computational costs. To achieve this high performance, SALT makes two essential assumptions. First, it assumes that for a fixed pump strength the electric field and polarization eventually reach a multi-periodic steady state,

$$\mathbf{E}^{+}(\mathbf{x},t) = \sum_{\mu=1}^{M} \mathbf{\Psi}_{\mu}(\mathbf{x}) e^{-ik_{\mu}t}, \tag{3.4}$$

$$\mathbf{P}^{+}(\mathbf{x},t) = \sum_{\mu=1}^{M} \mathbf{p}_{\mu}(\mathbf{x}) e^{-ik_{\mu}t}, \tag{3.5}$$

with M unknown lasing modes $\mathbf{\Psi}_{\mu}$ and *real* lasing frequencies k_{μ}. Second, SALT makes the stationary inversion approximation (SIA), i.e, $\dot{D} \approx 0$. In the single-mode regime the SIA is not necessary, as the average inversion in steady-state is exactly zero, but in the multimode regime the inversion is in general not stationary and only under certain conditions is $\dot{D} \approx 0$. However, the development of SALT was specifically oriented towards describing novel solid state microlasers and the necessary conditions are typically satisfied for such lasers, as we discuss in the following.

If the laser is operating in the multimode regime, then the term $E(t) \cdot P(t)$ in Eq. (3.3) above will drive the inversion at all beat frequencies of active modes, which is of order Δk, the free spectral range of the laser. In addition, the polarization can respond at the rate $\gamma_{\perp}$ and could additionally drive time variation in the inversion. However, if the condition $\Delta k, \gamma_{\perp} \gg \gamma_{\parallel}$ holds, then the inversion is being driven non-resonantly and responds quite weakly, except to the dc part of the drive which represents static gain saturation. The effects of the residual four-wave mixing can be included perturbatively if desired, as was done in Ref. [25], but are neglected in standard SALT. The condition $\gamma_{\perp} \gg \gamma_{\parallel}$ is satisfied in essentially all solid state lasers due to strong dephasing, but the condition $\Delta k \gg \gamma_{\parallel}$ depends on the linear dimensions and geometry of the laser cavity and is typically not satisfied for macro scale tabletop lasers. However for a linear cavity it typically would be satisfied for $L < 100\,\mu\text{m}$ and hence the SIA tends to be a good approximation for multimode lasing in micro lasers. This general argument was made by Fu and Haken [100] in 1991 and was applied to Fabry-Perot lasers, for which they provided a stability proof for the multimode state under these conditions. These assumptions leading to the SIA allow the derivation of the much more general SALT equations, which were then tested extensively in comparison to full FDTD simulations for many multimode lasing structures in Refs. [24, 25, 200]. A general linear stability analysis in the SALT framework is challenging due to the necessity of testing stability against all possible spatial fluctuations, something not ever done in standard analyses, where the spatial degrees of freedom are frozen. However, work in this direction is in progress and partial results have been obtained.

Using these well-motivated approximations, Eq. (3.1) can then be written for each lasing mode $\mathbf{\Psi}_\mu(\mathbf{x})$ as

$$\Big[-\nabla\times\nabla\times+k_\mu^2\varepsilon_c(\mathbf{x})+k_\mu^2\gamma(k_\mu)D\Big]\mathbf{\Psi}_\mu(\mathbf{x})=0, \tag{3.6}$$

where the two-level active gain material is described by the non-linear susceptibility $\gamma(k_\mu)D$. Here, $\gamma(k_\mu)$, is the Lorentzian gain curve, where

$$\gamma(k_\mu)\equiv\frac{\gamma_\perp}{k_\mu-k_a+i\gamma_\perp}, \tag{3.7}$$

and D the population inversion. The latter contains the spatial hole-burning term that nonlinearly couples all lasing modes,

$$D(\mathbf{x},d,\{k_\nu,\mathbf{\Psi}_\nu\})=\frac{D_0(\mathbf{x},d)}{1+\sum_{\nu=1}^{M}|\gamma(k_\nu)\mathbf{\Psi}_\nu(\mathbf{x})|^2}, \tag{3.8}$$

where the $\mathbf{\Psi}_\nu(\mathbf{x})$ are in their natural unit $e_c=2g/\hbar\sqrt{\gamma_\perp\gamma_\parallel}$.

The non-linear SALT equations, Eq. (3.6), for the electric field of the lasing modes, $\mathbf{\Psi}_\mu(\mathbf{x})$, and for the associated lasing frequencies k_μ can be conceived of as the limit of an amplifying scattering process in which the input goes to zero, corresponding to purely outgoing solutions with real frequency or, equivalently, to a pole in the relevant scattering matrix on the real axis. Until the external pump is strong enough for the gain to balance the loss there will be no solution of this type, i.e., $\boldsymbol{\Psi}_\mu(\mathbf{x}) = 0$. However, when increasing the pump strength, non-trivial solutions appear at a sequence of thresholds and at different frequencies k_μ. The non-linear interaction between these solutions is through the spatial hole-burning and depletion of the gain medium, Eq.(3.8): each lasing mode extracts energy from the pump in a space-dependent manner which in general makes it more difficult for subsequent modes to reach threshold, and also effectively changes the index of refraction of the gain medium.

As already noted, Eq. (3.6) has been solved in 1D and 2D geometries, where either the electric or the magnetic field can be treated as scalar, for diverse systems such as random, microdisk or photonic crystal lasers using an algorithm based on expansion of the solutions in the CF basis [26]. In the most recent and most efficient formulation, the *linear* non-Hermitian eigenvalue problem,

$$\left[-\nabla \times \nabla \times + k^2\varepsilon_c(\mathbf{x}) + k^2\eta_n(k)f(\mathbf{x})\right] u_n(\mathbf{x};k) = 0, \tag{3.9}$$

is used to define the optimal set of threshold CF states $u_n(\mathbf{x};k)$ and eigenvalues $\eta_n(k)$.

The function $f(\mathbf{x})$ adapts the basis to the spatial pump profile of the experiment of interest and is nonzero only inside the gain medium. The $u_n(\mathbf{x};k)$ form a complete basis and satisfy a biorthogonality relation at any frequency k. Equation (3.6) is solved by projecting the lasing modes $\Psi_\mu(\mathbf{x})$ into the CF basis. The resulting non-linear eigenvalue equation can only be satisfied at discrete frequencies which hence determine the lasing frequencies, k_μ. In principle one does not need to pre-calculate and store the CF basis at different real values of k but it is numerically favorable to do so in general. However, the wider the Lorentzian gain curve, Eq. (3.7), is compared to the free spectral range, the more memory intensive the storage of the CF basis becomes, which makes calculations problematic in two and three dimensions. Moreover, if the pump profile $f(\mathbf{x})$ is fixed and only its amplitude is varied experimentally, then CF states need only be calculated for various k values, but if the pump profile also varies along a pump trajectory then one has to calculate new CF states also for many values of d [200]. For a limited set of highly symmetric cavities, including piecewise-homogeneous 1D slabs and uniform cylinders, the solution of Eq. (3.9) is known semi-analytically at any k. However, for all other geometries, Eq. (3.9) must be solved numerically for all relevant k needed to build a basis. Consequently, for a fully-vectorial treatment of SALT in arbitrary cavities, CF bases cannot be used without significant computational costs. Our direct solution method eliminates the computation and storage of CF bases and scales easily to 3D geometries.

3.3 Solution method

3.3.1 Overview

The basic idea of our new solution method to obtain the lasing modes in the SALT is as follows: We discretize Eq. (3.6), using standard discretization techniques like FEM or FDFD, and iteratively solve for the lasing modes $\mathbf{\Psi}_\mu$ and their frequencies k_μ at successively increasing values of the pump parameter d. This nonlinear coupled problem is most conveniently solved by using the Newton-Raphson method. For initial guesses, we use the modes at threshold when we are close above threshold, and the modes at the previous pump step when we are far above threshold. In order to find the first threshold and the corresponding solution, Eq. (3.6) is initially solved for $d = 0$ as an eigenvalue problem (EVP). The solutions are the resonances or quasi bound states $\bar{\mathbf{\Psi}}_n$ of the passive cavity, corresponding to the poles of the passive scattering matrix (S matrix) [26] with frequencies $\bar{k}_n$ lying in the negative imaginary half plane (note that we will label all quantities below threshold with overbars throughout the paper). While increasing the pump d, Eq. (3.6) is solved without the nonlinearity in Eq. (3.8) and the nonlasing modes near the gain frequency k_a are tracked until the first $\bar{k}_{n_0}$ reaches the real axis and turns the corresponding mode into an active lasing mode, $\bar{\mathbf{\Psi}}_{n_0} \to \mathbf{\Psi}_1$. Once we have crossed the first threshold, we use the solutions for $\bar{\mathbf{\Psi}}_{n_0}$ and $\bar{k}_{n_0}$ of the eigenvalue problem at threshold as a first guess for the solution of $\mathbf{\Psi}_1$ and k_1 in the non-linear Newton solver slightly above threshold. The latter already includes the non-linearity $D(\mathbf{x}, d, \{k_1, \mathbf{\Psi}_1\})$ which, once the Newton solver has converged, we treat as a fixed function like $\varepsilon_c(\mathbf{x})$ to examine the remaining non-lasing modes $\bar{\boldsymbol{\Psi}}_n$ at the current pump strength d. This has to be done in order to verify if further modes cross the lasing threshold. For the non-lasing modes, Eq. (3.6) is thus only nonlinear in $\bar{k}_n$ and linear in $\bar{\boldsymbol{\Psi}}_n$, such that this problem can be cast into a nonlinear EVP [104]. The procedure of increasing the pump is now continued by tracking the lasing mode solving the nonlinear coupled SALT system, while the non-lasing modes are evaluated from the corresponding nonlinear EVP until a second mode reaches threshold. At this point the number of lasing modes is increased by 1 and the procedure continues with two and more lasing modes in essentially the same way.

To illustrate this approach in more detail, we apply it to the simple one-dimensional edge-emitting laser shown in Fig. 3-1(a) which already captures all the main features. We pump the 1D slab cavity uniformly along its length $L = 100\,\mu\text{m}$ with a pump strength $D_0(\mathbf{x}, d) = d$ which, above the first threshold, leads to emission to the right. Starting with $d = 0$, where the SALT system reduces to a simple resonance problem, we increase d and observe that the resonance poles move upwards in the complex plane; see Fig. 3-1(b) where the starting point $d = 0$ is marked by circles and the pump value at the first threshold, $d_1 = 0.267$, is marked by triangles. Below this first threshold no mode is lasing, such that the non-linear spatial hole-burning term is zero, resulting in the following PDE for all non-lasing modes,

$$\left\{-\nabla \times \nabla \times +\bar{k}_n^2\left[\varepsilon_c(\mathbf{x}) + \gamma(\bar{k}_n) D_0(\mathbf{x}, d)\right]\right\} \bar{\boldsymbol{\Psi}}_n(\mathbf{x}) = 0\,, \tag{3.10}$$

which is linear with respect to $\bar{\boldsymbol{\Psi}}_n$, but into which the resonance values $\bar{k}_n$ enter non-linearly. Starting at the first threshold, the terms $\boldsymbol{\Psi}_1$ and k_1 of the first lasing mode enter the spatial hole-burning denominator in Eq. (3.8) (where $M = 1$), resulting in the following equation for the first lasing mode $\boldsymbol{\Psi}_1$ and its wavenumber k_1,

$$\left\{ - \nabla \times \nabla \times +k_1^2\left[\varepsilon_c(\mathbf{x}) + \frac{\gamma(k_1) D_0(\mathbf{x}, d)}{1 + |\gamma(k_1)\boldsymbol{\Psi}_1(\mathbf{x})|^2}\right]\right\}\boldsymbol{\Psi}_1(\mathbf{x}) = 0 \tag{3.11}$$

which is now nonlinear with respect to both $\boldsymbol{\Psi}_1$ and k_1. When continuing to increase the pump, the frequencies corresponding to the active modes are forced to stick to the

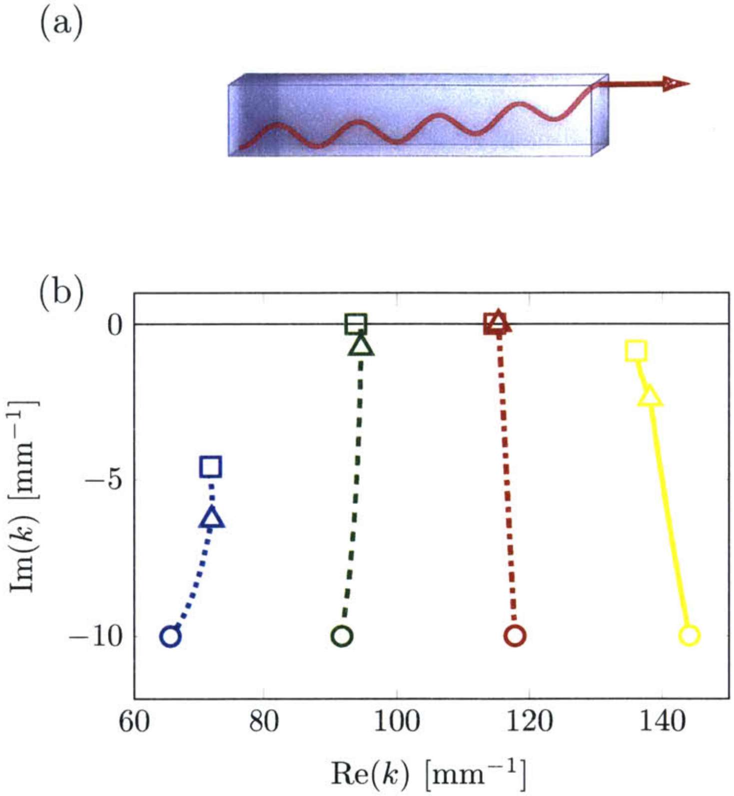


Figure 3-1: (a) 1D slab cavity laser of length $L = 100$ μm with purely reflecting boundary on the left side and open boundary on the right side. The mode shown in red (gray)corresponds to the intensity profile of the first lasing mode at threshold. (b) SALT eigenvalues corresponding to the scattering matrix poles for a uniform and linearly increasing pump strength $D_0(\mathbf{x}, d) = d$ applied inside the slab $[D_0(\mathbf{x}, d) = 0$ outside$]$. We use a refractive index $\sqrt{\varepsilon_c} = 1.2$ in the slab ($\sqrt{\varepsilon_c} = 1$ outside), a gain frequency $k_a = 100$ mm^{-1} and a polarization relaxation rate $\gamma_\perp = 40$ mm^{-1}. The trajectories start at $d = 0$ (circles) and move toward the real axis with different speed when increasing d. The first lasing mode (dash-dotted red line) activates at $d = 0.267$ (triangles) with $k_1 = 115.3$ mm^{-1}. The trajectories end at $d = 1$ (squares) where a second lasing mode (dashed green line) turns active and the two other non-lasing modes (blue dotted and yellow solid line) remain inactive.

real axis, while the eigenvalues associated to all other inactive modes continue moving upwards, see Fig. 3-1(b). To detect the activation of further modes, the inactive modes have to be recalculated again, however, this time by additionally taking into account the spatial hole burning contribution of the currently lasing active mode $(\boldsymbol{\Psi}_1, k_1)$ at a given pump strength d. For this, we insert the currently active mode into the denominator of Eq. (3.11) which turns the above nonlinear problem into another nonlinear (in $\bar{k}_n$) eigenvalue problem,

$$\Big\{ -\nabla\times\nabla\times + \bar{k}_n^2 \Big[\varepsilon_c(\mathbf{x}) + \frac{\gamma(\bar{k}_n) D_0(\mathbf{x}, d)}{1 + |\gamma(k_1)\boldsymbol{\Psi}_1(\mathbf{x})|^2}\Big]\Big\} \bar{\boldsymbol{\Psi}}_n(\mathbf{x}) = 0 \tag{3.12}$$

which, however, has the same structure as Eq. (3.10). As soon as the imaginary part of another eigenvalue $\bar{k}_n$ reaches the real axis, a new laser mode $\boldsymbol{\Psi}_2$ becomes active which increases the size of the nonlinear problem by 1. For even higher pump strength and a larger number of lasing modes this procedure continues accordingly. Note also, that the case when a mode shuts down during the pumping process can be incorporated without major effort.

To summarize, the solution of the SALT equations reduces essentially to computing the full nonlinear (in $\boldsymbol{\Psi}_\mu$ and k_μ) system of PDEs through a Newton-Raphson method and the computation of an EVP which is linear in $\bar{\boldsymbol{\Psi}}_n$ but which still remains nonlinear in $\bar{k}_n$. Details of how to obtain the *active* or *lasing* solutions $\{\boldsymbol{\Psi}_\mu, k_\mu\}$ of the Newton problem as well as the *inactive* or *non-lasing* solutions $\{\bar{\boldsymbol{\Psi}}_n, \bar{k}_n\}$ through the nonlinear EVP are provided in the following two sections.

3.3.2 Lasing modes

For modes that are lasing, Eq. (3.6) is nonlinear in the unknowns $\{\boldsymbol{\Psi}_\mu(\mathbf{x}), k_\mu\}$. As theses modes are all coupled together through the spatial hole-burning interaction, they must be solved simultaneously. In general, such systems of nonlinear equations can be written in the form

$$\mathbf{f}(\mathbf{v}) = 0 \tag{3.13}$$

where the vector of equations $\mathbf{f}$ is an analytic nonlinear function of the unknown solution vector $\mathbf{v}$ which again gathers all unknowns $\{\mathbf{\Psi}_\mu(\mathbf{x}), k_\mu\}$. This nonlinear problem can generally be solved by using the Newton-Raphson method [1]. The basic idea is that for a guess $\mathbf{v}_0$ for the solution $\mathbf{v}$, one can write

$$\mathbf{v} - \mathbf{v}_0 = -\mathcal{J}(\mathbf{v}_0)^{-1}\mathbf{f}(\mathbf{v}_0) + \mathcal{O}(|\mathbf{v} - \mathbf{v}_0|^2), \tag{3.14}$$

where $\mathcal{J}$ is the Jacobian matrix of partial derivatives of $\mathbf{f}$ with respect to $\mathbf{v}_0$. A solution $\mathbf{v}$ can usually be obtained by iterating Eq. (3.14) using only the linear terms. This iterative algorithm converges "quadratically" (squaring the errors on each step [1]) if $|\mathbf{v}_0 - \mathbf{v}|$ is small. Further, we use an analytic evaluation for the Jacobian $\mathcal{J}$ from Eq. (3.6), as described below, and do not need to compute it using numerical differentiation schemes. Since $\mathcal{J}$ is then a sparse matrix each iteration can exploit fast algorithms for sparse linear equations [112, 114].

To solve Eq. (3.13) on a discrete level, we project the complex fields $\mathbf{\Psi}_\mu(\mathbf{x})$ of each lasing mode onto a discrete N-component basis (for a FEM or FDFD approach). Unlike the CF basis, we use a *localized* basis generated once from a grid or mesh. This is the key to producing *sparse* matrices and hence makes the method scalable to the larger bases required in two and three dimensions. The discretizations on such a basis turn the fields $\mathbf{\Psi}_\mu$ into complex coefficient vectors $\mathbf{c}_\mu$, while k_μ is required to be purely real. Because the SALT equations are not differentiable in the complex fields (due to the complex conjugation), we split our unknown coefficient vectors $\mathbf{c}_\mu$ (and the vector function $\mathbf{f}$ accordingly) into their real and imaginary parts. The discretized version of $\mathbf{v}$ then consists of $(2N+1)M$ real unknowns (fields and frequencies). However, we only obtain $2NM$ real-valued equations from $\mathbf{f}$. The underspecification comes from the fact that the hole-burning term $D(\mathbf{x}, \{k_\nu, \mathbf{\Psi}_\nu\})$ happens to be invariant under global phase rotations $\mathbf{\Psi}_\nu(\mathbf{x}) \to e^{i\phi_\nu}\mathbf{\Psi}_\nu(\mathbf{x})$. In addition to the problem of underspecification, there is also a problem of stability: for lasing modes slightly above threshold, the amplitude is nearly zero, which would result in problems distinguishing between the solution we want and the trivial solution $\mathbf{\Psi}(\mathbf{x}) = 0$. We resolve both issues by normalizing the amplitude and fixing the phase of all lasing modes while keeping track of their amplitudes using a separate variable. This procedure results in both the number of real unknowns and the number of real equations being $(2N+2)M$.

Here, we provide further details on setting up the Newton-Raphson iteration for M lasing modes. First, we describe how to fix the phase and normalization for each mode (as mentioned in Sec. 3.3.2). We choose a point $\mathbf{x}_0$ and a constant unit vector $|\mathbf{a}| = 1$ such that $\mathbf{a} \cdot \boldsymbol{\Psi}_\mu(\mathbf{x}_0)$ is nonzero for all lasing modes. This condition is usually satisfied provided that $\mathbf{x}_0$ is neither far outside the cavity nor a point of high symmetry. We then define the quantity $s_\mu \equiv |\mathbf{a} \cdot \boldsymbol{\Psi}_\mu(\mathbf{x}_0)|$ and rescale the field such that the *physical* field becomes $s_\mu^{-1} \boldsymbol{\Psi}_\mu(\mathbf{x})$ and the rescaled field satisfies

$$\mathbf{a} \cdot \boldsymbol{\Psi}_\mu(\mathbf{x_0}) = 1. \tag{3.15}$$

With this redefinition, the rescaled field $\boldsymbol{\Psi}_\mu(\mathbf{x})$ has a fixed phase and a normalization that distinguishes it from the trivial solution $\boldsymbol{\Psi}_\mu(\mathbf{x}) = 0$. Further we treat the quantity s_μ as a separate unknown that contains the mode's amplitude. The spatial hole-burning [Eq. (3.8)] then becomes

$$D(\mathbf{x}, d, \{k_\nu, s_\nu, \boldsymbol{\Psi}_\nu(\mathbf{x})\}) = \frac{D_0(\mathbf{x}, d)}{1 + \sum |\gamma(k_\nu) \boldsymbol{\Psi}_\nu(\mathbf{x})|^2 s_\nu^{-2}}.$$

We note that as an extra optimization, we can also absorb the factor of $|\gamma(k_\nu)|$ into s_ν. This extra step simplifies the k_ν derivatives of D that will be taken later, and we can easily replace the extra factor of $|\gamma(k_\nu)|$ at the very end, when the problem has been solved. However, for continuity and clarity, we skip this step here.

Now, we describe how to construct the vector of unknowns $\mathbf{v}$ which, after rescaling, should contain $\boldsymbol{\Psi}_\mu(\mathbf{x})$, k_μ, and s_μ. First, the discretized fields $\boldsymbol{\Psi}_\mu(\mathbf{x})$ are described by N-component complex vectors $\mathbf{b}_\mu$. The $2N+2$ real unknowns for each mode can then be written in block form as

$$\mathbf{v}^\nu = \begin{pmatrix} \mathbf{v}_1^\nu \\ \mathbf{v}_2^\nu \\ \mathbf{v}_3^\nu \\ \mathbf{v}_4^\nu \end{pmatrix} = \begin{pmatrix} \mathrm{Re}[\mathbf{b}_\nu] \\ \mathrm{Im}[\mathbf{b}_\nu] \\ k_\nu \\ s_\nu \end{pmatrix}. \tag{3.16}$$

The vector $\mathbf{v}$ we use for the Newton-Raphson method contains all $\mathbf{v}^\mu$ in sequence, since the lasing modes are all coupled together through the spatial hole-burning interaction and thus must be solved simultaneously.

Next, we construct the equation vector $\mathbf{f}$ by discretizing the operator $-\nabla \times \nabla \times +k_\mu^2 \left[\varepsilon_c(\mathbf{x}) + \gamma(k_\mu)D\right]$ into a sparse complex matrix $\mathbf{S}_\mu$. In the discrete basis, Eq. (3.6) becomes $\mathbf{S}_\mu \mathbf{b}_\mu = 0$ which gives N complex scalar equations, and the normalization condition that fixes the phase [Eq. (3.15)] becomes the complex scalar equation $\mathbf{e}^T \mathbf{b}_\mu = 0$, where $\mathbf{e}^T$ is the discrete-basis representation of the vector function consisting of the unit vector $\mathbf{a}$ at point $\mathbf{x}_0$ and zero everywhere else. The real and imaginary parts of these $N+1$ complex equations can be written in block form as

$$\mathbf{f}^\mu = \begin{pmatrix} \mathbf{f}_1^\mu \\ \mathbf{f}_2^\mu \\ \mathbf{f}_3^\mu \\ \mathbf{f}_4^\mu \end{pmatrix} = \begin{pmatrix} \mathrm{Re}\left[\mathbf{S}_\mu \mathbf{b}_\mu\right] \\ \mathrm{Im}\left[\mathbf{S}_\mu \mathbf{b}_\mu\right] \\ \mathrm{Re}\left[\mathbf{e}^T \mathbf{b}_\mu\right] - 1 \\ \mathrm{Im}\left[\mathbf{e}^T \mathbf{b}_\mu\right] \end{pmatrix}. \tag{3.17}$$

The vector $\mathbf{f}$ we use for the Newton-Raphson method contains all $\mathbf{f}^\mu$ in sequence, due to the intermodal coupling.

Finally, we describe how to construct the (real) Jacobian matrix $\mathcal{J}$ (which is real), which consists of M^2 blocks $\mathcal{J}^{\mu\nu}$ that each have size $2N+2$ and have the block form

$$\mathcal{J}_{ij}^{\mu\nu} = \frac{\partial \mathbf{f}_i^\mu}{\partial \mathbf{v}_j^\nu}.$$

We explicitly construct these blocks by taking derivatives of column blocks of $\mathbf{f}_i^\mu$ with respect to row blocks of $(\mathbf{v}_j^\nu)^T$, as defined in Eqs. (3.16) and (3.17). First, we see that

$\mathcal{J}_{31}^{\mu\nu} = \mathcal{J}_{42}^{\mu\nu} = \mathbf{e}^T$, while all other blocks of $\mathcal{J}_{ij}^{\mu\nu}$ with $i = 3, 4$ are zero. Second, we have the columns

$$\begin{pmatrix} \mathcal{J}_{13}^{\mu\nu} \\ \mathcal{J}_{23}^{\mu\nu} \end{pmatrix} = \begin{pmatrix} \mathrm{Re} \\ \mathrm{Im} \end{pmatrix} \left[\frac{\partial \mathbf{S}_\mu}{\partial k_\nu} \mathbf{b}_\mu \right]$$

and

$$\begin{pmatrix} \mathcal{J}_{14}^{\mu\nu} \\ \mathcal{J}_{24}^{\mu\nu} \end{pmatrix} = \begin{pmatrix} \mathrm{Re} \\ \mathrm{Im} \end{pmatrix} \left[\frac{\partial \mathbf{S}_\mu}{\partial s_\nu} \mathbf{b}_\mu \right],$$

where the derivatives of $\mathbf{S}_\mu$ are diagonal complex matrices that can be obtained straightforwardly by discretizing the same derivatives of the complex scalar function $k_\mu^2 \left[\varepsilon_c(\mathbf{x}) + \gamma(k_\mu) D \right]$. (In the case that exact outgoing radiation conditions are used for $\mathbf{S}_\mu$, the matrix for $-\nabla \times \nabla\times$ may also depend on k_μ and this dependence must also be included in the derivative.)

Finally, the remaining blocks are given by

$$\begin{pmatrix} \mathcal{J}_{11}^{\mu\nu} & \mathcal{J}_{12}^{\mu\nu} \\ \mathcal{J}_{21}^{\mu\nu} & \mathcal{J}_{22}^{\mu\nu} \end{pmatrix} = \begin{pmatrix} \mathrm{Re} & -\mathrm{Im} \\ \mathrm{Im} & \mathrm{Re} \end{pmatrix} \mathbf{S}_\mu \delta_{\mu\nu} + \mathbb{S}_{\mu\nu}$$

where $\delta_{\mu\nu}$ is the Kronecker δ, and $\mathbb{S}_{\mu\nu}$ is the matrix discretization of the real 6×6 tensor function

$$2k_\mu^2 \gamma(k_\mu) \frac{\partial D}{\partial \left| \mathbf{\Psi}_\nu(\mathbf{x}) \right|^2} \mathbf{\Psi}_\mu(\mathbf{x}) \otimes \mathbf{\Psi}_\nu(\mathbf{x})$$

with the outer product $\otimes$ taken over the real and imaginary parts of the vector components of $\mathbf{\Psi}(\mathbf{x})$.

For finite-difference calculations shown in the main text, the discretization code implemented in [205,206] was used. The complex electric fields $\mathbf{\Psi}_\mu$ were discretized on an $N_x \times N_y \times N_z$ pixel grid of equally spaced points with the $-\nabla \times \nabla\times$ operator being conveniently discretized using second-order centered differences on a Yee lattice [37]. To impose outgoing boundary conditions, additional pixels of PML were added at the boundaries with the appropriate absorption, as explained in Sec. 3.3.4. For each mirror symmetry in a geometry, we were able to halve the computational domain by replacing the PML at the lower walls with the corresponding boundary conditions of the mirror plane. Furthermore, for the cases of TM ($E_{x,y} = 0$) and TE polarization ($E_z = 0$), the problem size can be reduced by factors of 3 and 3/2, respectively, by projecting $\mathbf{S}_\mu$ and $\mathbf{b}_\mu$ into the nonzero field components only. Additionally, 2D calculations were performed by setting $N_z = 1$ and the boundary condition in the z direction to be periodic. 1D calculations were performed by doing so for both the z and y directions.

Note that for the Newton-Raphson iteration to be scalable to higher dimensions and to high-resolution meshes, it is also important to use a scalable solver (in our case, the sparse direct solver [114] PaStiX [113] was called from the PETSc library [207] because the Jacobian is sparse). For very large-scale 3D systems, it may become necessary to use iterative linear solvers [112] for each Newton step instead, in which case it is important to select certain PML formulations [48].

3.3.3 Non-lasing modes

In order to find the first pump threshold and the corresponding lasing solution as well as to verify when a new mode activates, the non-lasing modes have to be monitored while changing the pump. These non-lasing modes $\bar{\boldsymbol{\Psi}}_n$ are defined as complex-frequency solutions to Eq. (3.6) that do not enter into the nonlinear hole burning term in $D(\mathbf{x}, d, \{k_\nu, \boldsymbol{\Psi}_\nu\})$, see Eq. (3.8). Due to causality constraints, the complex eigenvalues associated with non-lasing modes, $\bar{k}_n$ always feature $\mathrm{Im}(\bar{k}_n) < 0$, and usually approach the real axis as d is increased. When all lasing modes have been determined for a particular d, the function $D(\mathbf{x}, d, \{k_\nu, \boldsymbol{\Psi}_\nu\})$ is known and can be treated as a fixed function like $\varepsilon_c(\mathbf{x})$, see Eq. (3.12). As outlined in Sec. 3.3.1, this reduces Eq. (3.6) to a non-Hermitian, nonlinear eigenvalue problem (NEVP) which is *linear* in the eigenvectors $\bar{\boldsymbol{\Psi}}_n(\mathbf{x})$, but nonlinear in the complex eigenvalues $\bar{k}_n$.

For situations where we are only interested in the behavior of a few lasing modes in a small range of the pump parameter d, Newton's method is still a convenient approach to determine the non-lasing modes and, in fact, the only viable method in terms of computational cost for high resolution 2D or 3D computations. In this case, we typically use standard EVP algorithms to solve Eq. (3.6) first for $d = 0$ (which is usually either linear or quadratic in $\bar{k}$, depending on the method for implementing the outgoing radiation condition). This provides us all the modes of interest which we then track to threshold with Newton's method as d is increased. As in Sec. 3.3.2, convergence is "quadratic", but, unlike for the lasing modes, Eq. (3.14) can be used with complex unknowns and equations since Eq. (3.6) is differentiable in all unknowns once $D(\mathbf{x}, d, \{k_\nu, \mathbf{\Psi}_\nu\})$ is fixed. The downside of Newton's method is that, in the absence of a good initial guess, it can be very unpredictable and slow to converge. Such a situation arises, e.g., when the modes that can lase are not known *a priori* as in the case where a large number of near-threshold modes are clustered together, all with frequencies close to the gain center k_a. In this instance, a more general and comprehensive method for evaluating the non-lasing modes is required.

Such more general techniques exist in terms of NEVP solvers [104]. One conceptually simple method for our problem is to divide Eqs. (3.6) and 3.12) by $\gamma(k)$, turning the rational EVP into a cubic EVP which can then be linearized at the expense of making the problem three times as large and possibly also very ill-conditioned. Other, more sophisticated solution methods include "trimmed" linearization [2], Newton [3], Jacobi-Davidson [208], rational Krylov [209], and nonlinear Arnoldi [210]. Independently of the chosen solution strategy, we can take into account that only modes which have a spectral overlap with the gain curve $\gamma(k)$ near its center frequency k_a are expected to be candidates for active laser modes. In addition, the Lorentzian gain curve of width $\gamma_\perp$ produces a singularity in the NEVP at $k = k_a - i\gamma_\perp$ which may result in spurious numerical solutions. Combining these observations, we restrict our attention to those eigenvalues $\bar{k}_n$ that are in the following cropped subpart of the complex plane: $\{z \in \mathbb{C} \,|\, \mathrm{Im}(z) > -\gamma_\perp \wedge \mathrm{Re}(z) \in [k_a - \gamma_\perp, k_a + \gamma_\perp]\}$. A suitable method that allows us to conveniently include such auxiliary restrictions is the contour integral method presented recently in [211, 212]. There, the search for eigenvalues is restricted to a region within a smooth contour such as a circle or an ellipse. By using the residue theorem, all poles of the inverse of the differential operator, which are equivalent to the eigenvalues of the same operator, are obtained within the specified contour. This feature is not only useful for employing this method as a stand-alone solver for non-lasing modes, but also as a complementary tool to check if, in addition

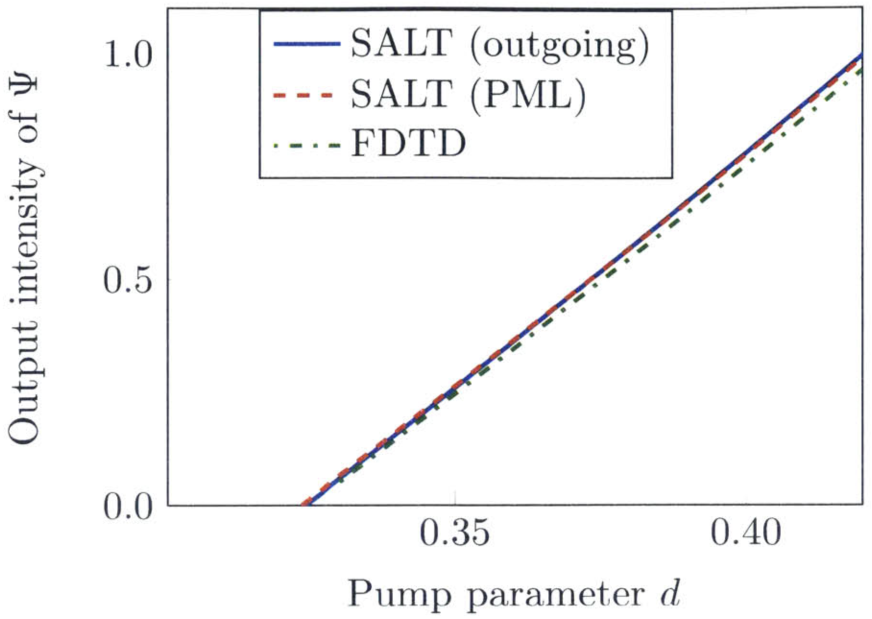


Figure 3-2: Comparison between the laser output using SALT with exact outgoing boundary conditions and PML absorbing layers, on the one hand, and a full time integration of the MB equations using FDTD, on the other hand. We study the first and second TM lasing modes of a 1D slab cavity which is similar to the one above. The applied pump is uniform, $D_0(\mathbf{x}, d) = d$, the cavity has a uniform dielectric $\sqrt{\varepsilon_c} = 2$ a length $L = 100$ μm, and gain parameters $\gamma_\perp = 3$ mm^{-1}, $k_a = 300$ mm^{-1}. For the FDTD simulations additionally $\gamma_\| = 0.001$ mm^{-1} was used. The PML method is nearly as accurate as the outgoing boundary condition, but has the advantage of being easily generalizable to two and three-dimensional calculations [37]. The times to reach $d = 0.11$ are shown for the two methods (with identical spatial resolution). The FDTD computation was done on the Yale BulldogK cluster with E5410 Intel Xeon CPUs, while the SALT computations were done on a Macbook Air.

to the limited set of non-lasing modes that are tracked with a Newton solver, no new modes have entered the region of interest within the chosen contour.

3.3.4 Outgoing radiation condition

For numerical computations, the outgoing radiation condition must be implemented within a truncated, finite domain. In one dimension, the radiation condition can be expressed exactly [213]. This also allows us to shift the boundary of the domain right to the border of the cavity, which decreases the computational cost. This method is, however, not easily generalizable to two and three dimensions [37]. An efficient and robust alternative is to use the standard perfectly matched layer (PML) technique [47, 70] in which an artificial material is placed at the boundaries. The material has a certain complex permittivity and permeability such that it is absorbing and analytically reflectionless. In one dimension, the PML technique can be tested against an exact outgoing boundary condition, and the two methods yield results that are nearly indistinguishable, as shown in Fig. 3-2. Also shown in Fig. 3-2 is a comparison with conventional methods of solving the MB equations using finite difference time domain (FDTD) simulations demonstrating the validity of the stationary inversion approximation used in the derivation of the SALT equations. Both the quantitative agreement between SALT and FDTD solutions as well as the former's substantial numerical efficiency over the latter have been previously documented [24, 25]. Of course, the precise computation times depend on many factors, including hardware details, parameter choices in the algorithms, and software implementation quality, but the magnitude of the difference here makes it unlikely that any FDTD implementation could be competitive with the SALT approach.

3.4 Assessment and application of the solution method

In this section we will validate our solution strategy against the traditional method based on CF states and we will show first results for prototypical laser cavities.

3.4.1 1D slab laser as test case

We demonstrate here the accuracy of the presented direct solver method by studying in more detail the 1D edge-emitting slab laser introduced in Sec. 3.3.1. One of the advantages of the direct solver, as compared to the CF method, is the accuracy of its solutions far above the threshold. In this regime the CF basis becomes a poorer match for the lasing modes and, as explained in Sec. 3.1, a large number N_{CF} of basis functions is required for convergence compared to near threshold. This is especially relevant for low-Q (short-lifetime) laser resonators such as random lasers or cavities featuring gain-induced states, as considered, e.g., in [214] . In Fig. 3-3 the intensity

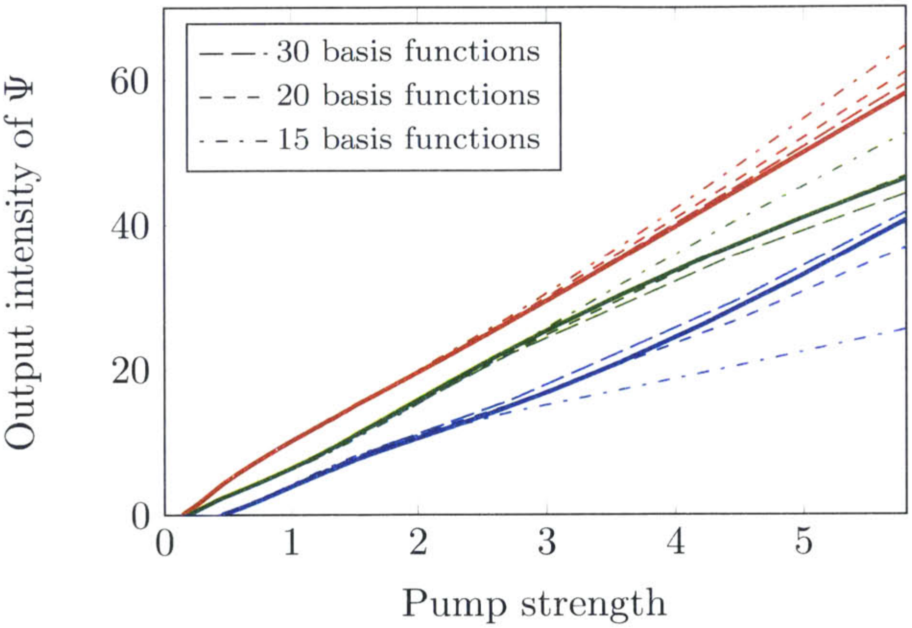


Figure 3-3: Output intensity vs pump strength in a 1D resonator with reflecting boundary on the left side and outgoing radiation on the right side; see Fig. 3-1(a). The cavity has length 100 μm with a refractive index $n = 1.01$. The gain curve has its peak at $k_a = 250$ mm^{-1} and a width $2\gamma_\perp = 15$ mm^{-1}. The output intensity is given by $|\mathbf{\Psi}|^2$ evaluated at the right boundary $x = L$. The pump is constant in the entire cavity. Solid lines describe the results of our solution method. Comparing them to the solutions of the CF-state formalism with 30 (long dashed), 20 (dashed), and 15 (dash-dotted) CF-basis functions, one observes that the two approaches converge towards each other for a sufficiently high number of CF states being included.

of such a low-Q cavity is plotted with respect to an overall pump strength d for a constant spatial pump profile. The figure contains both the results of the direct and of the CF state solver. For the latter the solution for different numbers N_{CF} of CF states are depicted, demonstrating that for a larger basis the solution converges towards the solution of the direct solver. Our solution method thus leads to an accuracy far above threshold which can only be achieved by the traditional approach with a considerably large number of CF states.

3.4.2 Scalability to full-vector 2D and 3D calculations

In this section we briefly explore the applicability of our solution strategy to 2D and 3D setups by considering the following prototypical examples: In the 2D case we investigate a circular dielectric resonator and in the 3D case a photonic-crystal slab.

In the former situation we study a circular disk with uniform index, which is routinely used in the experiment due to its long-lived resonances associated with "whispering gallery modes" [116]. For this system we study lasing based on TM polarized modes and compare the Newton method presented here (based on FDFD) with the previously developed CF-state method [26, 28, 29]. Due to the azimuthal symmetry, the resonant TM modes [28, 115] are exact solutions of the Bessel equation characterized by an azimuthal phase $e^{\pm i\ell\theta}$ (with ℓ being an integer angular-momentum quantum number) and subject to outgoing boundary conditions. Due to the circular symmetry, each of the modes with a given value of ℓ comes with a degenerate partner mode, characterized by the quantum number $-\ell$. In the presence of the lasing nonlinearities, a preferred superposition will typically be selected as the stable solution, e.g., the circulating modes $e^{\pm i\ell\theta}$, rather than the $\sin(\ell\theta)$ and $\cos(\ell\theta)$ standing waves. The determination of this stable solution in a degenerate lasing cavity is a complex problem that we plan to address in future work. For validation and demonstration purposes in this chapter, we simply select *a priori* a single solution from each degenerate pair by imposing corresponding symmetry boundary conditions. In the case of the circular cavity, we choose the circulating modes with a phase $e^{-i\ell\theta}$ for comparison with the CF solutions. We obtain these by solving for both the sine and cosine modes

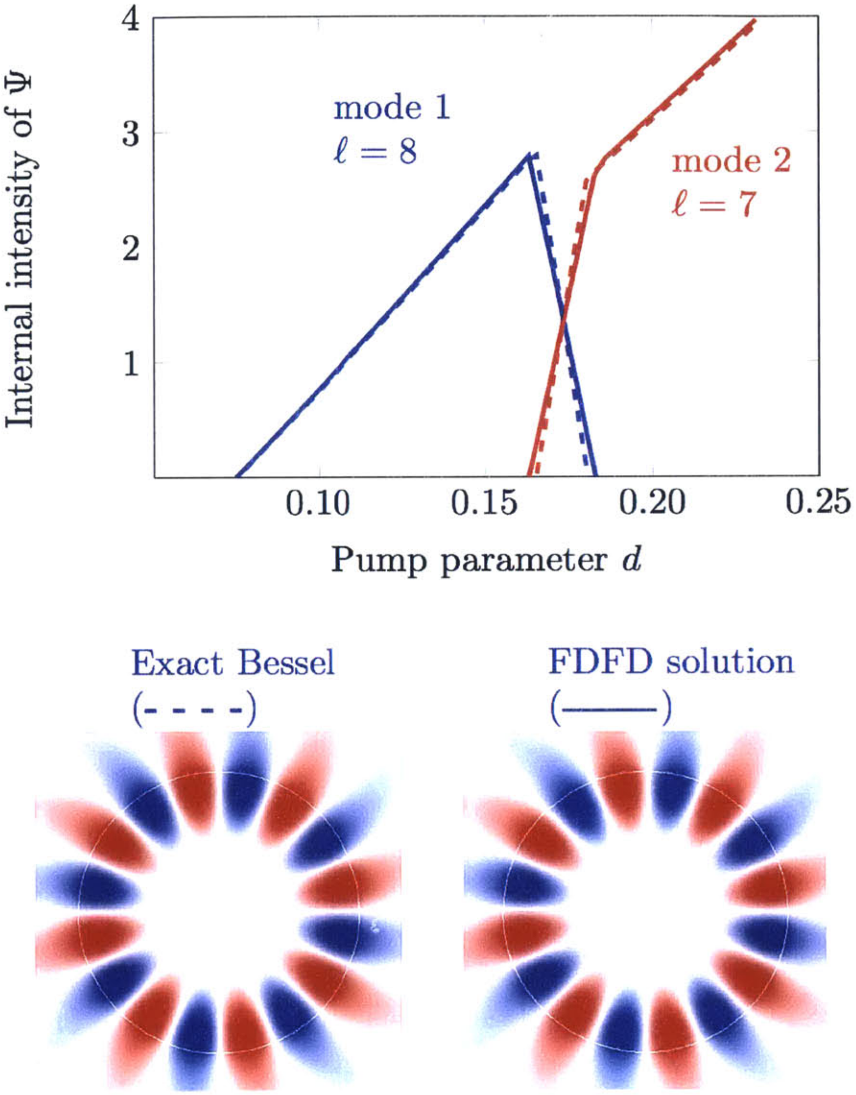


Figure 3-4: Validation of the 2D Newton solver based on FDFD against the CF-state approach (using 20 CF basis states) in a circular cavity with radius $R = 100\ \mu\text{m}$ and dielectric index $n = \sqrt{\varepsilon_c} = 2 + 0.01i$. TM-polarized modes are considered and the following gain parameters are used: $\gamma_\perp = 10\ \text{mm}^{-1}$, $k_a = 48.3\ \text{mm}^{-1}$. Increasing the strength of the uniform pump $D_0(\mathbf{x}, d) = d$, we encounter strong non-linear modal competition between the first two lasing modes with the result that for sufficiently large pump strength the second lasing mode is found to suppress the first one (see top panel). The internal intensity is defined as the integral over the cavity $\int |\mathbf{\Psi}(\mathbf{x})|^2 d\mathbf{x}$. The real part of the lasing mode profile $\Psi(\mathbf{x})$ at the first threshold is shown for both the exact Bessel solution ($\Psi \sim e^{-i\ell\theta}$) and for the finite difference solution (see bottom panel, where blue/white/red color corresponds to negative/zero/positive values). As the pump strength is increased, this profile does not change appreciably apart from its overall amplitude.

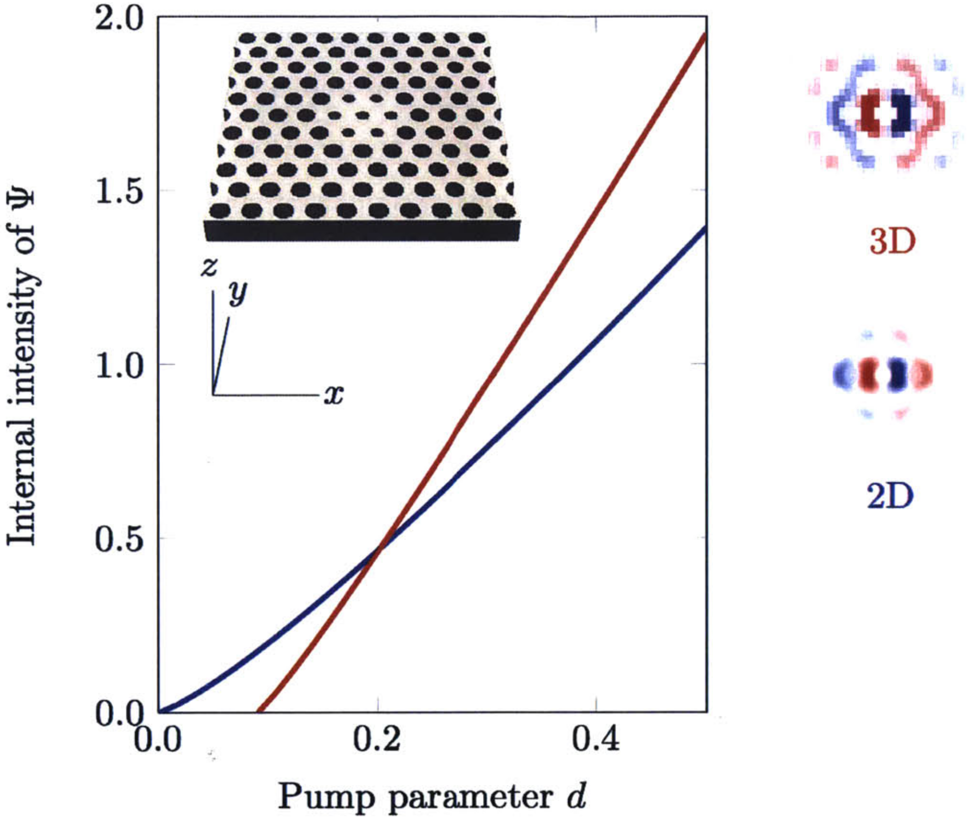


Figure 3-5: 3D calculation of a lasing mode created by a "defect" in a photonic-crystal slab [34]: a period-a hexagonal lattice of air holes (with $a = 1$ mm and radius 0.3 mm) in a dielectric medium with index $n = \sqrt{\varepsilon_c} = 3.4$ with a cavity formed by seven holes of radius 0.2 mm in which a doubly-degenerate mode is confined by a photonic bandgap (one of these degenerate modes is selected by symmetry, see text). The gain has $\gamma_\perp = 2.0$ mm^{-1}, $k_a = 1.5$ mm^{-1}, and non-uniform pump $D_0(\mathbf{x}, d) = f(\mathbf{x})d$, where the pump profile $f(\mathbf{x}) = 1$ in the hexagonal region of height 2 mm in the y-direction, and $f(\mathbf{x}) = 0$ outside that region and in all air holes. The slab has a finite thickness 0.5 mm with air above and below into which the mode can radiate (terminated by PML absorbers). The inset shows magnetic field H_z ($\sim \partial_x E_y - \partial_y E_x$) of the TE-like mode at the $z = 0$ plane.

(using the appropriate boundary conditions at the $x = 0$ and $y = 0$ symmetry planes) and by combining them to construct the exponentially circulating mode.

Under these premises, we find that for uniform pump the first mode turns on at $d \approx 0.075$ and increases linearly in intensity, as seen in Fig. 3-4. The second mode turns on at about twice the pump strength as the first threshold. As the intensity of the second mode increases, we observe a reduction and ultimately a complete suppression of the first mode intensity. This mode competition can be attributed to the following two effects: The two modes have a significant spatial overlap, such that they compete for the same gain through non-linear spatial hole burning which is fully incorporated in SALT. In addition, as being spectrally closer to the peak of the gain curve $\gamma(k)$, the second mode can profit more strongly from the gain in the disk than the first mode. As a result, the second mode prevails against the first mode in this non-linear competition. This behavior of interaction-induced mode switching is general and can be found in other laser configurations and nonlinear media as well [31]. In Fig. 3-4 we show that this behavior is faithfully reproduced with our approach, not only in terms of the modal intensities as a function of the applied pump (see top panel), but also in terms of the corresponding lasing modes which mirror those obtained with the CF-state technique very accurately (see bottom panel).

The second example we consider is a photonic crystal slab with a "defect" (see inset Fig. 3-5) engineered to efficiently trap a mode [215]. The photonic crystal is formed in a dielectric slab by holes which are arranged in a hexagonal lattice and the defect is created by decreasing the radius of seven of the holes in the center. In our study, we focus on a TE-like lasing mode, situated at the defect (spatially) and in the bandgap of the lattice (spectrally). To select one of the degenerate standing-wave solutions, we impose even and odd symmetry at $x = 0$ and $y = 0$, respectively, as well as an even symmetry at $z = 0$. Staying close to a potential experimental realization, we choose the pump profile $D_0(\mathbf{x}, d)$ to be uniform inside the slab material's defect region but zero outside and in the air holes. Increasing the overall amplitude of this pump profile, we find the lasing behavior shown in Fig. 3-5 (main panel). This calculation was performed with 16 nodes (using one CPU per node) of the Kraken Cray XT5 at the University of Tennessee. With 144 × 120 × 40 pixels (the mirror conditions effectively halve these), the total wall-clock time for the computation, from passive resonance at $d = 0$ to lasing above threshold at $d = 0.18$, was 5.9 min. Pump steps of $\delta d = 0.02$ were taken, with three to four Newton iterations per pump value.

3.5 Concluding Remarks

In this chapter, we have presented an algorithm for solving the SALT equations which describe the steady-state lasing modes and frequencies of lasers with a free spectral range and a dephasing rate that are both large as compared to the population decay rate and the relaxation oscillation frequency. These conditions are typically satisfied by microlasers with a linear dimension that does not exceed a few hundred wavelengths. Our solution strategy proceeds by a direct discretization using standard methods as FEM or FDFD, without the need for an intermediate CF basis. The resulting increase in efficiency lets our approach scale to complex 2D and 3D lasing structures, which paves the way for future work in a number of directions.

First, it is now possible to study lasing in much more complex geometries than could previously be readily simulated, offering the possibility of discovering geometries that induce unexpected new lasing phenomena. Going one step further, future computations could search a huge space of lasing structures via large-scale optimization ("inverse design"), which has already been applied to the design of *linear* microcavities [92, 216, 217]. Since our approach is only more expensive than the solution of linear cavity modes by a small constant factor (e.g., the number of modes and the number of Newton iterations) it will be the ideal tool for this purpose. More complicated gain profiles, lineshapes, and other material properties can easily be incorporated into our approach as well. SALT can, e.g., be coupled to a diffusion equation in order to model the migration of excited atoms in molecular-gas lasers [218, 219]. Based on the mathematical relation of the multimode lasing equations to incoherent vector solitons, we believe that numerical methods commonly used in soliton theory can also be adopted to efficiently solve the multimode SALT equations. Another intriguing direction of research is the development of a more systematic approach to modeling lasers with degenerate linear modes, which requires a technique to evaluate the stability of the solution and evolve an unstable mode to a stable mode. Finally, many refinements are possible to the numerical methods, such as efficient iterative solvers and preconditioners for the Newton iterations of the lasing modes or criteria to alternate between systematic contour-integral evaluation and simpler Newton-inverse tracking of the non-lasing modes. In this sense our approach has more in common with standard sparse discretization methods used to solve other nonlinear PDEs than the CF-basis approach (which is specialized to the SALT problem) and thus opens the door for more outside researchers and numerical specialists to study lasing problems.

Chapter 4

Degenerate modes in SALT

In chapter 3, we presented a direct method to solve the equations of steady-state ab-initio lasing theory (SALT). So far, most cases in which SALT has been applied has dealt with either single lasing modes or multimode regimes in which frequencies are far apart. This is necessary for SALT because the stationary-inversion approximation requires that any beating-terms average out over time . When frequencies are close but not exactly degenerate, there is non-negligible beating and SALT is invalid. However, when two lasing modes are *exactly* degenerate, it turns out that SALT is still perfectly valid, because there is an exact steady-state solution of the MB equations (for a single-mode pair), provided that interference between the two degenerate modes is taken into account. Of course, it is possible that a degeneracy in the linear regime may split in the presence of the laser nonlinearity above threshold. However, if a degeneracy persists (and we have consistently observed this for symmetry-induced degeneracies), our method will find it. The literature on degenerate lasing modes has almost invariably dealt with whispering-gallery modes in microdisks and ring resonators [31,32,220,221]. Many of these earlier works discussed the stability of travelling-wave modes in ring resonators under perturbations that break the symmetry [222–226]. A very limited number of other works on degenerate and nearly-degenerate lasing modes exist, such as Ref. [227]. However, the microdisk is only a single example of a setting where one can find degenerate resonant modes that can lase: there is a great variety of other symmetric geometries where degeneracies can occur [228]. So far, the problem of degenerate modes in lasers has not been studied systematically for the general case.

The Maxwell–Bloch equation describing the inversion is

$$\dot{D} = D_0 - D - D\sum_{\mu\nu}\left(\mathbf{E}_\mu \cdot \mathbf{E}_\nu^\star e^{i(\omega_\nu-\omega_\mu)t} - cc\right). \tag{4.1}$$

Currently, SALT assumes that all distinct modes have distinct frequencies, i.e. $\omega_\mu \neq \omega_\nu$ when $\mu \neq \nu$, which gives the stationary-inversion expression

$$D = \frac{D_0}{1+\sum|\mathbf{E}_\mu|^2}. \tag{4.2}$$

when higher-frequency $\omega_\mu - \omega_\nu$ $(\nu \neq \mu)$ terms are dropped. However, when there are degeneracies, Eq. 4.1 will have terms of the form $\mathbf{E}_\mu \cdot \mathbf{E}_\nu^\star$ where $\mu \neq \nu$, since $\omega_\mu = \omega_\nu$ and one can no longer drop the $e^{i(\omega_\mu-\omega_\nu)t}$ term. The correct expression for the stationary inversion will then be

$$D = \frac{D_0}{1+\sum'\mathbf{E}_\mu \cdot \mathbf{E}_\nu^\star}, \tag{4.3}$$

where $\sum'$ indicates a summation over all μ and ν for which $\omega_\mu = \omega_\nu$, not just for $\mu = \nu$. To illustrate the difference between the two, we examine a case in which there are three lasing modes, two of which are degenerate with each other $(\omega_1 = \omega_2 \neq \omega_3)$. Eq. 4.2 will have

$$|\mathbf{E}_1|^2 + |\mathbf{E}_2|^2 + |\mathbf{E}_3|^2$$

in the denominator, while eq. 4.3 will have

$$|\mathbf{E}_1 + \mathbf{E}_2|^2 + |\mathbf{E}_3|^2. \tag{4.4}$$

From Eq. 4.4 we see that the degenerate pair can be thought of as a single-mode that is a superposition of $\mathbf{E}_1$ and $\mathbf{E}_2$. This equivalence means that the solution to the lasing degenerate problem can be thought of in two equivalent pictures. First, we can think of the linear combination $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ as a single mode that satisfies the equation

$$\begin{aligned} -\nabla \times \nabla \times \mathbf{E} &= \omega^2 \left(\varepsilon_c + DF\gamma(\omega)H\right)\mathbf{E} \\ H &\equiv \frac{1}{1+|\mathbf{E}|^2}, \end{aligned} \tag{4.5}$$

where ε_c is the "cold cavity" permittivity. However, if there were originally a degenerate pair of modes at threshold and we only have one lasing on pole on the real-ω axis, it may be unstable if the other pole is above the real axis. Second, we can think of the two modes as *separately* satisfying the equations

$$\begin{aligned} -\nabla \times \nabla \times \mathbf{E}_\mu &= \omega^2 \left(\varepsilon_c + DF\gamma(\omega)H\right)\mathbf{E}_\mu \\ H &\equiv \frac{1}{1+|\mathbf{E}_1+\mathbf{E}_2|^2}. \end{aligned} \tag{4.6}$$

Eq. 4.5 is a less strict condition than Eq. 4.6. Not all solutions that satisfy the former satisfy the latter. If a lasing mode $\mathbf{E}$ satisfies the former there may be a second pole above the real-ω axis, rendering the lasing mode unstable to the second pole coming in. However, if a pair of solutions satisfy Eq. 4.6, then both degenerate modes are accounted for and lie on the real-ω axis, and there is at least no instability due to the degeneracy. We now describe this idea in detail.

Prior to lasing, suppose that we have a 2-fold degenerate solution, corresponding to a double pole in the Green function. As the gain increases, and even when the system passes threshold and becomes nonlinear, poles can shift (and degeneracies may split) but poles do not appear or disappear discontinuously, so we should always expect there to be two poles (in the linearized Green's function around the SALT solution) arising from the original degenerate pair. Given this fact, if we solve the single-mode SALT equations as in Eq. 4.5, there is the danger that the other pole is pushed above the real-ω axis, in which case the lasing solution is unstable. We given an example of this (an unstable "standing-wave" SALT solution) in Sec. 4.1.5. On the other hand, if we find a solution of the two-mode SALT equations as in Eq. 4.6, then by construction we have placed both poles together on the real-ω axis and the degeneracy is not a source of instability (and the overall stability of the Maxwell–Bloch equations can be checked as in Ref. [229]). However, Eq. 4.6 has a drawback: the hole-burning term now depends on the relative phase of $\mathbf{E}_1$ and $\mathbf{E}_2$. In the original SALT equations, even for multimode problems, the phase was irrelevant and was chosen arbitrarily in order to obtain a solvable system of equations. If we remove the arbitrary phase choice, our equations (from chapter 3) become undetermined, so in this chapter we will need to devise a modified approach to solve Eq. 4.6. Equivalently, if we solve the single-mode equation (Eq. 4.5) but simultaneously constrain the other pole (the linearly independent degenerate partner) to be degenerate with the lasing pole, then we will effectively have solved Eq. 4.6, and in the following sections we will explain how to implement this constraint.

Overall, our approach to solving for degenerate SALT solutions has three conceptual pieces: (i) using perturbation theory to find the approximate solution near threshold in Sec. 4.1, (ii) solving the nonlinear single-mode problem 4.5 above threshold in cases where symmetry guarantees the degeneracy, and (iii) restoring the degeneracy in cases where it is artificially broken (e.g. by the discretization of space in the numerical solver) in Sec. 4.2.2. As described in chapter 3, we can employ Newton's method to solve the SALT equations because we can continuously vary the solution starting from the linear solution at threshold, which guarantees convergence if a small enough increment of the pump strength is chosen (since Newton always converges whenever the initial guess is sufficiently close to the root). In the case where the threshold solution is degenerate, however, in order to obtain an appropriate initial guess for the Newton iteration we must choose the *correct* linear combination of the degenerate modes. We find this linear combination *analytically* in Sec. 4.1 by employing perturbation theory in the pump strength around threshold. Furthermore, we analyze the stability, and we show that mirror-symmetric standing-wave solutions are always unstable, and that "circulating" solutions (like $e^{im\phi}$, but more generally a $\psi_1 + i\psi_2$ combination of orthogonal standing-wave solutions) are the correct (stable) choice in typical $C_{n\mathrm{v}}$ symmetry groups. Given this initial solution, in cases where the degeneracy arises from symmetry it turns out that the above-threshold nonlinear solutions preserve the degeneracy. However, a problem arises when the Maxwell-Bloch equations are discretized (e.g. with finite-difference or finite-element methods): often, the discretization will break the symmetry and hence the degeneracy, preventing us from finding a degenerate solution to Eqs. 4.5 or 4.6. For example, a finite-difference grid typically has square symmetry, but breaks the sixfold symmetry of a hexagonal cavity as in the photonic-crystal example from chapter 3. To solve this problem, we describe a general and convergent technique in Sec. 4.2.2 to restore the degeneracy: we find the *smallest* perturbation $\delta\varepsilon(\mathbf{x})$ that coalesces the poles (with only slightly more computational effort than the ordinary SALT solver). For simplicity, most of our discussion will focus on the case of a single degenerate lasing mode. Once this case is solved, however, we show in Sec. 4.5 that the extension to multiple degenerate and non-degenerate lasing modes is straightforward, and is closely analogous to standard multi-mode SALT.

4.1 Threshold perturbation theory

Because degenerate pairs of lasing modes can always be thought of as a single-mode superposition, we solve the single mode case slightly above threshold. The equation is

$$\begin{aligned} 0 &= -\nabla \times \nabla \times \mathbf{E} + \omega^2 \tilde{\varepsilon} \mathbf{E} && (4.7) \\ \tilde{\varepsilon} &\equiv \varepsilon_c + \frac{D_0 \gamma(\omega) F}{1 + |\mathbf{E}|^2} && (4.8) \\ \gamma(\omega) &\equiv \frac{\gamma_\perp}{\omega - \omega_a + i\gamma_\perp}. \end{aligned}$$

Suppose that at threshold ($D_0 = D_0^t$), the solutions are $\mathbf{E}_1^t$ and $\mathbf{E}_2^t$, both with frequency ω_t and zero amplitude. Now perturb the pump to $D_0 = D_0^t\,(1+d)$, and we expect the single-mode solution to be

$$\begin{aligned} \mathbf{E} &= \sqrt{d}\left(a_1 \mathbf{E}_1^t + a_2 \mathbf{E}_2^t\right) + \delta\mathbf{E} \\ \omega &= \omega_t + \delta\omega. \end{aligned}$$

Insert into Eq. 4.8 and we have

$$\tilde{\varepsilon} = \varepsilon_c + D_0^t \gamma(\omega_t) F \left[1 + d - d \left| \sum a_\eta \mathbf{E}_\eta^t \right|^2 + \frac{\gamma'(\omega_t)}{\gamma(\omega_t)} \delta\omega \right] + \mathcal{O}(d^2).$$

Now multiply both sides of Eq. 4.7 by $\mathbf{E}_\mu^t$ and integrate. Note that

$$\int \mathbf{E}_\mu^t \cdot \nabla \times \nabla \times \delta\mathbf{E} = \int \mathbf{E}_\mu^t \cdot \left[\omega^2 \varepsilon_c + D_0^t \gamma(\omega_t) F \right] \delta\mathbf{E}.$$

This is because the operators are complex-symmetric and can be transposed to act to the left, annihilating $\mathbf{E}_\nu^t$ because it solves the SALT equation at threshold. Hence, all $\delta\mathbf{E}$ terms vanish after integration, and we have

$$0 = \sum_\nu \left[\frac{\delta\omega}{d} \left(A_{\mu\nu} + \Gamma B_{\mu\nu} \right) + B_{\mu\nu} - \sum_{\rho\eta} C_{\rho\eta\mu\nu} a_\rho a_\eta^\star \right] a_\nu, \tag{4.9}$$

which is a 2×2 nonlinear eigenproblem in the unknown eigenvector a_μ and eigenvalue $\delta\omega$, in terms of the known coefficients:

$$\begin{aligned} \Gamma &\equiv \frac{2}{\omega_t} + \frac{\gamma'(\omega_t)}{\gamma(\omega_t)}, \\ A_{\mu\nu} &\equiv \frac{2}{\omega_t D_0^t \gamma(\omega_t)} \int \mathbf{E}_\mu^t \cdot \varepsilon_c \mathbf{E}_\nu^t, \\ B_{\mu\nu} &\equiv \int \mathbf{E}_\mu^t \cdot F \mathbf{E}_\nu^t, \\ C_{\rho\eta\mu\nu} &\equiv \int \left(\mathbf{E}_\rho^t \cdot \mathbf{E}_\eta^{t\star} \right) \left(\mathbf{E}_\mu^t \cdot F \mathbf{E}_\nu^t \right). \end{aligned} \tag{4.10}$$

4.1.1 Existence

We now show that only certain linear combinations of the threshold modes solve the single-mode SALT equation. Eq. 4.9 is an eigenvalue problem that is linear in the eigenvalue $\delta\omega$ (which in this case is constrained to be real) and *cubic* in the components of the eigenvector a_ν. In general, Eq. 4.9 could be solved numerically by a low-dimensional root-finding algorithm. In fact, it has the same structure (number of unknowns) as the SALT equation (with two "spatial" degrees of freedom) and can therefore be solved with the same techniques as chapter 3 (but much more quickly). In practice, however, realistic degeneracies will almost always arise from symmetry, and in this case we can solve Eq. 4.9 analytically. (Degeneracies that do *not* arise from symmetry are known as "accidental" [117, 118], and because accidental degeneracies are broken by generic perturbations they will generally not survive above threshold unless there is some other constraint that enforces the degeneracy as in Appendix A.) In this case, $\mathbf{E}_1$ and $\mathbf{E}_2$ at threshold are "partner functions" [117] of an irreducible representation of the symmetry group, and several of the integrals in Eq. 4.9 vanish by symmetry.

In particular, the most common symmetry groups exhibiting degeneracies (via two-dimensional irreducible representations [117, 118]) are n-fold rotational symmetries combined with mirror symmetries, the $C_{n\text{v}}$ groups: the $C_{\infty\text{v}}$ group of the circle, the $C_{3\text{v}}$ group of the equilateral triangle, the $C_{4\text{v}}$ group of the square, and so on. (The photonic-crystal cavity of chapter 3 has $C_{6\text{v}}$ symmetry.) For these symmetry groups, there is always a mirror symmetry plane such that $\mathbf{E}_1$ is even and $\mathbf{E}_2$ is odd (or vice versa). The reason is that the $C_{n\text{v}}$ group is a subgroup of $C_{\infty\text{v}}$, and the latter has partner functions $\sin m\phi$ and $\cos m\phi$, which form 2d irreps of $C_{\infty\text{v}}$ for all $m > 0$. Since $\sin m\phi$ and $\cos m\phi$ always have opposite mirror symmetry about the line $\phi = 0$, and the 2d irreps of $C_{n\text{v}}$ induce 2d irreps of $C_{\infty\text{v}}$ and use the same partner functions [230], we can always choose the line with angle $\phi = 0$ such that the partner functions of $C_{n\text{v}}$ have opposite mirror symmetries. For example, Figs. 4-1 and 4-2 show examples of degenerate pairs from a metallic square cavity and a dielectric cylinder satisfying opposite mirror conditions. Hence, those integrals in Eq. 4.9 with an *odd* number of both 1's and 2's are zero. Eq. 4.9 then becomes

$$C_{2112} a_1^\star a_2^2 = \left[\frac{\delta\omega}{d} \left(A_{11} + \Gamma B_{11} \right) + B_{11} - C_{1111} \left| a_1 \right|^2 - \left(C_{2211} + C_{1212} \right) \left| a_2 \right|^2 \right] a_1. \quad (4.11)$$

(All equations in this section hold also when 1's and 2's are exchanged.) We list all

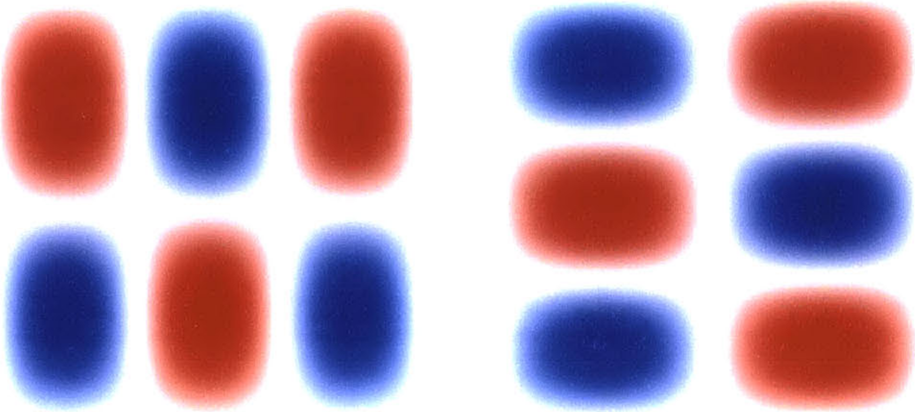


Figure 4-1: Threshold modes $\mathbf{E}_1^t = \psi_{32}\hat{\mathbf{z}}$ and $\mathbf{E}_2^t = \psi_{23}\hat{\mathbf{z}}$ for metallic square. The modes are $\frac{\pi}{2}$ rotations from each other, and this operation is also exact symmetry of the discretized grid, so there is an exact degeneracy even for the numerical solution.

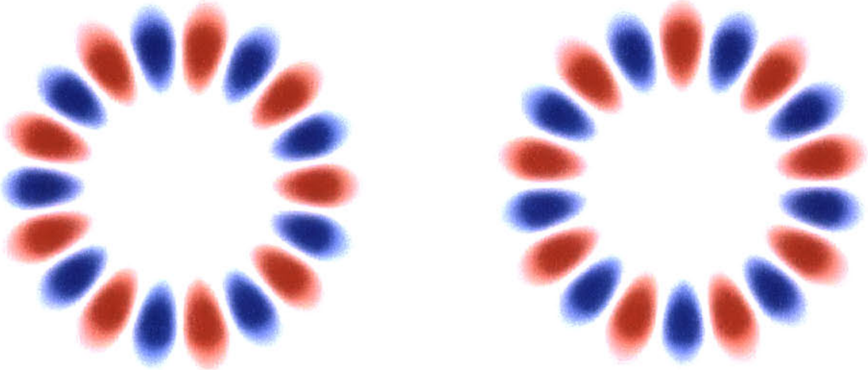


Figure 4-2: Odd-m threshold modes for dielectric cylinder. The real part of $\mathbf{E}$ has been plotted. Like in the metallic square, the modes are $\frac{\pi}{2}$ rotations from each other so there is an exact degeneracy even for the numerical solution. This is true for all odd-m whispering gallery modes, but not for even-m modes. We treat the latter in Sec. 4.2.

the solutions of this equation. First, the a_1-only solution is

$$\begin{aligned} |a_1|^2 &= \frac{\frac{\delta\omega}{d}\left(A_{11}+\Gamma B_{11}\right)+B_{11}}{C_{1111}} \\ a_2 &= 0 \\ \frac{\delta\omega}{d} &= -\left(\mathrm{Im}\frac{A_{11}+\Gamma B_{11}}{C_{1111}}\right)^{-1}\mathrm{Im}\frac{B_{11}}{C_{1111}}, \end{aligned} \tag{4.12}$$

where the last line can be obtained from noting that the first line is real.

The case in which both a_μ are nonzero can be examined by writing $a_\mu = |a_\mu| e^{i\theta_\mu}$, and Eq. 4.11 becomes

$$C_{1111}|a_1|^2 + \left[C_{2211}+C_{1212}+C_{2112}e^{2i(\theta_2-\theta_1)}\right]|a_2|^2 = \frac{\delta\omega}{d}\left(A_{11}+\Gamma B_{11}\right)+B_{11}.$$

and the solution is

$$|a_\mu|^2 = \sum_\nu \left(G^{-1}\right)_{\mu\nu}\left[\frac{\delta\omega}{d}\left(A_{\nu\nu}+\Gamma B_{\nu\nu}\right)+B_{\nu\nu}\right] \tag{4.13}$$

where we define the matrix

$$G \equiv \begin{pmatrix} C_{1111} & C_{2211}+C_{1212}+C_{2112}e^{2i(\theta_2-\theta_1)} \\ C_{1122}+C_{2121}+C_{1221}e^{-2i(\theta_2-\theta_1)} & C_{2222} \end{pmatrix}. \tag{4.14}$$

The relative phase $\theta_2-\theta_1$ and frequency shift $\delta\omega$ can then be determined as the real numbers that result in the Eq. 4.13 being real and positive. (G is almost always invertible, except for the case of $e^{2i(\theta_2-\theta_1)}=1$ for a geometry with $C_{\infty\mathrm{v}}$ symmetry, such as the uniform cylinder which we will treat later). These conditions almost always mean that there are only discrete sets of linear combinations that satisfy the single-mode SALT equations; i.e. there is not a continuous manifold of linear combinations (not including the overall phase choice or an arbitrary continuous rotation in $C_{\infty\mathrm{v}}$).

4.1.2 Stability

Not all solutions are stable, however. To determine stability of a lasing solution with coefficients a_ρ, we need to solve for the "passive" mode arising from the other degenerate partner: it is necessary to solve the eigenvalue problem

$$0 = \sum_\nu \left[\frac{\delta\omega'}{d} \left(A_{\mu\nu} + \Gamma B_{\mu\nu} \right) + B_{\mu\nu} - \sum_{\rho\eta} C_{\rho\eta\mu\nu} a_\rho a_\eta^\star \right] a'_\nu, \tag{4.15}$$

which resembles Eq. 4.9 except for two differences: the eigenvalue $\delta\omega'$ is now allowed to be complex, and the equation is *linear* in the eigenvector coefficients a'_ν because the lasing coefficients a_ρ are already known. By inspection, $a'_\nu = a_\nu, \delta\omega' = \delta\omega$ is clearly a solution to Eq. 4.15. The question is whether the *other* eigenvalue $\delta\omega'$ is above the real axis. If so, then the lasing coefficients a_ρ give an *unstable* linear combination. Another (stable) possibility would be if $\delta\omega'$ were below the real axis, but have not encountered any such cases. This situation is not expected to occur frequently, because it would require a passive pole to hit the threshold and then immediately turn back, which is unlikely given the analytical nature of the SALT equations. A third possibility is for $\delta\omega'$ to lie on the real axis, but with value differing from $\delta\omega$. This situation occurs if the threshold degeneracy was forced (accidental) by tuning (e.g. the geometry), with the splitting arising from the spatial hole-burning breaking the degeneracy. For such small splittings, SALT (stationary inversion) breaks down and new physics is needed. We discuss the details and implications of this case in a later section. The final case is if the second eigenvalue is *also* $\delta\omega' = \delta\omega$. This case is the stable two-mode degenerate lasing solution: a passive pole coincides with the lasing pole.

For mirror-symmetric threshold modes, Eq. 4.15 becomes

$$\left(C_{2112} a_2 a_1^\star + C_{1212} a_1 a_2^\star \right) a'_2 = \left[\frac{\delta\omega'}{d} \left(A_{11} + \Gamma B_{11} \right) + B_{11} - C_{1111} \left| a_1 \right|^2 - C_{2211} \left| a_2 \right|^2 \right] a'_1. \tag{4.16}$$

We now show an important result about purely standing waves with mirror symmetry, in which $a_2 = 0$ (or alternatively $a_1 = 0$), which has the solution given in Eq. 4.12.

Inserting these into Eq. 4.16 and its counterpart with 1's and 2's exchanged, we have

$$0 = \frac{\delta\omega' - \delta\omega}{d} (A_{11} + \Gamma B_{11}) a_1'$$
$$0 = \left[\frac{\delta\omega'}{d} (A_{22} + \Gamma B_{22}) + B_{22} - \frac{C_{1122}}{C_{1111}} \left(\frac{\delta\omega}{d} [A_{11} + \Gamma B_{11}] + B_{11} \right) \right] a_2'$$

The trivial eigenpair has $\delta\omega' = \delta\omega$ and $a_2' = 0$, but the other eigenvalue is

$$\frac{\delta\omega'}{d} = \frac{\frac{C_{1122}}{C_{1111}} \left[\frac{\delta\omega}{d} (A_{11} + \Gamma B_{11}) + B_{11} \right] - B_{22}}{A_{22} + \Gamma B_{22}}. \tag{4.17}$$

with $a_1' = 0$. It appears that this eigenvalue is always above the real axis (we can confirm this fact on a case-by-case basis, but a general proof is beyond the scope of the thesis). Hence, one of the standing modes will never by itself be the stable lasing field because it will always be unstable to the second mode (which has the opposite mirror symmetry) coming in. In the absence of a general proof that $\text{Im}\delta\omega' > 0$, one can easily check Eq. 4.17 as a small part of solving degenerate SALT problems. The intuitive reason why standing-wave modes are unstable is known from the ring-resonator case [32]: the nodes in the standing wave represent unused gain that the degenerate partner "wants" to grow into. There are, however, exceptions: the authors of Ref. [229] found interesting examples of single-mode solutions with passive poles *above* the real axis that were shown to be stable from both stability analysis and brute-force time-domain simulation. These examples suggest many new possibilities for stable single-mode lasing, and we plan to examine them in future work.

4.1.3 Stability of circulating solutions

In the case of the ring resonator, it is known that the "circulating" solutions $e^{\pm im\phi}$ are the stable solutions, and it turns out that a close analogue of this result holds for any C_{nv} symmetry group. As above, we use $\mathbf{E}_1$ and $\mathbf{E}_2$ to denote the degenerate eigenfunctions (at threshold) chosen to be even/odd with respect to one of the mirror planes, analogous to $\cos m\phi$ and $\sin m\phi$ in the $C_{\infty v}$ group. In terms of these eigenfunctions, the generalized "circulating" modes are $\mathbf{E}_\pm \equiv \mathbf{E}_1 \pm i\mathbf{E}_2$, and it turns out that writing the solution in terms of this always yields a stable lasing mode with a degenerate passive mode.

Before we show stability, we must first check that the circulating modes are solutions. One possible method is to work in the basis $\mathbf{E}_1$ and $\mathbf{E}_2$ and exploit the mirror symmetry so that many matrix elements $A_{\mu\nu}$, $B_{\mu\nu}$, and $C_{\rho\eta\mu\nu}$ vanish by symmetry. This method has two disadvantages: first, in cases in which $\mathbf{E}_1$ and $\mathbf{E}_2$ are not simple $\frac{\pi}{2}$ rotations of each other (for example, dipole partners in $C_{n\mathrm{v}}$ for any n that is not a multiple of 4) the relation between matrix elements such as C_{1111} and C_{2222} is more complicated. Second, when $\mathbf{E}_1$ and $\mathbf{E}_2$ are no longer mirror symmetric (as we will see in Sec. 4.5, in the case where gain is already chiral, with C_n symmetry rather than $C_{n\mathrm{v}}$), the matrix elements that usually vanish by symmetry no longer do.

Fortunately, working in the basis $\mathbf{E}_\pm$ greatly simplifies the calculations. Analogous to Sec. 4.1.1, we look for the solution $a_+\mathbf{E}_+ + a_-\mathbf{E}_-$ with $a_+ \neq 0$ and $a_- = 0$ or vice versa. First, we see that $A_{\pm\pm}$ and $B_{\pm\pm}$ both involve integrals of $\mathbf{E}_\pm \cdot \mathbf{E}_\pm = \mathbf{E}_\pm \cdot (\mathbf{E}_\pm^\star)^\star$. Since $\mathbf{E}_\pm^\star$ transforms as the irrep with opposite chirality from $\mathbf{E}_\pm$, and partner functions of different irreps have zero (conjugated) inner product due to the great orthogonality theorem [117,118], both $A_{\pm\pm}$ and $B_{\pm\pm}$ vanish. Inserting these facts and $a_- = 0$ into Eq. 4.9, we have the two equations ($\mu = 1, 2$)

$$\begin{aligned} 0 &= C_{++++} \\ 0 &= \frac{\delta\omega}{d}\left(A_{-+} + \Gamma B_{-+}\right) + B_{-+} - C_{++-+}\left|a_+\right|^2 . \end{aligned}$$

The first equation is satisfied because $C_{++++} = \int F \left|\mathbf{E}_+\right|^2 \mathbf{E}_+ \cdot \left(\mathbf{E}_+^\star\right)^\star = 0$ by the great orthogonality theorem again (since $\left|\mathbf{E}_+\right|^2$ has C_n symmetry; equivalently, one could use product representation theory [117]). The real and imaginary parts of the second equation exactly specify the values of the two real unknowns $\delta\omega$ and $|a_+|$.

Now, to show that $\mathbf{E}_+$ is stable, we can exploit some group theory combined with a recent result regarding degeneracies in chiral structures [119] that is reviewed and generalized in appendix A. In representation-theory terms, the circulating modes $\mathbf{E}_\pm$ diagonalize the rotation operations R_n^k with eigenvalues $e^{\mp 2\pi i m k/n}$, or equivalently are partners of 1d irreducible representations (irreps) of the C_n subgroup (the rotation symmetries without the mirror planes). In order to understand the passive mode of the $\mathbf{E}_+$ lasing solution above threshold, the key point is to realize that $|\mathbf{E}_+|^2$ has C_n symmetry (because rotation is just a phase shift) but does *not* in general have $C_{n\mathrm{v}}$ symmetry: the intensity pattern of the radiation from a circulating mode is typically a "spiral" pattern (as shown e.g. in Sec. 4.1.5 for a dielectric-square cavity) that breaks the mirror symmetries. Naively, since the C_n group does not have 2d irreps [117, 118], we should not expect a degenerate passive mode, or any degeneracies for that matter. However, as was recently shown in Ref. [119] and is reviewed in appendix A, the C_n symmetry group combined with Lorentz reciprocity (the fact that the Maxwell operator is complex-symmetric for symmetric matrices ε and μ) means that the circulating modes are in fact degenerate. This degeneracy (which is technically "accidental" since it does not arise from symmetry alone [34]) guarantees stability of the $\mathbf{E}_-$ partner.

More specifically, as the pump strength varies continuously from $d = 0$ to $d > 0$ and C_n (but not $C_{n\mathrm{v}}$) symmetry is preserved, the solutions cannot jump discontinuously from one irrep of C_n to another. The $\mathbf{E}_+$ and $\mathbf{E}_-$ degenerate partners at threshold fall into the $D(k) = \exp\left(\mp\frac{2\pi imk}{n}\right)$ irreps of the C_n group and the above-threshold solutions therefore stay in these complex-conjugate irreps. As reviewed in appendix A , whenever there is an eigenfunction of the reciprocal Maxwell operator in a complex irrep $D(k)$ (such as the $\mathbf{E}_+ + \delta\mathbf{E}_+$ lasing solution), there is always a degenerate (equal-frequency) eigenfunction in the complex-conjugated irrep $\bar{D}(k)$. (Remarkably, these degenerate eigenfunction are not mirror flips of one another, because the degeneracy does not arise arise from geometric symmetry alone.) So, there must a passive pole coinciding with the lasing solution. Since the poles move continuously with pump strength, and at $d = 0$ $\mathbf{E}_+$ was degenerate with $\mathbf{E}_-$, this other passive pole must be the $\mathbf{E}_- + \delta\mathbf{E}_-$ solution (which is a partner of the conjugated irrep). Hence, the poles must stick together as the pump is increased continuously, even though the two solutions are no longer mirror images of one another.

Of course, there is also a mirror-image lasing solution that can be found, exactly as above, by starting with $a_+ = 0$. Because of the symmetry of the underlying problem, this is simply the mirror flip of the $\mathbf{E}_+ + \delta\mathbf{E}_+$ lasing solution. It is important not to confuse this mirror-flipped lasing solution with the $\mathbf{E}_- + \delta\mathbf{E}_-$ passive solution above, which is not the mirror flip of the lasing solution. The system is nonlinear, so the passive left-circulating solution of the right-circulating lasing mode is not a solution for the left-circulating lasing mode.

As a practical matter, in order to obtain the $\mathbf{E}_+$ and $\mathbf{E}_-$ solutions at threshold, one must remember that a numerical eigensolver will generally return a random an arbitrary superposition $\mathbf{E}$ of $\mathbf{E}_1$ and $\mathbf{E}_2$. But, given such a superposition, one can obtain $\mathbf{E}_+$ merely by acting the projection operator [117] of that irrep of the C_n group on $\mathbf{E}$:

$$\mathbf{E}_+ = \sum_{k=0}^{n-1} \exp\left(-\frac{2\pi i m k}{n}\right) R_n^k \mathbf{E},$$

where R_n is the operator rotating by $2\pi/n$. In a numerical grid that breaks the C_n symmetry as in Fig. 4-22, implementing R_n requires rotating the grid and then interpolating back to the original grid, but the small interpolation errors introduced by this process are irrelevant—for the Newton solver to converge, all we need is a solution that is sufficiently close to the "true" $\mathbf{E}_+$. The key point, as discussed in Sec. 4.1.1, is that the stable and unstable solutions above threshold are not connected by a continuous manifold, so if you are close to a stable solution then there is no nearby unstable solution for Newton to find. Note also that the choice of m in the projection depends on which irrep one is projecting into, and this can be determined easily by inspection of the $\mathbf{E}$: $m = 1$ for a dipole-like mode, $m = 2$ for quadrupole-like, etcetera. Alternatively, a completely automated procedure would be to simply try projecting for $m = 1, 2, \ldots, \text{floor}(\frac{n-1}{2})$: for all but the correct m, the projected function should be nearly zero.

Given $\mathbf{E}_+$ via the projection operator, the mirror flip of $\mathbf{E}_+$ is $\mathbf{E}_-$ (because at threshold these are partners of $C_{n\text{v}}$, not just C_n). From these two eigenfunctions, one can construct $\mathbf{E}_1 = \frac{\mathbf{E}_+ + \mathbf{E}_-}{2}$ and $\mathbf{E}_1 = \frac{\mathbf{E}_+ + \mathbf{E}_-}{2}$ if desired, e.g. for checking the instability of the standing-wave mode from Sec. 4.1.2. It is best to apply the perturbation theory to either even/odd or circulating partner functions, and not just to whatever random superposition a numerical eigensolver produces, because this basis leads to the simplest analytical results.

4.1.4 Obtaining the stable superposition directly from $\mathbf{E}_1$ and $\mathbf{E}_2$

In this section, we provide an alternative method to get the stable lasing superposition in terms of the standing partner functions $\mathbf{E}_1$ and $\mathbf{E}_2$. Although this treatment is redundant with that given in Sec. 4.1.3, it has the advantage of not requiring implementation of rotational operators (and the associated interpolation, as explained in Sec. 4.1.3) to construct the $\mathbf{E}_+$ and $\mathbf{E}_-$ circulating solutions—for mirror symmetric geometries, $\mathbf{E}_1$ and $\mathbf{E}_2$ can be obtained much more easily than $\mathbf{E}_\pm$ by using mirror-flip operations, which are much simpler to implement and require no interpolation. However, a disadvantage is that $\mathbf{E}_2$ obtained this way, without any further processing, is usually not normalized correctly [the inner products $(\mathbf{E}_i, \mathbf{E}_i)$ are different for $i = 1, 2$] and does not have the correct overall phase ($\mathbf{E}_2 = e^{i\theta} \frac{\mathbf{E}_+ - \mathbf{E}_-}{2i}$, where θ is some arbitrary phase found by the solver). Consequently, the results in this section are mainly valuable as a validity check on results obtained by the method of Sec. 4.1.3.

For a 2×2 generalized linear eigenvalue problem such as Eq. 4.15 to have degenerate eigenvalues, the matrix elements must individually satisfy the eigenequation, which is a stricter condition than the eigenequation itself. Hence, for the stable combination we have

$$0 = \frac{\delta\omega}{d} \left(A_{\mu\nu} + \Gamma B_{\mu\nu} \right) + B_{\mu\nu} - \sum_{\rho\eta} C_{\rho\eta\mu\nu} a_\rho a_\eta^\star \tag{4.18}$$

In practice, if we wanted the stable linear combination, we can simply solve this equation directly rather than finding all the solutions of Eq. 4.9 and categorizing their stability using Eq. 4.15. For mirror-symmetric partner functions, 4.18 becomes

$$0 = \frac{\delta\omega}{d}\left(A_{11} + \Gamma B_{11}\right) + B_{11} - C_{1111}\left|a_1\right|^2 - C_{2211}\left|a_2\right|^2 \tag{4.19}$$

$$0 = C_{1212} a_1 a_2^\star + C_{2112} a_2 a_1^\star, \tag{4.20}$$

Eq. 4.19 has the closed-form solution

$$\begin{pmatrix} |a_1|^2 \\ |a_2|^2 \end{pmatrix} = \frac{\begin{pmatrix} C_{2222} & -C_{2211} \\ -C_{1122} & C_{1111} \end{pmatrix}\left[\frac{\delta\omega}{d}\begin{pmatrix} A_{11} + \Gamma B_{11} \\ A_{22} + \Gamma B_{22} \end{pmatrix} + \begin{pmatrix} B_{11} \\ B_{22} \end{pmatrix}\right]}{C_{1111}C_{2222} - C_{2211}C_{1122}}. \tag{4.21}$$

From this expression, $\delta\omega$ can also be obtained: it is simply the real frequency shift that makes the right-hand side real, since by definition $|a_\mu|^2$ must be a real number. Eq. 4.20 gives the phase between the two coefficients for the stable linear combination.

Eqs. 4.20 and 4.21 are very most useful because they provide a guess for the Newton solver slightly above threshold. Once we find the two orthogonal modes at threshold, we form the linear combination provided by these equations and then use it as an initial guess. Since there is only a discrete set of solutions that solve the single-mode SALT equations, there is no threat of converging to a nearby unstable solution; as long as our initial guess is reasonably close, Newton will always converge to the stable linear combination.

We do not generally expect the intensity pattern of the stable lasing mode to be mirror symmetric unless ψ is real or can be separated into a complex factor that is unchanged under R_n and a real factor that changes under R_n. From this fact, we can conclude that most degenerate lasing pairs will exhibit a nonzero amount of chirality, not just in their mode profiles, but in their intensity patterns as well (with the exception of the whispering-gallery modes of $C_{\infty v}$, whose partner functions can be separated into a complex part that depends only on the radial coordinate, and hence is invariant under all rotations, and a azimuthal dependence that can be chosen to be entirely real, as in $\sin m\phi$ and $\cos m\phi$). High-Q pairs will only be slightly chiral because they field profiles can be chosen to be nearly real. This chirality will be very evident in low-Q degenerate lasing modes, as we will see later in Sec. 4.1.5, because their field patterns have a significant amount of both real and complex parts.

4.1.5 Threshold perturbation examples

Now, we illustrate the ideas of threshold perturbation theory with several simple examples of symmetric geometries with degeneracies: a metallic square, a dielectric cylinder, and a dielectric square. We first present the threshold modes (solutions of a linear problem at threshold) which are exactly degenerate. Then, we compute the matrix elements $A_{\mu\nu}$, $B_{\mu\nu}$, and $C_{\rho\eta\mu\nu}$ in the perturbation theory. Next, we solve Eq. 4.11 for the existing lasing modes, then use Eq. 4.16 to find the passive poles and check their stability. Alternatively, we apply Eq. 4.21 to $\mathbf{E}_1$ and $\mathbf{E}_2$ and check that it produces the circulating modes predicted in Sec. 4.1.3. We compare the results obtained by perturbation theory to numerical solution of the SALT equation.

Metallic square

For example, consider TM modes ($\mathbf{E} = \psi\hat{\mathbf{z}}$) in a metallic $L_x \times L_y$ rectangle with uniform gain $F(\mathbf{x}) = 1$ and uniform lossy dielectric $\varepsilon_c(\mathbf{x}) = 1 + i\sigma$. For $\sigma = 0$ and $D_0 = 0$, the modes and frequencies would be

$$\begin{aligned} \psi_{mn} &= \sin\left(\frac{m\pi x}{L_x}\right)\sin\left(\frac{n\pi y}{b}\right) \\ \omega_{mn} &= \pi\sqrt{\frac{m^2}{L_x^2}+\frac{n^2}{L_y^2}}. \end{aligned} \tag{4.22}$$

Choose $L_x = L_y = 1$, and we have $\omega_{mn} = \omega_{nm}$. Now we choose a pair $m \neq n$, and set the gain center to be $\omega_a = \omega_{mn}$ (the results in this section are independent of the choice of m and n, as long as they are different integers). The profiles are shown in Fig. 4-1 for $m = 3$, $n = 2$. Since the gain and loss profiles are the same, the threshold for this pair of modes will be $D_0^t = \sigma$, with frequency $\omega_t = \omega_{mn}$. We also have $\gamma(\omega_t) = -i$ and $\gamma'(\omega_t) = 1/\gamma_\perp$, which result in $\Gamma = \frac{2}{\omega_t} + \frac{i}{\gamma_\perp}$. The threshold modes are $\mathbf{E}_1^t = \psi_{mn}\hat{\mathbf{z}}$ and $\mathbf{E}_2^t = \psi_{nm}\hat{\mathbf{z}}$. Note that the following analysis can alternatively be done with any linear combination of these two modes as a basis, as long as we use the general equations 4.9, 4.15, and 4.18. However, this *particular* choice is convenient because it falls under the category of degenerate pairs satisfying mirror symmetries.

The nonzero matrix elements in the threshold perturbation theory are

$$\begin{aligned} A_{11} &= \frac{1}{2\omega_t}\left(\frac{i}{\sigma}-1\right) \\ B_{11} &= \frac{1}{4} \\ C_{1111} &= \frac{9}{64} \\ C_{1122} = C_{1212} = C_{1221} &= \frac{1}{16}, \end{aligned}$$

and the same equations hold with 1's and 2's exchanged. Eq. 4.11 then becomes

$$a_1^\star a_2^2 = \left[\frac{i\Gamma_0\delta\omega}{d} + 4 - \frac{9}{4}\left|a_1\right|^2 - 2\left|a_2\right|^2\right] a_1, \tag{4.23}$$

where we have defined $\Gamma_0 \equiv \frac{8}{\omega_t\sigma} + \frac{4}{\gamma_\perp}$. We now solve this equation to obtain the lasing solutions, including unstable ones. First, if $a_2 = 0$, then $\delta\omega = 0$ and $16 = 9\left|a_1\right|^2$, giving the lasing mode $\mathbf{E} = \sqrt{d}\frac{4}{3}\psi_{mn}\hat{\mathbf{z}}$ (a second lasing mode with $\mathbf{E} = \sqrt{d}\frac{4}{3}\psi_{nm}\hat{\mathbf{z}}$ can trivially be obtained by setting $a_1 = 0$ instead). From Eq. 4.17, we see that this mode is unstable to the second mode coming in, with a frequency eigenvalue $\delta\omega' = \frac{20id}{9\Gamma_0}$, as expected.

Now consider the case that both a_1 and a_2 are nonzero. Eq. 4.13 then gives

$$\begin{pmatrix} \left|a_1\right|^2 \\ \left|a_2\right|^2 \end{pmatrix} = \frac{\frac{i\Gamma_0\delta\omega}{d} + 4}{\frac{1}{16} - 4\cos(2\theta_2 - 2\theta_1)} \begin{pmatrix} \frac{1}{4} - e^{2i(\theta_2-\theta_1)} \\ \frac{1}{4} - e^{2i(\theta_1-\theta_2)} \end{pmatrix}.$$

For this expression to be real, we must have $\delta\omega = 0$ and $2\left(\theta_2 - \theta_1\right)$ must be an integer multiple of π. For even multiplies of π, we have $\left|a_\mu\right|^2 = \frac{16}{21}$, giving the lasing mode $\mathbf{E} = 4\sqrt{\frac{d}{21}}\left(\psi_{mn} \pm \psi_{nm}\right)\hat{\mathbf{z}}$, and for odd multiples of π we have $\left|a_\mu\right|^2 = \frac{16}{13}$, giving the lasing mode $\mathbf{E} = 4\sqrt{\frac{d}{13}}\left(\psi_{mn} \pm i\psi_{nm}\right)\hat{\mathbf{z}}$. Fig. 4-3 shows simulation results compared to theory for the modes from Fig. 4-1.

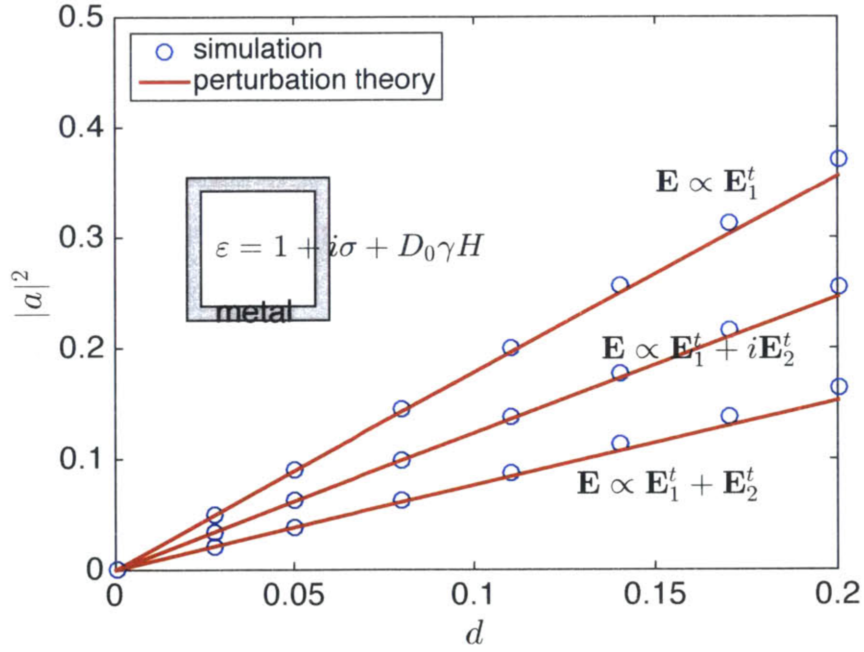


Figure 4-3: Lasing amplitudes for metallic square. Three types of lasing modes were used as initial guesses at threshold: a single standing (vertically and horizontally mirror-symmetric) mode lasing by itself, a circulating mode (two standing waves with a relative phase of i), and a "diagonal" standing-wave mode (diagonally mirror symmetric). The theoretical lines were obtained from Eq. 4.23: as expected, the perturbation theory is correct to first order in d.

We test the stability of these solutions using Eq. 4.16. Inserting the matrix elements, we have

$$(a_2 a_1^\star + a_1 a_2^\star)\, a_2' = \left[\frac{i\Gamma_0 \delta\omega'}{d} + 4 - \frac{9}{4}\left|a_1\right|^2 - \left|a_2\right|^2\right] a_1' \tag{4.24}$$

For $\mathbf{E} = 4\sqrt{\frac{d}{21}}\,(\psi_{mn} + \psi_{nm})\,\hat{\mathbf{z}}$, the other eigenvalue is $\delta\omega' = \frac{64id}{21\Gamma_0}$, with an eigenvector of $\binom{1}{-1}$. Hence, this mode is unstable to the second mode proportional to $\psi_{mn} - \psi_{nm}$ coming in. The unstable intensity patterns are shown in Fig. 4-5. For $\mathbf{E} = 4\sqrt{\frac{d}{13}}\,(\psi_{mn} + i\psi_{nm})\,\hat{\mathbf{z}}$, on the other hand, Eq. 4.24 reduces to $0 = \delta\omega' a$, so both eigenvalues vanish and this lasing mode is stable. This stability is guaranteed by symmetry (as described in appendix A) since the passive solution is a circulating mode of the opposite chirality. Fig. 4-4 shows simulation results compared to theory for passive eigenpairs of the lasing modes from Fig. 4-3. While the above is an exhaustive procedure that is guaranteed to find all lasing modes and categorize them into stable and unstable, in practice it is much easier to directly find the stable circulating superposition as in Sec. 4.1.3.

Dielectric cylinder

As another example, we consider an infinite cylinder with uniform "cold cavity" dielectric $\varepsilon_c = n^2$ and gain profile $F = 1$ inside $\rho < 1$ and air outside (we use cylindrical coordinates ρ, ϕ, z here). Here, we consider the example of z-invariant TM ($\mathbf{E} = \psi\hat{\mathbf{z}}$) solutions. The passive resonances are exactly solvable [115] and are given by $\psi^{\pm}_{Nm}(\rho, \phi) = e^{\pm im\phi} B_{Nm}(\rho)$, where

$$B_{Nm}(\rho) \equiv \begin{cases} J_m\left(n\omega_{Nm}\rho\right), & \rho < 1 \\ \eta_{Nm} H_m\left(\omega_{Nm}\rho\right), & \rho > 1 \end{cases}, \tag{4.25}$$

and J_m is a Bessel function of the first kind, H_m is an outgoing Hankel function, and η_{Nm} and ω_{Nm} are chosen so that both ψ and $\frac{\partial\psi}{\partial\rho}$ are continuous at ρ. For $D_0 = 0$, the poles ω_{Nm} will lie below the real axis. Now suppose that we choose ω_a and $\gamma_\perp$ such

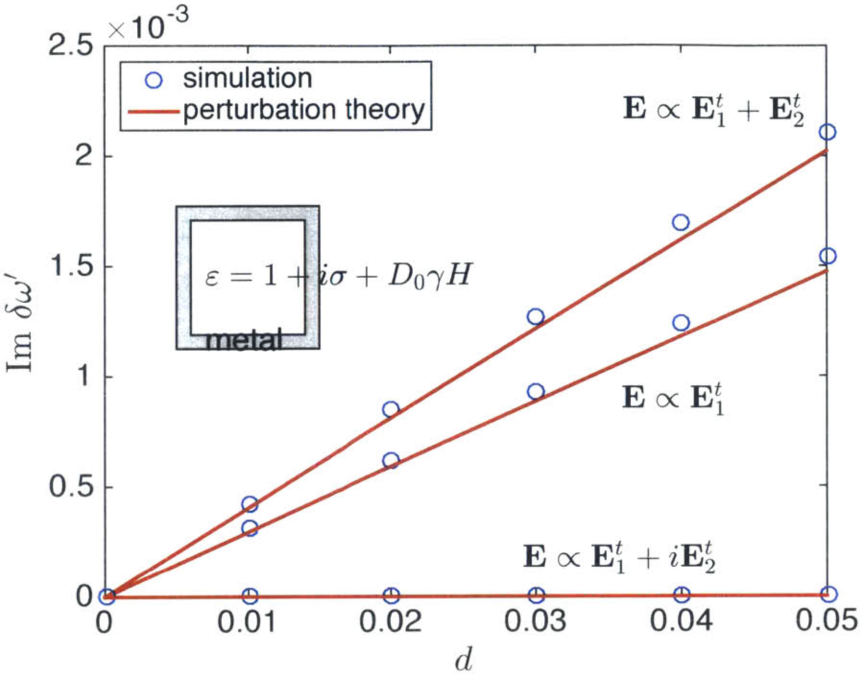


Figure 4-4: Imaginary part of the "passive" pole (degenerate partner of the lasing mode at threshold) for metallic square. For the simulation data points, the nonlinear problem for **E** was first solved with the initial guesses in Fig. 4-3, and the result was used in the spatial hole-burning term for a linear problem. Only the lasing mode proportional to $\mathbf{E}_1^t + i\mathbf{E}_2^t$ is stable. The theoretical lines were obtained from Eq. 4.24: as expected, the perturbation theory is correct to first order in d.

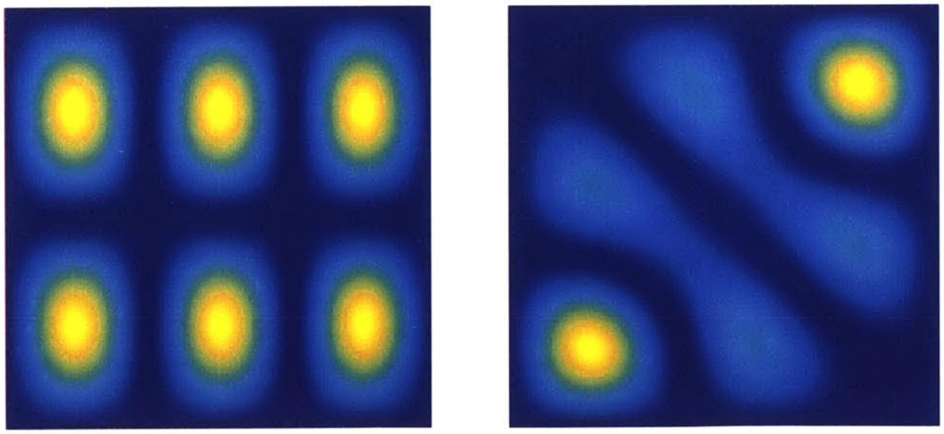


Figure 4-5: Intensity profiles for unstable lasing modes of metallic square. The intensity $\sqrt{d}\frac{4}{3}\psi_{32}$ (left) has nodes that make it unstable to a ψ_{23} mode. The intensity for $4\sqrt{\frac{d}{21}}\left|\psi_{32}+\psi_{23}\right|^2$ (right) has nodes that makes it unstable to a $\psi_{32}-\psi_{23}$ mode. Note that both intensity patterns break the symmetry that connected the two degenerate modes in Fig. 4-1.

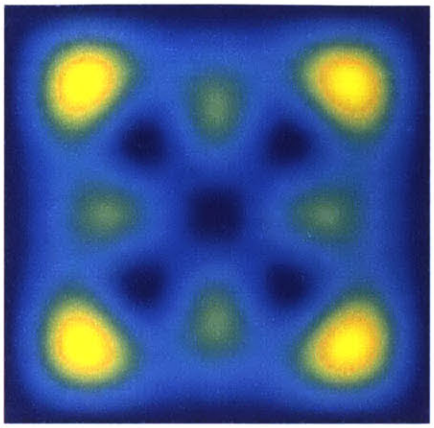


Figure 4-6: Intensity profile for stable lasing mode $4\sqrt{\frac{d}{13}}\,(\psi_{32} + i\psi_{23})$ of square. This pattern maintains the $C_{4\text{v}}$ symmetry of the original system that connected the two threshold modes, unlike those in Fig. 4-5.

that a certain ψ_{nm} lases first at some pump strength D_0^t, and $\omega_{Nm} = \omega_t$ hits the real axis. The threshold lasing mode will have the radial dependence given in Eq. 4.25, except with the index inside given by the complex number $n_t \equiv \sqrt{n^2 + D_0^t\gamma(\omega_t)}$.

To analyze the lasing modes above threshold, we choose $\mathbf{E}_1^t = B_{Nm}(\rho)\cos(m\phi)$ and $\mathbf{E}_2^t = B_{Nm}(\rho)\sin(m\phi)$. The nonzero matrix elements are then

$$\begin{aligned} A_{11} &= \frac{2n^2\beta_{\text{in}} + 2\beta_{\text{out}}}{\omega_t D_0^t\gamma(\omega_t)} \\ B_{11} &= \beta_{\text{in}} \\ C_{1111} &= 3\xi \\ C_{1122} = C_{1212} &= C_{1221} = \xi \end{aligned}$$

where the equalities still hold with 1's and 2's exchanged, and we have defined

$$\begin{aligned}\beta_{\text{in}} &\equiv \pi \int_0^1 \rho d\rho J_m \left(n_t\omega_t\rho\right)^2 \\ \beta_{\text{out}} &\equiv \pi \int_1^\infty \rho d\rho \eta_{Nm}^2 H_m \left(\omega_t\rho\right)^2 \\ \xi &\equiv \frac{\pi}{4} \int_0^1 \rho d\rho J_m \left(n_t\omega_t\rho\right)^2 \left|J_m \left(n_t\omega_t\rho\right)\right|^2 .\end{aligned}$$

We now find the lasing solutions, and test their stability. First, consider the pure cosine mode. From Eq. 4.12, we have the lasing mode

$$\begin{aligned}\mathbf{E} &= \hat{\mathbf{z}}\sqrt{\frac{\lambda d}{3}} B_{Nm} \cos m\phi \\ \frac{\delta\omega}{d} &= -\left(\text{Im}\frac{A_{11}+\Gamma\beta_{\text{in}}}{\xi}\right)^{-1} \text{Im}\frac{\beta_{\text{in}}}{\xi},\end{aligned} \tag{4.26}$$

where we have defined

$$\lambda \equiv \frac{\delta\omega}{\xi d}\left(A_{11}+\Gamma\beta_{\text{in}}\right) + \frac{\beta_{\text{in}}}{\xi}.$$

From Eq. 4.17, we see that there will be unstable pole with eigenfrequency

$$\frac{\delta\omega'}{d} = \frac{\delta\omega}{3d} - \frac{2\beta_{\text{in}}}{3(A_{11}+\Gamma\beta_{\text{in}})}. \tag{4.27}$$

We now turn to Eq. 4.13, which becomes

$$\lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 & 2+e^{2i(\theta_2-\theta_1)} \\ 2+e^{-2i(\theta_2-\theta_1)} & 3 \end{pmatrix} \begin{pmatrix} |a_1|^2 \\ |a_2|^2 \end{pmatrix} \tag{4.28}$$

By inspection, $2(\theta_2-\theta_1)$ must be an integer multiple of π, and the frequency shift is given again by Eq. 4.26. For even integer, the matrix G in Eq. 4.14 becomes singular, and we have $|a_1|^2+|a_2|^2 = \frac{\lambda}{3}$, giving the manifold of lasing modes $\mathbf{E} = \hat{\mathbf{z}}\sqrt{\frac{\lambda d}{3}} B_{Nm}(\rho)\cos(m\phi+\phi_0)$, where ϕ_0 is an arbitrary angle. These modes are just rotations of the purely standing-wave solutions above, so are unstable to a passive pole

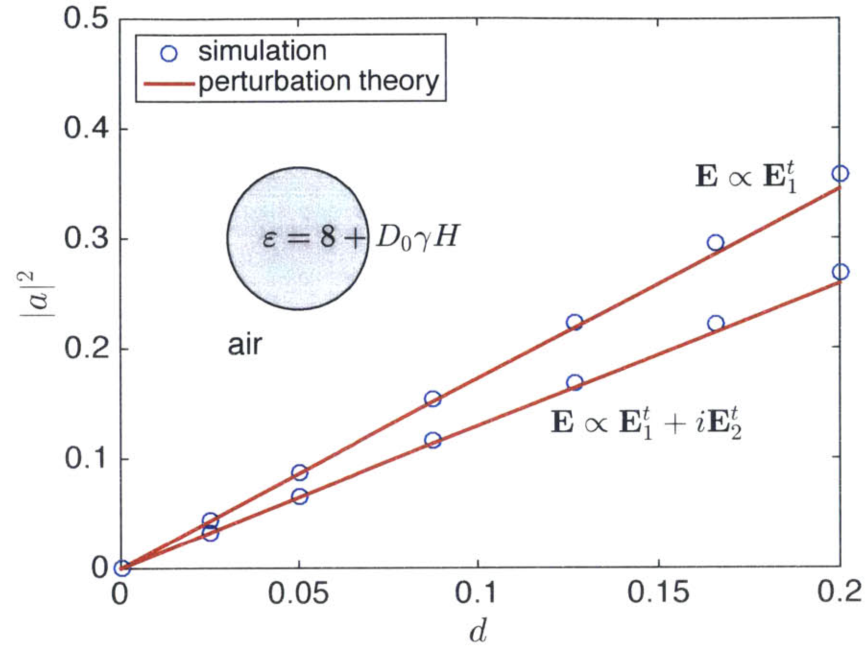


Figure 4-7: Lasing amplitudes for dielectric cylinder. The simulation was performed with a grid of 121×121, with a homogeneous cylinder of radius 1 and dielectric index $\varepsilon_c = 8$. The gain was chosen to be $\omega_a = 4.3$ and $\gamma_\perp = 1$. Both threshold modes had frequency $\omega_t = 4.267$. The purely sinusoidal lasing mode has a slightly higher lasing amplitude than the circulating solution. The theoretical lines were obtained from Eq. 4.12 and Eq. 4.28.

with eigenvalue given by Eq. 4.27 and profile $\sin(m\phi+\phi_0)$. The case in which $2(\theta_2-\theta_1)$ is an odd multiple of π gives $|a_\mu|^2 = \frac{\lambda}{4}$, giving the lasing modes $\mathbf{E} = \hat{\mathbf{z}}\sqrt{\frac{\lambda d}{4}}B_{Nm}(\rho)e^{\pm im\phi}$. Now check the stability of this lasing mode. From Eq. 4.16, we have $0 = (\delta\omega' - \delta\omega)a'_\mu$, so both eigenvalues here are $\delta\omega' = \delta\omega$, and this circulating mode is stable. Again, a much quicker way to obtain the stable linear combination here is to use Eq. 4.20 to see that the relative phase between a_1 and a_2 must be $\pm i$, and Eq. 4.21 directly gives the amplitude $|a_\mu|^2 = \frac{\lambda}{4}$.

Dielectric square

To see an example in which the interference term of the two-mode SALT equation has a very noticeable effect, we look at a pair of low-Q modes of a homogeneous dielectric

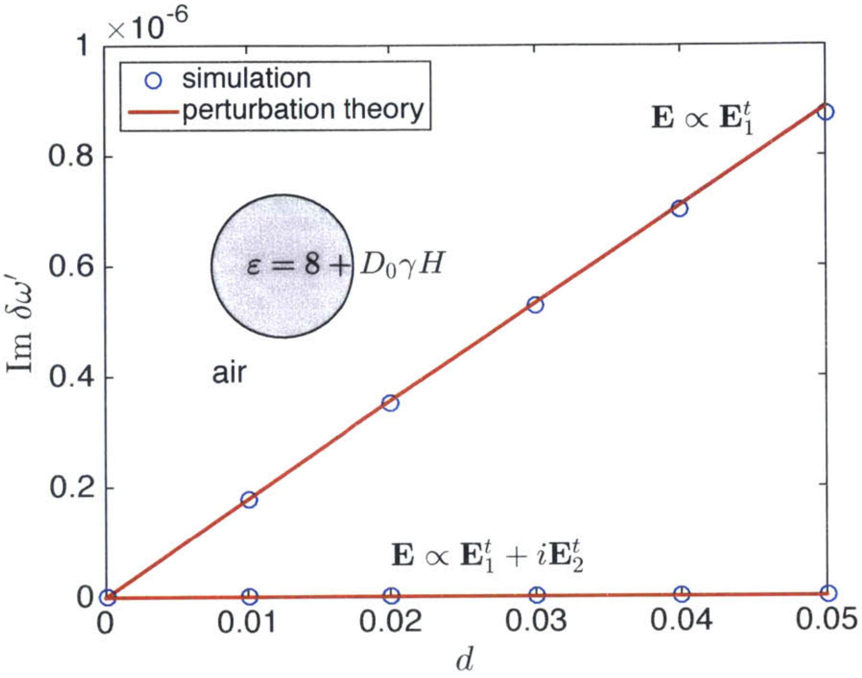


Figure 4-8: Stability eigenvalues for dielectric cylinder. The simulation data points we obtained using the same method as in Fig. 4-4. The circulating mode is clearly stable while the sinusoidal mode is not. The theoretical lines were obtained from Eq. 4.16.

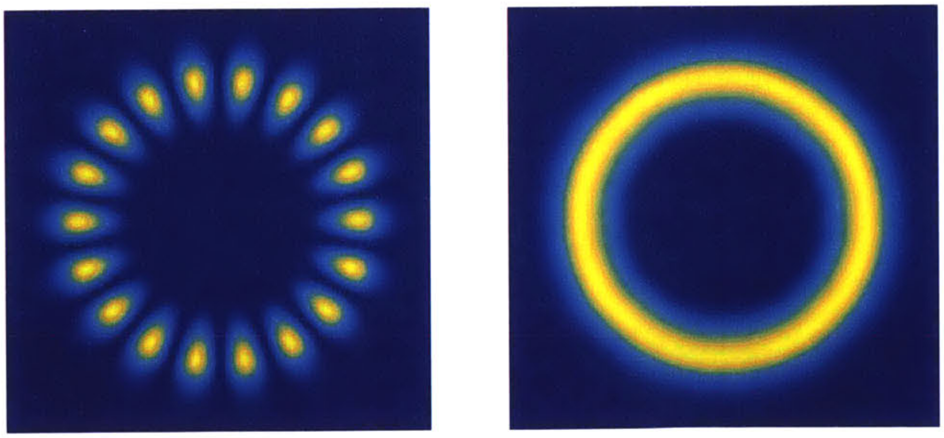


Figure 4-9: Intensity profiles for cylinder lasing modes. The intensity of the cosine mode (left) has nodes that make it unstable to the sine mode. The intensity for the circulating mode (right) does not have such nodes. It is also stable because it maintains the original symmetry that connected the two threshold modes ($C_{\infty\mathrm{v}}$ in the ideal system and $C_{4\mathrm{v}}$ for the discretized geometry).

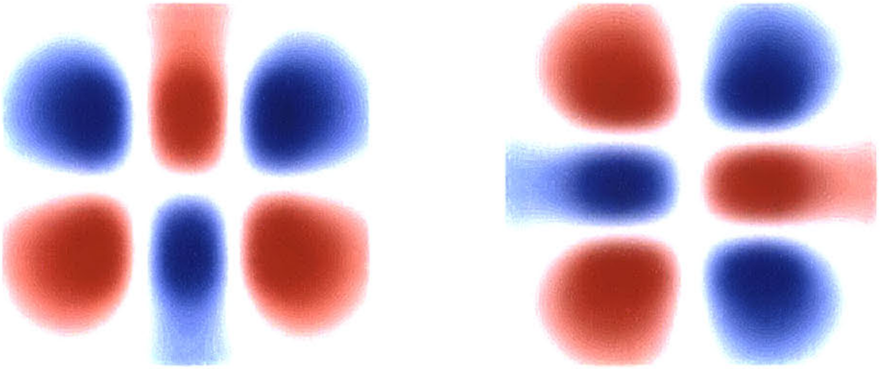


Figure 4-10: A pair of low-Q modes ($Q \approx 15$) at threshold for dielectric square and homogeneous dielectric $\varepsilon_c = 5$. The square has a side of 1, and the lasing parameters have been chosen to be $\omega_a = 3.5$, $\gamma_\perp = 1.0$. The $\frac{\pi}{2}$ rotation is an exact symmetry of the geometry, so there is an exact degeneracy even for the numerical grid.

square (Q is defined as the ratio $\mathrm{Re}\omega'/\mathrm{Im}\omega'$ of the cold cavity at zero pump strength). Unlike the metal square, the equation for the electric field is not separable in the x and y directions. The modes $\mathbf{E}_1$ and $\mathbf{E}_2$ are shown in Fig. 4-10.

Numerically performing the perturbation theory to obtain the stable mode, we verify that Eq. 4.20 gives a prediction for the stable linear combination of $\mathbf{E} \propto \mathbf{E}_1 + i\mathbf{E}_2$, which is the circulating mode predicted by Sec. 4.1.3. Although the field is neither real nor separable, the relative phase is still i because of the four-fold rotation symmetry. From the fact that $\mathbf{E}_2 = R_4\mathbf{E}_1$ (where R_4 is a four-fold rotation), we see that

$$\mathbf{E}_1 + i\mathbf{E}_2 \propto \left(1 + iR_4 - R_4^2 - iR_4^3\right)\mathbf{E}_1,$$

and this intensity pattern is chiral (with C_4 symmetry rather than $C_{4\mathrm{v}}$). This chirality is an important consequence of including the interference terms in the degenerate formulation of SALT. Fig. 4-11 shows the intensity pattern for the result without interference compared to the correct solution. Fig. 4-12 shows the intensity pattern of the degenerate partner. Since this degeneracy does not come from geometric symmetry along, there is no clear symmetry operation that takes the lasing mode (right panel of Fig. 4-11) to its degenerate partner (Fig. 4-12), but there is still an exact degeneracy

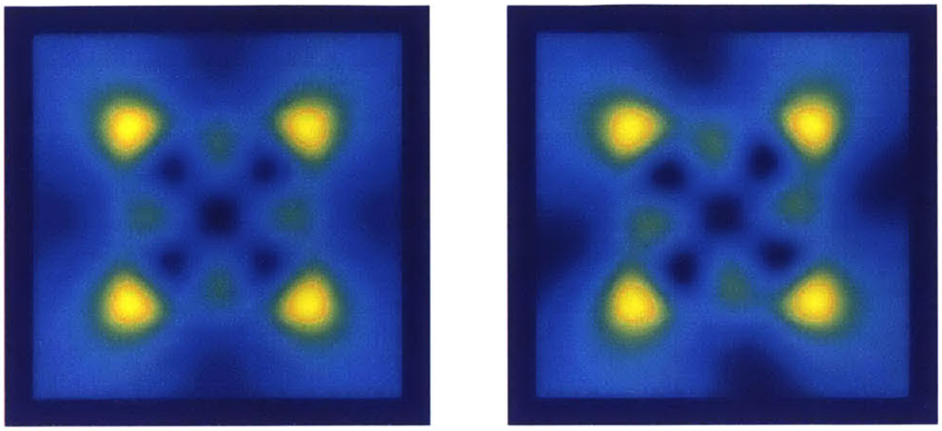


Figure 4-11: The incorrect (left) intensity pattern and correct (right) intensity pattern for the pair of low-Q dielectric square modes, slightly above threshold. The left intensity pattern was obtained by solving two-mode SALT without any interference effects, while the right pattern was obtained by constructing the stable linear combination predicted by symmetry and perturbation theory and then solving single-mode SALT. The correct pattern clearly has a chirality, which the incorrect pattern lacks.

as explained in appendix A.

4.2 Symmetries broken by discretization

In many cases the geometry we are trying to solve has some symmetry, which results in a pair of degenerate modes, but the degeneracy is broken when the geometry is projected onto the discretized grid, since the grid no longer has the symmetry. For linear equations, this unphysical splitting is not an issue because it is usually straightforward to tell whether a pair of modes is "really" degenerate, and since all linear superpositions solve the equation in the infinite-resolution limit, we can construct arbitrary superpositions *after* solving for both of the modes. However, for SALT (which is nonlinear), the coefficients of the superposition is a physical quantity that must be found by our solution method, as explained in Secs. 4.1.3 and 4.1.4. As explained in chapter 3, the process for solving for lasing modes begin with the linear problem for the passive poles. Because both the real and imaginary parts of the passive poles are split by the discretization error, the modes will lase at different pump strengths, and even after both modes lase we cannot construct a

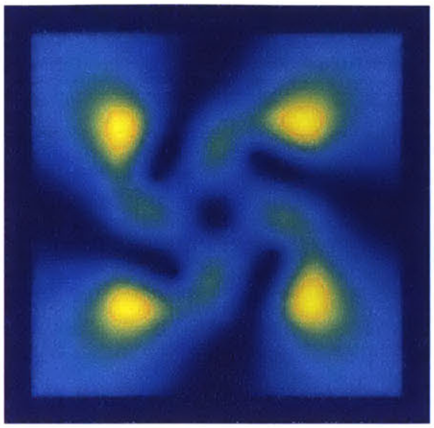


Figure 4-12: Degenerate passive mode of opposite chirality of low-Q dielectric square for lasing very high above threshold ($D_0 = 100D_0^t$). The profile is not simply a mirror flip of the lasing intensity pattern in Fig. 4-11, since there is no symmetry that connects the two modes.

linear combination of them because the two modes satisfy equations with different real eigenfrequencies. Consequently, degeneracy-splitting caused by discretization is a major practical problem for our SALT solver and requires a nontrivial solution.

A good example is the uniform cylinder projected onto a square grid, which has $C_{4\text{v}}$ symmetry but not the $C_{\infty\text{v}}$ symmetry of the cylinder (another example, as we will see later in Sec. 4.2.2 and in Fig. 4-22, is a hexagonal photonic crystal defect cavity that has $C_{6\text{v}}$ symmetry). For odd angular momentum whispering-gallery modes, the sine mode is simply a 90-degree rotation of the cosine mode, as seen in Fig. 4-2. This symmetry is still exact for the grid, so odd angular momentum modes are exactly degenerate even when discretized. However, for even angular momentum modes, the operation that takes sine to cosine is not a member of $C_{4\text{v}}$. Hence, there is a small splitting in both the real and imaginary parts of the frequencies. Fig. 4-13 gives an example of this effect. There are two potential methods to solve this problem: one is to solve a two-mode equation with an interference term in the spatial hole-burning denominator, while allowing the frequencies to split. This method is conceptually simpler, but is more difficult to use in practice because it is difficult to choose the correct relative phase. The second method is to construct a small perturbation to the dielectric function to force a degeneracy in both the pump strength and the frequencies

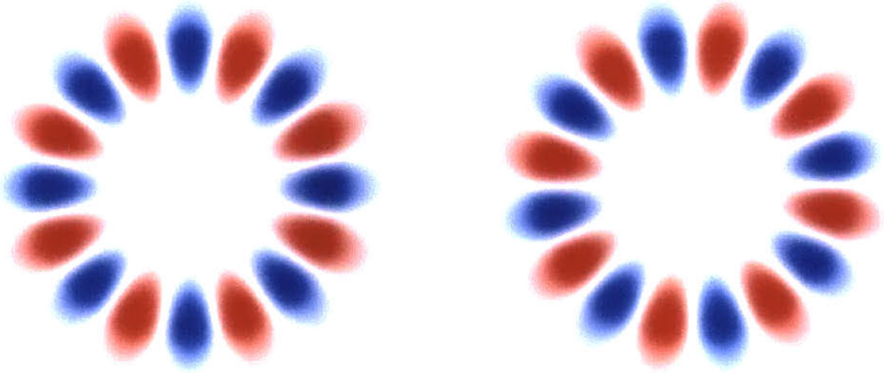


Figure 4-13: Even-m threshold modes for dielectric cylinder. The real part of $\mathbf{E}$ has been plotted. The geometry and parameters are the same as in Fig. 4-2, except with the gain shifted to $\omega_a = 3.8$, which is near the threshold frequency for the even-m modes. Unlike in the previous case, however, the discretized modes are *not* $\frac{\pi}{2}$ rotations from each other. Consequently, there is an unphysical splitting of 0.11% in $\text{Re}(\omega_1 - \omega_2)$ and 11.5% in $\text{Im}(\omega_1 - \omega_2)$ (the latter being larger only because these are high-Q modes and $\text{Im}\omega_\mu$ is already very small). A difference in imaginary parts also means a splitting in the threshold pump strength D_t.

at threshold. We discuss both methods below.

4.2.1 Solving a two-mode equation with splitting

Because the symmetry is broken by the grid, the two modes are only nearly degenerate and cannot be combined in the stable linear combination, or *any* linear combination for that matter, to solve *single*-mode SALT. However, the modes will satisfy a modified *two-mode* SALT equation

$$
\begin{aligned}
-\nabla \times \nabla \times \mathbf{E}_\mu &= \omega_\mu^2 \left[\varepsilon_c + D\gamma(\omega_\mu) F H\right] \mathbf{E}_\mu \qquad (4.29) \\
H &\equiv \frac{1}{1 + \left|\mathbf{E}_1 + e^{i\theta}\mathbf{E}_2\right|^2},
\end{aligned}
$$

with the two frequencies $\omega_1 \neq \omega_2$. The relative phase θ here is a physical quantity because we have fixed the phase of both $\mathbf{E}_1$ and $\mathbf{E}_2$, as explained in the previous chapter on the solution method for SALT. Note that here we are including the interference term $\mathbf{E}_1 \cdot \mathbf{E}_2^\star e^{-i\theta} + cc$ even though the frequencies are slightly different (the splitting

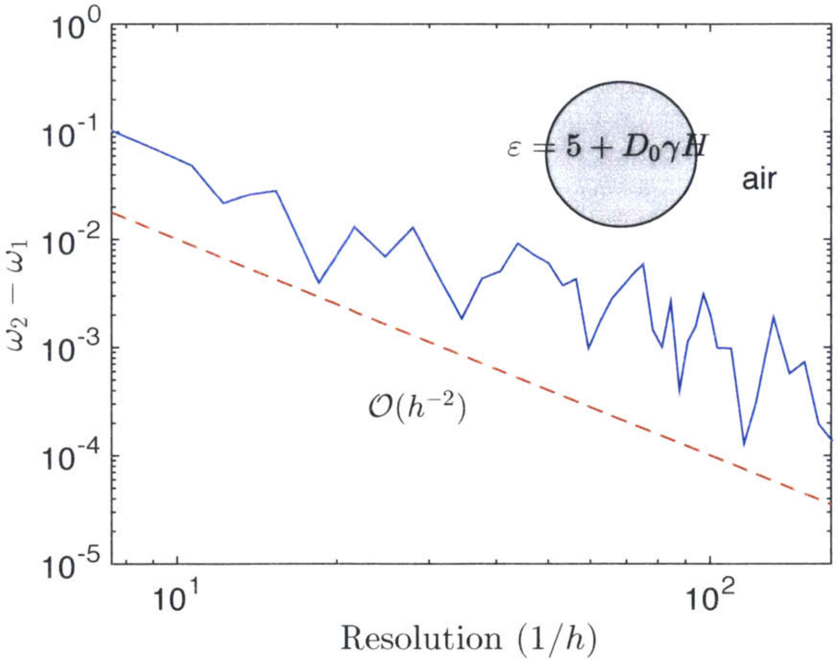


Figure 4-14: Splitting in degeneracy due to discretization error for even-m modes of dielectric cylinder versus the resolution $1/h$ of the discretization, where h is the distance between adjacent gridpoints. The oscillations, which are due to the discontinuous interfaces between dielectric and air that "jump" when the resolution is changed, could in principle be smoothed by using subpixel averaging techniques for the discretization [231].

will be on the order of the discretization error), because this is an approximation for the *true* (physical) symmetrical system when the modes are degenerate, and we are solving Eq. 4.3.

Eq. 4.29 is difficult to solve because of the relative phase θ. If we incorporate it as an extra unknown, the nonlinear problem becomes underspecified, making Newton's method much less straightforward to apply. Another option is to incorporate θ as a "known" parameter that we specify before solving. This method keeps the Jacobian square and the problem exactly specified, but greatly constricts the two-mode solution space in which we are searching for a root, since we are essentially specifying the exact value, to machine precision, of one of the components of $\mathbf{E}_2$ at the point $\mathbf{x}_2$ where we fixed its phase. Consequently, some ranges of values of θ often do not have a solution to Eq. 4.29, and even when there *is* a solution, the resulting intensity pattern $\left|\mathbf{E}_1 + e^{i\theta}\mathbf{E}_2\right|^2$ will vary significantly with θ and it is unclear which θ is the closest one to the stable lasing intensity pattern. Fig. 4-15 shows the intensity patterns for two choices of θ.

In principle, there may be some additional equation that determines the value of θ, but it is not clear what this equation may be. In order to discover that equation is needed to constrain theta, one possible approach is to simply fix θ at some arbitrary value and solve Eq. 4.29 in the remain unknowns by a Newton method analogous to chapter 3. Then, by comparing these solutions for different values of θ, we might hope to notice some obvious reason why one solution is "more correct" (or more stable) than the other solutions, and use this observation to determine theta (ideally as an additional equation introduced into the Newton solver). Even this approach is immediately problematic, because we find that Newton does not even converge for all values of theta, and it is not clear how to find which θ are solvable except by brute-force search. Even when there is a solution, we often find solutions that are "stable" (two on-axis poles of Eq. 4.29) for multiple values of θ, but whose intensity patterns vary greatly. (If the numerical solver *does* have an exact degeneracy, then the choice of θ is irrelevant: because $\mathbf{E}_1$ and $\mathbf{E}_2$ solve the same eigenproblem with the same eigenvalue, the solver can just replace $\mathbf{E}_1$ and $\mathbf{E}_2$ by arbitrary superpositions of

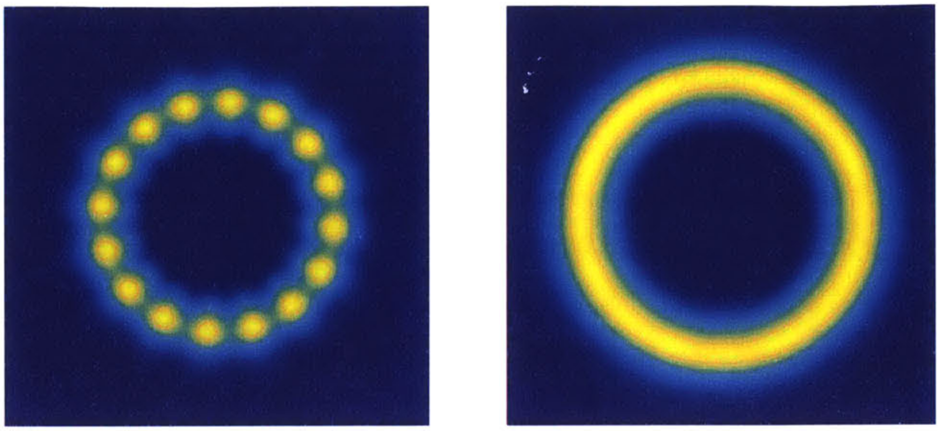


Figure 4-15: Intensity profiles for solutions Eq. 4.29 with $\theta = 1.3$ (left) and $\theta = 1.57$ (right). The intensity pattern on the left is unphysical because we have chosen the wrong θ as a parameter. On the right, the intensity pattern looks almost rotationally invariant, and that is because we chose a value of θ close to the known correct value of $\theta = \frac{\pi}{2}$.

one another as needed.)

For example, Fig. 4-15 shows two different intensity patterns for the dielectric cylinder lasing solutions (for even m where the grid breaks the degeneracy) obtained for different values of θ in Eq. 4.29. We know from previous work [32] that the physical solution (corresponding to an exact degeneracy for the circular symmetry) is the solution shown at the right (the "circulating" mode with a rotation-invariant intensity pattern). The mode shown at the left is closer to a standing-wave solution and is not a solution of the single-mode SALT equations in the exact symmetry group. Because both of these solutions are (numerically) non-degenerate, however, and it is difficult to determine which one of the two converges to an exact stable solution of the single-mode SALT equations in the limit of infinite resolution (where symmetry is restored). The basic problem is that Eq. 4-15 is not a physical SALT equation except in the limit of an exact degeneracy, and so we can obtain unphysical solutions for an incorrect θ.

For many cases we encountered, there is a simple heuristic that often leads to a value of θ that is very close to the known stable value. For the exactly degenerate case ($\omega_1 = \omega_2$), the intensity pattern $\left|\mathbf{E}_1 + e^{i\theta}\mathbf{E}_2\right|^2$ will be completely independent of θ; the solver can freely mix between $\mathbf{E}_1$ and $\mathbf{E}_2$ because the two equations have the same frequency anyway. For the non-degenerate case, this will no longer be true. However, if we were to find a θ in which H is *stationary*, i.e. $\frac{dH}{d\theta} = 0$, then both the frequencies ω_μ will also be stationary as well. To see this, we expand 4.29 for some phase $\theta + \delta\theta$. The new modes with this small phase change will be $\mathbf{E}_\mu + \delta\mathbf{E}_\mu$, with frequencies $\omega_\mu + \delta\omega_\mu$. Inserting into Eq. 4.29, multiplying both sides by $\mathbf{E}_\mu$ and integrating, we have

$$\delta\omega_\mu = -\frac{\omega_\mu \int (\mathbf{E}_\mu \cdot \mathbf{E}_\mu)\, D\gamma(\omega_\mu) F \delta H}{2\int (\mathbf{E}_\mu \cdot \mathbf{E}_\mu)\left(\varepsilon_c + D\left[\gamma(\omega_\mu) + \frac{\omega_\mu}{2}\gamma'(\omega_\mu)\right] FH\right)}.$$

Hence, if $\delta H(\mathbf{x}) = 0$, then also $\delta\omega_\mu = 0$. Using this as a heuristic, we can often obtain the correct phase between the pair of nearly degenerate lasing modes. For example, Fig. 4-16 shows this method applied to the even-m cylinder modes from Fig. 4-13. However, this method is difficult to implement, because we need to solve Eq. 4.29 at many values of θ (many for which the solution may not even exist, as explained above) and identify the extremal point. Second, this extremal point may not even exist in some cases. Third, we have no proof that this always works: it is based purely on intuition and experimentation. For these reasons, we look for a more reliable solution to the problem of unphysical degeneracy splittings.

4.2.2 Forcing the degeneracy using a dielectric perturbation

The basic idea is that we construct an artificial perturbation $\delta\varepsilon(\mathbf{x})$ to the dielectric that forces the degeneracy in both the frequency and threshold, and then we solve the perturbed single-mode SALT equation. There are infinitely many possible functions that do this, so we look for the one with the smallest L_2 norm $\|\delta\varepsilon(\mathbf{x})\|_2^2 = \int |\delta\varepsilon(\mathbf{x})|^2$. This is also reasonable because in the limit of infinite resolution, the perturbation required to force the degeneracy must approach zero. We construct the perturbation by solving a *quadratic program* with linear constraints which we obtain using perturbation theory.

Not only does this uniquely (and cheaply) determine $\delta\varepsilon$, as described below, but it

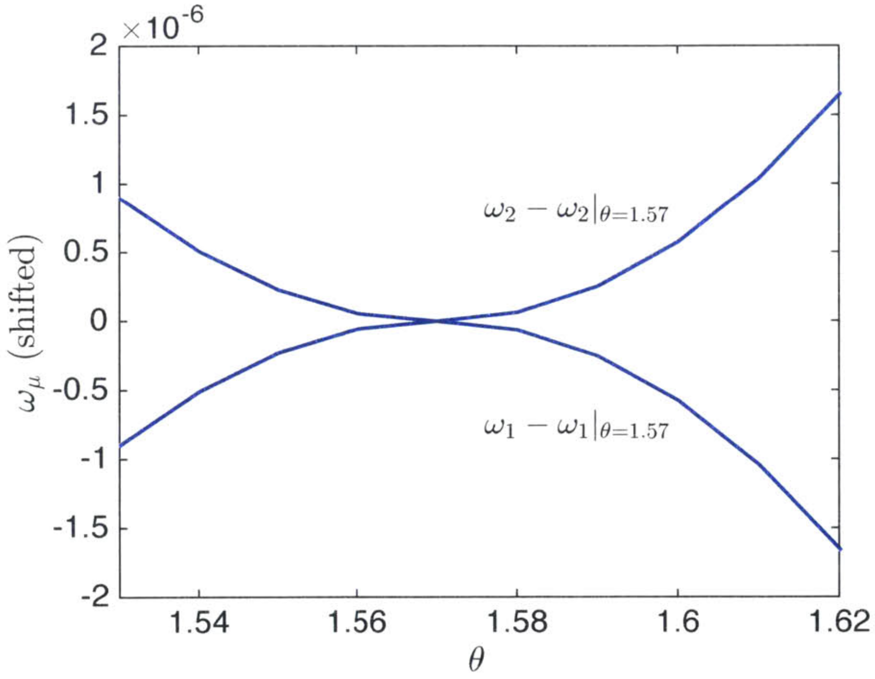


Figure 4-16: Lasing frequencies vs. relative phase for even-m cylinder modes above threshold obtained from numerical solution of Eq. 4.29. The actual splitting $\omega_2 - \omega_1 \approx 0.004$ is much greater than the variation shown here. For clarity, the frequencies plotted have been shifted and centered at their values at $\theta = 1.57$, which is near the stationary point (as expected, since we know the correct phase to be $\theta = \frac{\pi}{2}$).

also guarantees convergence to the solution of the unperturbed (physical) single-mode SALT equation in the limit of infinite resolution. The reason it guarantees convergence is that the frequency splitting vanishes in the limit of infinite resolution, and so the minimum-norm $\delta\varepsilon$ to force a degeneracy also vanishes in the limit of infinite resolution, recovering the unperturbed SALT.

It turns out that determining the minimum-norm $\delta\varepsilon$ requires only that we solve a sequence of "quadratic programming" (QP) [91] problems: minimizing a convex quadratic function ($\|\delta\varepsilon\|_2^2$) of $\delta\varepsilon$ subject to an affine constraint on $\delta\varepsilon$. QPs are "convex" optimization problems with a unique global minimum that can be efficiently found simply by solving a system of linear equations [91]. In particular, the affine constraint (the equation on $\delta\varepsilon$ that makes it coalesce the eigenvalues) can be derived from perturbation theory as explained below. Because the perturbation theory is only first-order, however, the $\delta\varepsilon$ that we find by solving the QP only approximately eliminate the degeneracy, but we can simply re-solve SALT and solve a new QP, iterating the process a few times (twice is typically enough) to force a degeneracy to machine precision.

We analyze the effect of a small $\delta\varepsilon$ on the eigenfrequencies (of the lasing mode and the passive pole) by well-known first-order perturbation theory for Maxwell's equations [34], but some modification is required to handle the nonlinearity of the hole-burning term above threshold. However, we begin by forcing the degeneracy below threshold (repeating as needed as the pump strength is increased), so that both passive poles reach threshold simultaneously. (In practice, we achieved the fastest convergence by allowing passive poles to have positive imaginary parts, and then setting the pump strength so that the two poles "straddle" the real axis; this way, when they meet in the middle they are both exactly at threshold.) Below threshold, the eigenproblem is linear, and we can apply standard perturbation theory (albeit for a complex-symmetric operator, not a Hermitian operator) as follows:

Consider two nonlasing modes that satisfy

$$0 = -\nabla \times \nabla \times \mathbf{E}_\mu + \omega_\mu^2 \varepsilon_\mu \mathbf{E}_\mu$$
$$\varepsilon_\mu \equiv \varepsilon_c + D_0 \gamma(\omega_\mu) F$$

Adding a perturbation to the dielectric $\delta\varepsilon$ will result in corresponding responses $\delta\mathbf{E}_\mu$ and $\delta\omega_\mu$. As in the threshold perturbation theory, we multiply both sides by $\mathbf{E}_\mu$ and keep only first-order terms. Terms involving $\delta\mathbf{E}_\mu$ again vanish because the operators act to the left, and we are left with [232–234]

$$\delta\omega_\mu = -\frac{\int \mathbf{E}_\mu \cdot \delta\varepsilon \mathbf{E}_\mu}{\int \mathbf{E}_\mu \cdot \left(\frac{2\varepsilon_\mu}{\omega_\mu} + \frac{\partial \varepsilon_\mu}{\partial \omega_\mu}\right) \mathbf{E}_\mu}. \tag{4.30}$$

We write this frequency shift as an inner product

$$\delta\omega_\mu = -p_\mu^T \delta\varepsilon. \tag{4.31}$$

As an aside, while it is fine to use a scalar $\delta\varepsilon$ function for this procedure, in the case when the $\mathbf{E}_\mu$ are TE modes or fully-vectorial fields, then it is also possible to allow $\delta\varepsilon(\mathbf{x})$ to be a diagonally anisotropic tensor

$$\overleftrightarrow{\delta\varepsilon}(\mathbf{x}) = \begin{pmatrix} \delta\varepsilon_{xx}(\mathbf{x}) & 0 & 0 \\ 0 & \delta\varepsilon_{yy}(\mathbf{x}) & 0 \\ 0 & 0 & \delta\varepsilon_{zz}(\mathbf{x}) \end{pmatrix}. \tag{4.32}$$

The column-vector form of $\delta\varepsilon$ in Eq. 4.31 would then have as its elements all the real and imaginary components of $\overleftrightarrow{\delta\varepsilon}(\mathbf{x})$ at each Yee point [37] [$\delta\varepsilon_{xx}(\mathbf{x})$, $\delta\varepsilon_{yy}(\mathbf{x})$, and $\delta\varepsilon_{zz}(\mathbf{x})$ for all the grid points $\mathbf{x}$], while the row-vector p_μ^T would have as its elements the real and imaginary parts of $E_x(\mathbf{x})^2$, $E_y(\mathbf{x})^2$, and $E_z(\mathbf{x})^2$ at all the grid points. If we take this option, then the norm we minimize would be

$$\|\delta\varepsilon\|_2^2 = \int d^3\mathbf{x} \left\| \overleftrightarrow{\delta\varepsilon}(\mathbf{x}) \right\|_F^2,$$

where the Frobenius norm [112] at each point $\mathbf{x}$ is defined as

$$\left\| \overleftrightarrow{\delta\varepsilon}(\mathbf{x}) \right\|_F^2 \equiv |\delta\varepsilon_{xx}(\mathbf{x})|^2 + |\delta\varepsilon_{yy}(\mathbf{x})|^2 + |\delta\varepsilon_{zz}(\mathbf{x})|^2 .$$

Whether we take $\delta\varepsilon$ to be scalar or tensor, the degeneracy-forcing condition $\omega_1 + \delta\omega_1 = \omega_2 + \delta\omega_2$ then becomes

$$(p_2 - p_1)^T \delta\varepsilon = \omega_2 - \omega_1.$$

It turns out that the solution of a quadratic program with equality constraints can be obtained directly by solving a linear *dual* problem [91], which in this case is

$$\begin{pmatrix} 1 & 0 & \mathrm{Re}q & \mathrm{Im}q \\ 0 & 1 & -\mathrm{Im}q & \mathrm{Re}q \\ \mathrm{Re}q^T & -\mathrm{Im}q^T & 0 & 0 \\ \mathrm{Im}q^T & \mathrm{Re}q^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{Re}\delta\varepsilon \\ \mathrm{Im}\delta\varepsilon \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \mathrm{Re}\,(\omega_2 - \omega_1) \\ \mathrm{Im}\,(\omega_2 - \omega_1) \end{pmatrix}.$$

Here, we have defined $q \equiv p_2 - p_1$, and the $\lambda_{1,2}$ are Lagrange multipliers that are not needed. When ω_2 is very close to ω_1, we can improve the condition number of the matrix by freely multiplying the second-to-last row and column of the matrix by a constant factor, provided that the second-to-last element of the right-hand side is *divided* by the same factor. The same can be done for the last row and column, with the last element of the right-hand side. The resulting $\delta\varepsilon$ of this procedure applied to the odd-m threshold modes in Fig. 4-13 are shown in Fig. 4-17, and the convergence of the splitting to zero is shown in Fig. 4-18. As verified in Fig. 4-19, the L_2 norm decreases with higher resolution, satisfying our requirement that the dielectric perturbation should go to zero in the continuum limit.

Note that even after the thresholds and threshold frequencies have been made exactly degenerate using the QP procedure illustrated above, we are still in principle forcing the degeneracy. Above threshold, the delicate balance created by $\delta\varepsilon$ to force the frequencies together is slightly broken. This results in an *approximate* degeneracy that is maintained very far above threshold, as shown in Fig. 4-20, with only a 10^{-8}

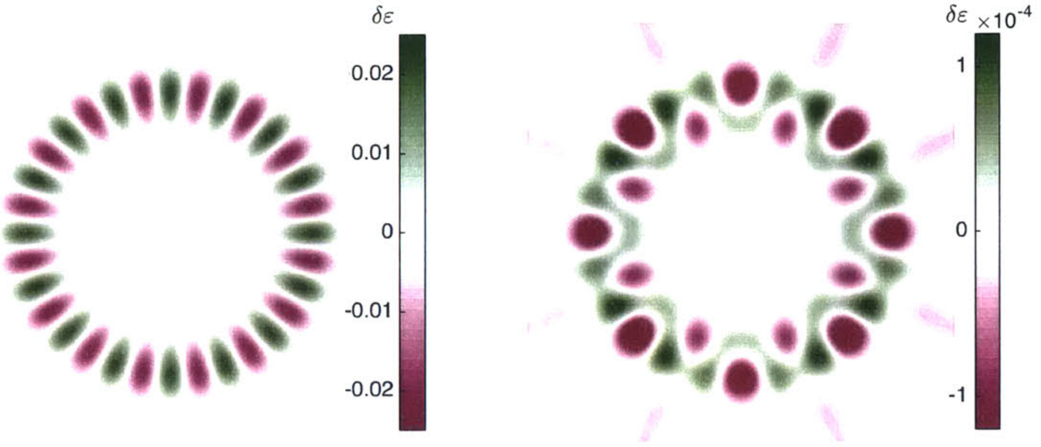


Figure 4-17: Dielectric perturbation $\delta\varepsilon$ obtained by solving QP for even-m threshold modes. The real part (left) has a dependence $\cos(2m\phi)$, while the imaginary part (right) is a more complicated function. Only two iterations of QP were required to obtain a 10^{-15} degeneracy in both the threshold frequency ω_t and the threshold pump strength D_t.

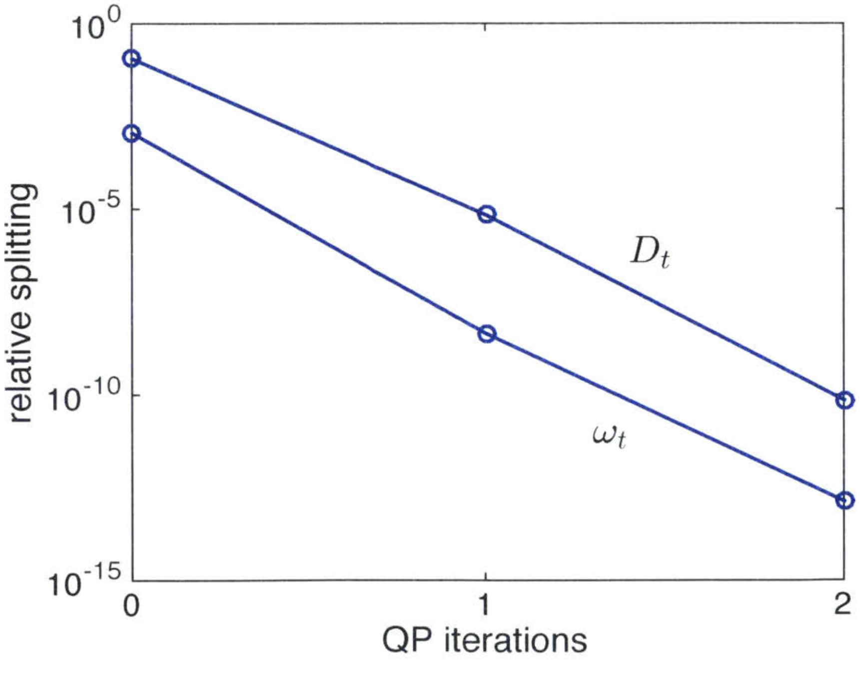


Figure 4-18: Relative splitting in threshold pump strength and frequency for even-m cylinder modes after QP iterations. The relative splitting in frequency is defined in the usual way as $2\left|\frac{\omega_1-\omega_2}{\omega_1+\omega_2}\right|$, and similarly for the pump strength. Two solves for $\delta\varepsilon$ is all that is needed to force the degeneracy.

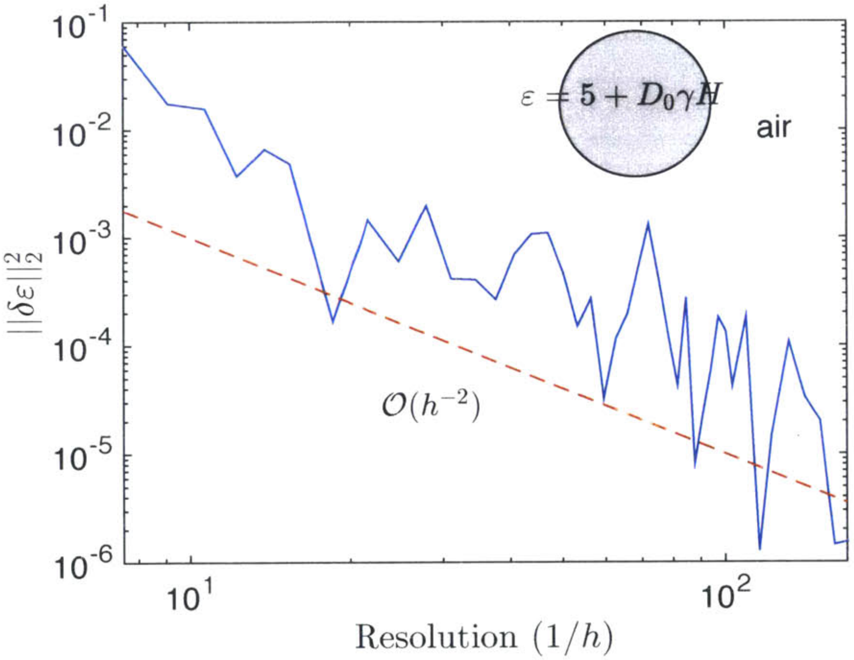


Figure 4-19: L_2 norm of resulting $\delta\varepsilon(\mathbf{x})$ function obtained from QP procedure versus discretization resolution $1/h$ for nearly degenerate even-m modes of cylinder, where h is the spacing between adjacent gridpoints. The same resolutions as in Fig. 4-14 were used, and the oscillations resemble the curve for splitting very closely. This is because the larger the splitting $\omega_2 - \omega_1$, the larger the $\delta\varepsilon(\mathbf{x})$ function needed to enforce the degeneracy. The fact that $||\delta\varepsilon||_2^2$ appears to be going to zero as the resolution increases indicates that our QP procedure is convergent.

splitting for pump strengths up to 100 times threshold. In practice, these results are already accurate enough to give all the desired physical information about the degenerate pair. If we wanted to be absolutely correct and force the degeneracy to machine precision (as it was in the exactly symmetric case for odd-m modes), we could simply perform QP again at some given $d > 0$ to force $\delta\omega$ and $\delta\omega'$ back together. One extra caveat in this case is that $\delta\omega$ is now a lasing pole, so the spatial hole-burning term need to be taken account for the perturbation theory ($\delta\omega'$ is still a passive pole, so the previous perturbation theory still applies), and instead of Eq. 4.30 we now have

$$\begin{aligned}
\delta\omega &= -\frac{\int \mathbf{E} \cdot (\delta\varepsilon + D_0\gamma(\omega_\mu)F\delta H)\,\mathbf{E}}{\int \mathbf{E} \cdot \left(\frac{2\varepsilon}{\omega} + \frac{\partial\varepsilon}{\partial\omega}\right)\mathbf{E}} \\
\varepsilon &\equiv \varepsilon_c + D_0\gamma(\omega)FH \\
\delta H &\equiv \frac{1}{1 + |\mathbf{E} + \delta\mathbf{E}|^2} - \frac{1}{1 + |\mathbf{E}|^2}.
\end{aligned}$$

Here, δH is the change in the spatial-hole burning term arising from the dielectric perturbation $\delta\varepsilon$. However, since there is no easy way to determine $\delta\mathbf{E}$ without numerically solving the full problem, δH is hard to determine semi-analytically. A simple work-around is to set $\delta H = 0$ above, which makes this procedure no longer a true first-order perturbation theory. However, since the splitting is already so small as shown in Fig. 4-20, the $\delta\varepsilon$ needed is also extremely small, so δH is also negligible. Although $\delta\omega$ is not zero to first order, the $\delta H = 0$ approximation is enough to find a $\delta\varepsilon$ that greatly decreases $\delta\omega$. We find empirically that it usually takes one iteration of this above-threshold QP procedure to restore the degeneracy of the lasing pole ω and its passive mode ω' to machine precision, since $\delta\omega$ is already very small. Practically speaking, this entire extra step is rarely needed since the solutions obtained from $\delta\varepsilon$ for the linear problem below threshold are already close enough for most pump strengths of physical interest.

We show another example, of a hexagonal cavity as shown in Fig. 4-22. This geometry was adapted from an infinite lattice of period a with air holes of radius $0.3a$. A single hole in the middle has a reduced radius $0.2a$ to create a defect in the band

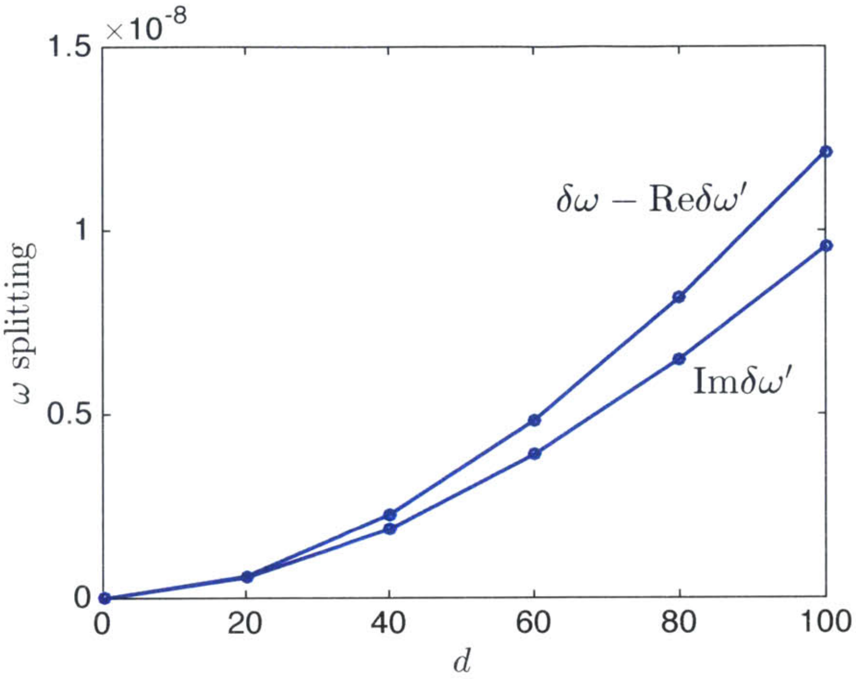


Figure 4-20: Above-threshold splitting in real and imaginary parts of $\delta\omega'$ after performing QP procedure for even-m modes. The magnitude is very small because the intensity profile, as shown in Fig. 4-21, is very close to rotationally symmetric.

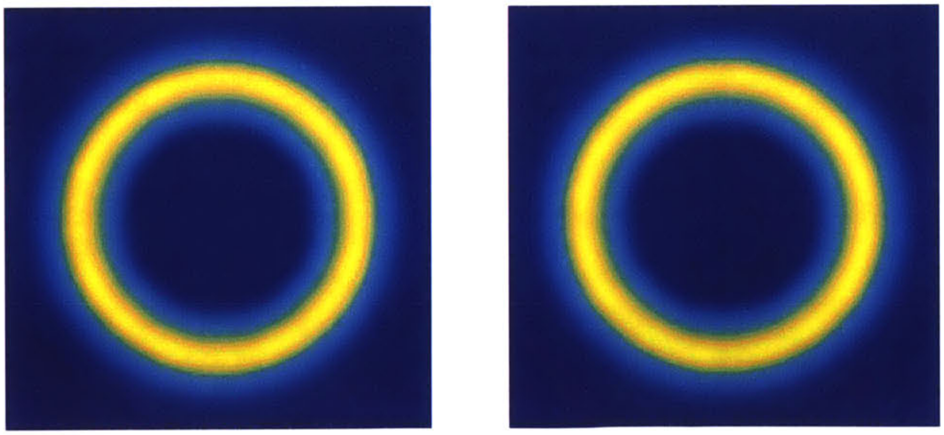


Figure 4-21: Intensity profiles for even-m lasing mode and its second passive pole at $d = 100$. The intensity of the lasing mode (left) differs imperceptibly from that of that of the passive eigenfunction. This difference is unphysical because it is solely due to discretization error, and is usually so small that it can be neglected.

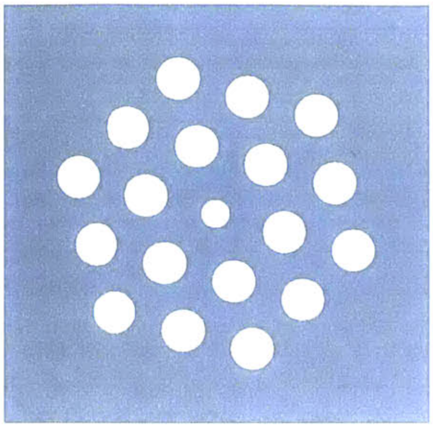


Figure 4-22: Dielectric function for hexagonal cavity. This cavity has $C_{6\mathrm{v}}$ symmetry and supports a pair of degenerate TE ($\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}}$) modes that transform as x and y under symmetry operations. All but two rows of holes have been removed to create a lower-Q structure. A PML is added to the boundaries to capture the radiation loss. The axes of the hexagon have been aligned with the diagonals rather than the x and y axes, because the finite-difference discretization happens to only have mirror symmetry along the diagonals.

gap. The dielectric is $\varepsilon_c = 11.56$ everywhere except in the holes, where there is air. Here, the lasing modes are TE (electric field in-plane and magnetic field out of plane), but there is a pair of degenerate threshold modes from the hexagon's $C_{6\mathrm{v}}$ symmetry, as shown in Fig. 4-23.

The two are related to each other by linear combinations of sixfold and threefold rotations. However, these operations do not belong in the $C_{4\mathrm{v}}$ symmetry group of the discretized grid, so the degeneracy here is again only approximate. For a 100×100 finite-difference discretization, there is about a 1.5% splitting between the threshold eigenvalues, so again we must use the QP procedure to force the threshold degeneracy. Since these are TE modes, we now have *two* components of the electric field, and consequently we may treat $\delta\varepsilon$ as a tensor, as in Eq. 4.32. We only consider the diagonal components $\delta\varepsilon_{xx}$ and $\delta\varepsilon_{xy}$ here for simplicity. Again, only two iterations of QP are necessary to force the degeneracy down to machine precision. We use Eq. 4.21 to predict the stable linear combination above threshold to use as an initial guess. The intensity pattern obtained in simulation is shown in Fig. 4-25.

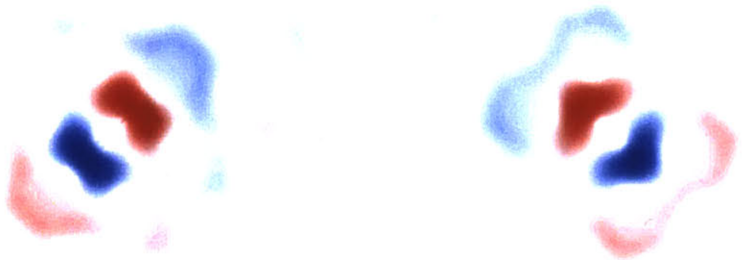


Figure 4-23: Magnetic fields H_z for pair of degenerate TE threshold modes for hexagonal cavity. The modes are not simple 90-degree rotations of one another, because that rotation is not a member of the $C_{6\mathrm{v}}$ symmetry group. However, these modes *can* be constructed by taking linear combinations of threefold and sixfold rotations of each other. Not shown are the imaginary parts of H_z which are non-negligible because this is not a high-Q cavity.

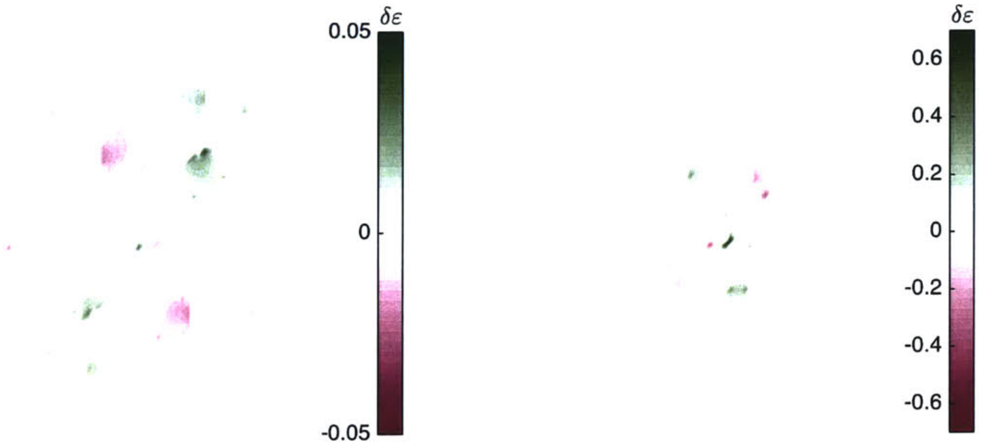


Figure 4-24: Dielectric perturbation obtained from QP procedure for hexagonal cavity. Since the mode is TE ($\mathbf{E} = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}}$), we have allowed the perturbation to be a diagonally-anisotropic tenosr, as in Eq. 4.32. Shown here are the real (left) and imaginary (right) parts of $\delta\varepsilon_{xx}$. The $\delta\varepsilon_{yy}$ looks similar except rotated by 60 degrees.

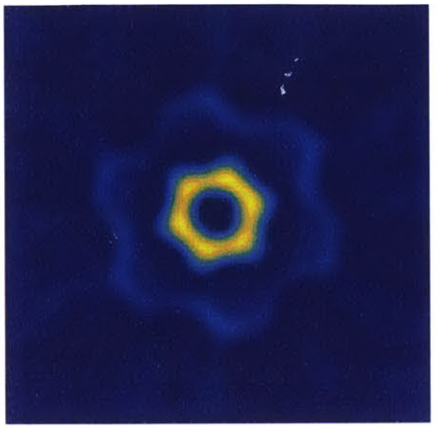


Figure 4-25: Intensity pattern for stable lasing mode for hexagonal cavity. The pattern appears to be six-fold symmetric, which is expected. Unlike in the right panel of Fig. 4-11 however, the chirality is not significant enough to be visible because the $Q \approx 100$ is much higher. In the ideal system, the second pole $\delta\omega'$ stays degenerate with the lasing eigenvalue $\delta\omega$, and this linear combination stays stable for all pump strengths above threshold. In the discretized system, there is not a true $C_{6\text{v}}$ symmetry, so there is a small splitting similar to that of the even-m cylinder modes. Again, this splitting is too small to affect physically meaningful results of the simulation, but can be removed using the QP procedure if desired.

4.3 Forced degeneracies

There are two types of degeneracies: those that come from symmetry, and forced degeneracies. For the passive cavity, if both the dielectric $\varepsilon_c(\mathbf{x})$ and the gain profile $F(\mathbf{x})$ are invariant under operations in some symmetry group, then pairs of degenerate modes will stay degenerate for all pump strengths under threshold. When a pair starts lasing, the gain profile will become

$$\frac{F(\mathbf{x})}{1 + |\mathbf{E}_1(\mathbf{x}) + \mathbf{E}_2(\mathbf{x})|^2} \tag{4.33}$$

If the total intensity $|\mathbf{E}_1 + \mathbf{E}_2|^2$ maintains the original symmetry or a sufficiently large subgroup (e.g. C_n as in Sec. 4.1.3), then the degeneracy will also be maintained. A simple example is the dielectric cylinder, which has $C_{\infty\text{v}}$ symmetry. The stable linear combination has an azimuthal dependence $e^{im\phi}$, so its intensity pattern is also azimuthally symmetric. This can be shown using the threshold perturbation theory,

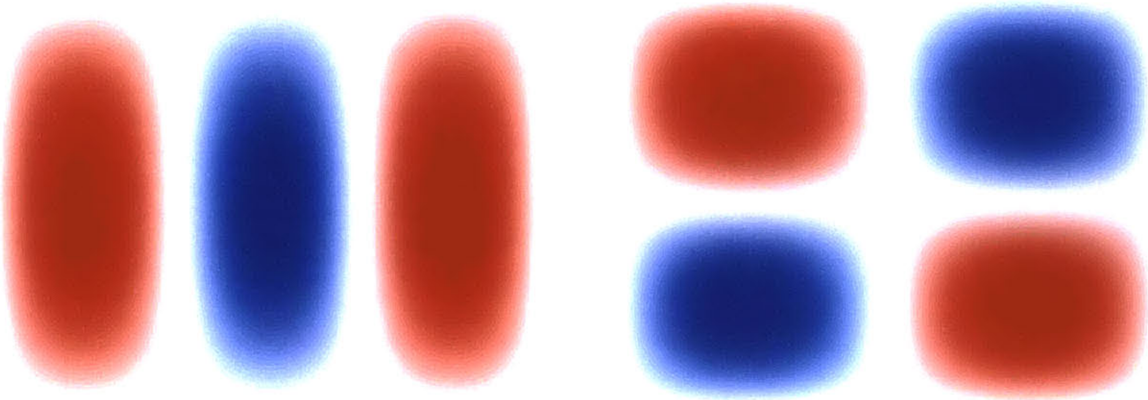


Figure 4-26: Threshold modes for metallic rectangle with forced degeneracy. A 90 × 70 cell with Dirichlet boundary conditions was used to simulate the cavity. A uniform loss of $\sigma = 0.01$ was chosen while the gain was chosen to be $\omega_a = \omega_t$ and $\gamma_\perp = 1$. While the rectangle's dimensions are chosen so that the modes ψ_{13} and ψ_{22} have the same frequency at threshold, there is no symmetry operation that takes one mode to the other; this is forced degeneracy.

except with $F(\mathbf{x})$ replaced by Eq. 4.33. Hence, whispering gallery modes will always stay degenerate above threshold, as long as there are no imperfections in the geometry. See also Sec. 4.1.3 for more general circulating modes in $C_{n\text{v}}$ symmetry structures.

A forced degeneracy, on the other hand, will in general split with lasing. This is because whatever delicate tuning in $\varepsilon(\mathbf{x})$ and $F(\mathbf{x})$ necessary to cause the degeneracy will most likely be broken by the spatial hole-burning. For example, we again consider the modes of the metallic rectangle in Eq. 4.22. This time we choose $L_x = 1$ and $L_y = \sqrt{7}$, such that $\omega_{13} = \omega_{22}$. The threshold modes ψ_{13} and ψ_{22} are shown in Fig. 4-26. Threshold perturbation theory predicts an intensity pattern proportional to $|\psi_{13} + i\psi_{22}|^2$. However, this spatial hole-burning breaks the delicate tuning required to force the degeneracy. Hence, the two modes will immediately split above threshold (as seen in Fig. 4-27) and SALT will no longer be valid because there will be slowly beating terms in the inversion. This splitting also has a second consequence: the $\psi_{13} + i\psi_{22}$ mode does not have a second linear eigenvalue $\delta\omega'$ that is exactly zero, as it did in the case of the square in Fig. 4-4 and the cylinder in Fig. 4-8. Fig. 4-27 shows this second value is small, but it has a positive imaginary part which grows. The intensity pattern for this passive pole and the lasing mode are shown in Fig. 4-28.

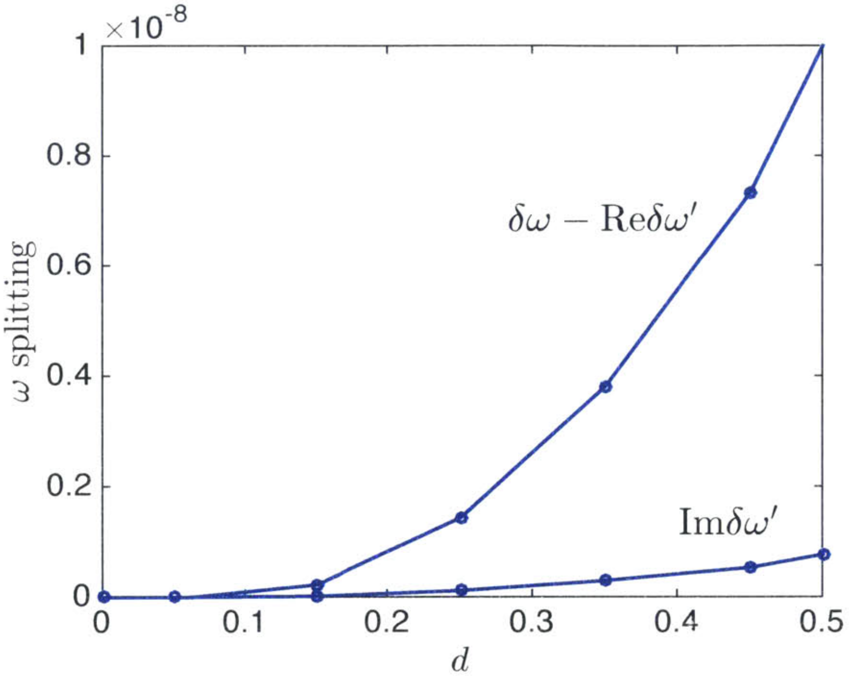


Figure 4-27: Above-threshold splitting in real and imaginary parts of $\delta\omega'$ for forced degeneracy in metallic rectangle. While perturbation theory predicts that this mode will be stable to first order in d slightly above threshold, the broken symmetry in the intensity profile gives it a very small but nonlinearly growing instability as the pump strength is further increased (there is not a simple power-law dependence of the instability on d, however). The real part of $\delta\omega'$ also splits away from $\delta\omega$ above threshold due to the forced degeneracy being broken. The perturbation theory prediction is consistent with the fact that the slopes of both curves are close to zero near $d = 0$.

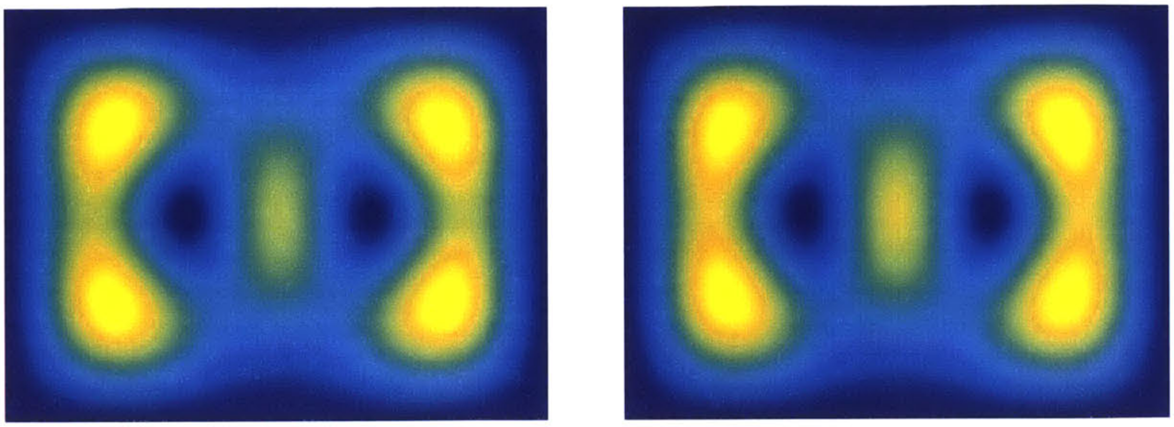


Figure 4-28: Intensity profiles for lasing mode and its second passive pole at $d = 100$. The intensity pattern of the lasing mode $\psi_{13} + i\psi_{22}$ (left) differs slightly from that of that of the passive eigenfunction. The latter has an eigenvalue $\delta\omega'$ that grows slowly with pump strength, as seen in Fig. 4-27.

4.4 Dynamics of slightly broken degeneracies

To illustrate what physically happens in the case that two lasing modes have slightly different frequencies, we perform brute-force finite-difference time-domain (FDTD) simulations of a 1d version of the Maxwell–Bloch equations [33] (similar to those performed in Ref. [235]) for scalar fields $E(x,t)$, $P(x,t)$ and $D(x,t)$ on a 1d ring $0 < x < 1$ with periodic boundary conditions:

$$\begin{aligned} \ddot{P}^+ &= \frac{d^2E^+}{dx^2} - \varepsilon_c \ddot{E}^+ \\ \dot{P}^+ &= -(\gamma_\perp + i\omega_a)P^+ - iE^+D \\ \dot{D} &= \gamma_\parallel(D_0 - D) + 2i\left[E^+P^{+\star} - cc\right], \end{aligned} \tag{4.34}$$

where ε_c is the "cold cavity" permittivity and the plus symbols indicate the "positive frequency" components, i.e. $E = E^+ + cc$ where $E^+(x,t) \equiv E_0^+(x,t)e^{-i\omega_a t}$ with E_0^+ varying much less rapidly than $e^{-i\omega_a t}$ (and similar for P). The dielectric is given by $\varepsilon = 1 + i\sigma$, where $\sigma > 0$ provides a loss mechanism. The corresponding SALT equation [28] is given by

$$-\frac{d^2\psi}{dx^2} = \omega^2 \left(\varepsilon_c + \frac{D_0 \gamma(\omega) F}{1 + |c\psi|^2} \right) \psi,$$

and the steady-state solution to Eq. 4.34 is

$$E^+(x,t) = \frac{c}{|\gamma(\omega)|} \sqrt{\frac{\gamma_\perp \gamma_\parallel}{2}} \psi(x) e^{-i\omega t}.$$

For the FDTD simulation, we numerically integrated Eqs. 4.34 on a grid of 20 points (to keep track of only the real, physical fields E and P, we simply added the complex conjugates of Eq. 4.34 and numerically integrated the equations for $\dot{E}$ and $\dot{P}$). This system is numerically tractable and captures many of the essential characteristics of degenerate lasing modes in SALT. Since the number of grid points was small, a simple second-order Euler method was enough to integrate with satisfactory accuracy. In Fig. 4-29, we show the dynamics of the field after it has reached a steady state. Initially, random noise was used for the fields E, P and D, and the simulation was run for a long time so that all transient behavior died out and only the steady-state lasing mode (if any) remained. As expected, in the degenerate case we obtain a steady-state circulating mode (a travelling sine wave in the $+x$ direction). In Fig. 4-30, we show the dynamics, for a pump strength chosen to be slightly above the threshold given by SALT, of the envelope (peaks in E of all the fast oscillations at the optical frequency) of the field at an arbitrarily chosen point x_0 when a small perturbation is added to the dielectric, so that the latter is $\varepsilon_c(x) = 1 + i\sigma + 0.01\cos(4\pi x)$. The field exhibits beating, in which there is a slow oscillation at approximately $0.0047\omega_a$ between the left and right-circulating SALT solutions with different amplitudes. Since there are two frequencies that are very near to but not exactly equal to one another, the stationary-inversion approximation is invalid because there is a slowly beating term in Eq. 4.1. However, the *magnitude* of this term is small, which leads to the beating being nearly sinusoidal. On the other hand, when we are further above threshold and the magnitude of the beating term in the inversion is larger, the beating is no longer sinusoidal, but rather exhibits limit cycle behavior as seen in Fig. 4-31.

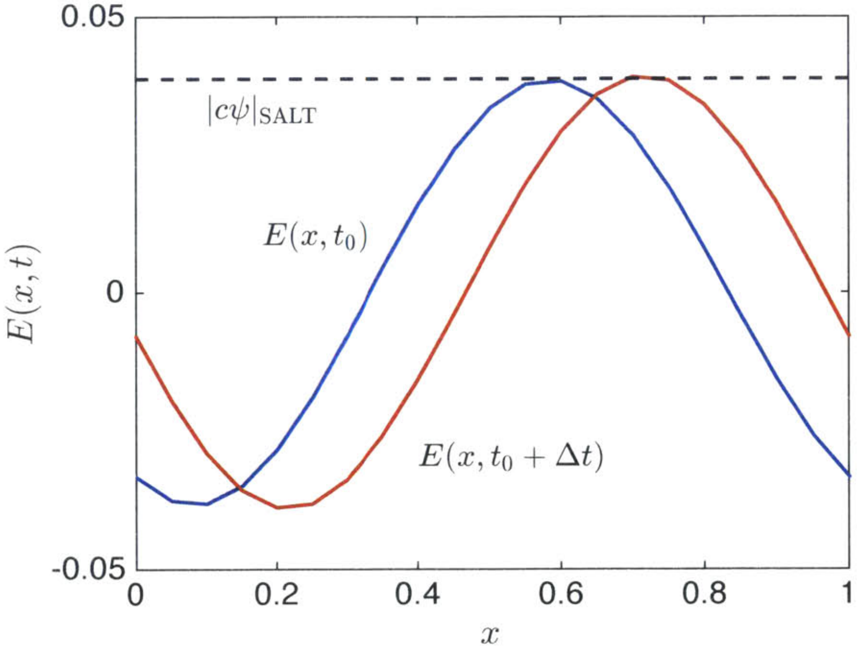


Figure 4-29: Steady-state behavior (after running for a long time) of FDTD simulation of 1d lasing ring with two-fold degeneracy. The field E is shown at two different times, and the prediction of the amplitude from SALT is shown in dotted line. The geometry is 1d with 20 grid points and periodic boundary conditions. The SALT parameters were $\omega_a = 6.25810$, $\gamma_\perp = 0.05$, $\gamma_\parallel = 0.01$, $\sigma = 0.01/\omega$, and pump strength $D_0 = 2 \times 10^{-4}$.

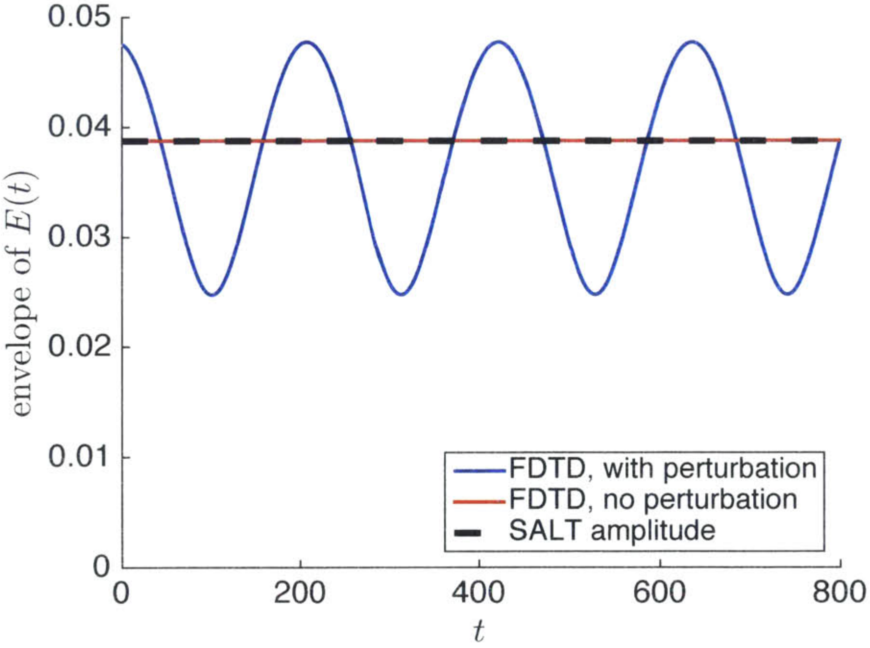


Figure 4-30: Envelopes (max E over each optical cycle) for electric field $E(x_0, t)$ chosen at arbitrary point $x_0 = 0.1$ for 1d ring obtained in FDTD for D_0 slightly above threshold. For the blue curve, there was a small perturbation $\delta\varepsilon = 0.01\cos(4\pi x)$ that splits the frequencies between the sine and cosine modes. The beating frequency here is $\omega_{\text{beating}} \approx 0.0294$, while two-mode SALT predicts a frequency splitting of $\omega_{\sin} - \omega_{\cos} \approx 0.0284$. The beating is an oscillation between left and right-circulating SALT-like solutions (but not with the correct SALT amplitudes). Not shown are the rapid oscillations at $\omega_a \approx 2\pi$. The same envelope with no perturbation (and hence no beating) is shown, as well as the amplitude predicted by SALT.

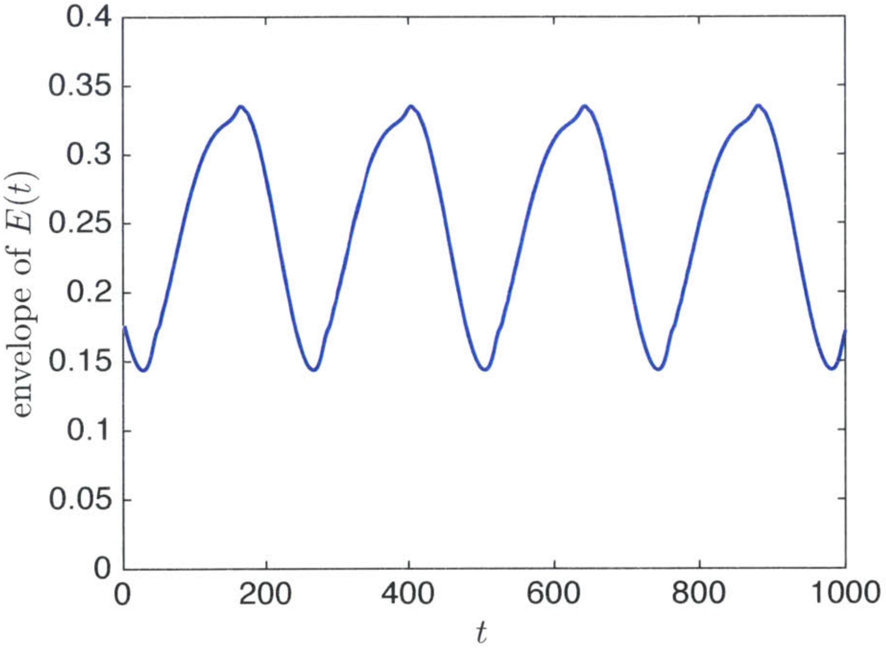


Figure 4-31: Envelopes (max E over each optical cycle) at same point $x_0 = 0.1$ and same parameters as in Fig. 4-30 (with perturbed ε to split the degeneracy), except with pump strength D_0 ten times higher. Unlike the previous figure in which the beating is a simple sinusoid, the oscillations here have a sawtooth shape due to the strong nonlinearities. The behavior is not a SALT-like steady state, but appears to be a stable limit cycle.

The dynamics shown in Fig. 4-30 and Fig. 4-31 indicate that the degeneracy-breaking perturbation causes there to no longer be a stable single-mode lasing solution above threshold. This can be seen from the slow beating periods in those figures: if there *were* a single-mode solution, then the optical period $2\pi/\omega_a \approx 1$ would be the only time scale in the problem. However, Ref. [229] has shown an interesting result in which a point scatterer was placed on the ring (which breaks the degeneracy between sine and cosine modes as expected), but a pair of degenerate "semi-circulating" lasing modes (SALT solutions) branch off from one of the unstable standing modes at higher pump strengths. We have confirmed this fact using our direct solver. Further, the authors of that work have shown, using both stability analysis and FDTD simulation, that under certain parameter regimes, these single-mode solutions are actually stable. These discoveries provide an interesting and unexpected example of degenerate modes when the symmetry has been broken.

4.5 Multi-mode degenerate lasing

So far, the discussion and examples in this chapter have dealt with the case in which only one pair of degenerate modes are lasing. The generalization to the case of multimode lasing (i.e. multiple nondegenerate and degenerate lasing modes all lasing simultaneously) is straightforward. Since our method combines degenerate pairs into a single mode that is the stable linear combination (as given by the perturbation theory in Sec. 4.1, the multimode treatment is exactly the same as for SALT without degeneracies: the degenerate pairs are always treated as a single mode. As in previous work on SALT [26, 28, 197], all lasing modes are solved simultaneously at first, and there the collective effect of their spatial hole-burning is used to track the passive modes and add any mode (degenerate or nondegenerate) that crosses threshold to the list of lasing modes. The only aspects of our method requiring generalization are the threshold perturbation theory of Sec. 4.1 and the QP method of Sec. 4.2.2. For both aspects, we describe small tweaks to the methods presented in those sections that make them valid for the case of multimode lasing.

A general situation in which multimode-lasing is occuring can be described by the equations from chapter 3 [26, 28]

$$\begin{aligned} 0 &= \left(\nabla \times \nabla \times -\omega_\mu^2 \varepsilon_\mu\right) \mathbf{E} \\ \varepsilon_\mu &\equiv \varepsilon_c + D_0 F \gamma(\omega_\mu) H \\ H &\equiv \frac{1}{1+\sum |\mathbf{E}_\nu|^2}, \end{aligned} \tag{4.35}$$

where there are M lasing modes ($\mu = 1, 2, \ldots, M$). If we start with $C_{n\text{v}}$ symmetry and have lasing modes that are either circulating modes (as in Sec. 4.1.3) or non-degenerate modes [partners of *real* 1d irreps $D(k)$ without a corresponding complex-conjugate irrep $\bar{D}(k)$ of the opposite chirality], then each of the $|\mathbf{E}_\nu|^2$ terms has at least C_n symmetry, so H has C_n symmetry. Suppose that the pump strength is at the threshold of mode M so that this mode has just started lasing, and the sum in H ranges from $\nu = 1$ to $\nu = M - 1$. Then all results in Sec. 4.1 still hold, except with gain profile $F(\mathbf{x})$ replaced by FH. While the gain profile is now C_n symmetric rather than $C_{n\text{v}}$, the presence of chiral degenerate pairs (which requires only C_n symmetry and Lorentz reciprocity, as explained in appendix A) still remains, as explained in Sec. 4.1.3 (the arguments in that section do not assume $C_{n\text{v}}$ symmetry, so they still hold even if $F(\mathbf{x})$ is replaced by a function with only C_n symmetry.

Now we consider how to treat the problem of discretization-broken symmetry (Sec. 4.2) in the case of multimode lasing. The method of Sec. 4.2.2 gives a $\delta\varepsilon(\mathbf{x})$ that forces the threshold degeneracy for a *single* mode pair. When there are multiple pairs of nearly degenerate modes the generalization is straightforward: we allow each pair to have its *own* $\delta\varepsilon_\mu$, so that their degeneracies can be forced *independently*. As a result, the dielectric for each pair in Eq. 4.35 will become

$$\varepsilon_\mu = \varepsilon_c + \delta\varepsilon_\mu + D_0 F \gamma(\omega_\mu) H.$$

Since our QP method finds the $\delta\varepsilon_\mu$ with the lowest L_2 norm (as described in Sec. 4.2.2) and the splitting decreases with resolution as seen in Fig. 4-14, each $\delta\varepsilon_\mu$ will independently go to zero as we increase the resolution, so this generalized method is also *convergent*: the unphysical frequency-dependent $\delta\varepsilon_\mu$ vanishes with increasing resolution.

Appendix A

We review the result, given in Ref. [119], of the fact that there are two-fold degeneracies (due to Lorentz reciprocity) in geometries with C_n but not $C_{n\mathrm{v}}$ symmetry, even though there are only one-dimensional irreps. We give a slightly simpler and more general proof by exploiting the differential form of Maxwell's equations, as opposed to the integral form in Ref. [119].

Consider a field $\mathbf{E}^+$ that satisfies the equation $\hat{L}(\omega^+)\mathbf{E}^+ = 0$, where we define the linear operator (as in Ref. [234])

$$\hat{L}(\omega) \equiv -\nabla \times \frac{1}{\mu(\mathbf{x},\omega)}\nabla \times +\omega^2\varepsilon(\mathbf{x},\omega)$$

where ω is the eigenfrequency and ε and μ have C_n symmetry: that is, $R_n\varepsilon R_n^{-1} = \varepsilon$, where R_n is an n-fold rotation and $R_n^n = 1$, the identity operator. Suppose that the field transforms like one of the chiral irreps of C_n: that is, $R_n\mathbf{E}^+ = \exp\left(-\frac{2\pi i m}{n}\right)\mathbf{E}^+$, with $0 < |m| \leq \text{floor}(\frac{n-1}{2})$. We want to show that there exists some other function $\mathbf{E}^-$ that transforms according to the irrep of the opposite chirality *and* has the same eigenfrequency: that is, $R_n\mathbf{E}^- = \exp\left(\frac{2\pi i m}{n}\right)\mathbf{E}^-$ and $\hat{L}(\omega^+)\mathbf{E}^- = 0$.

The key step is to use the right basis: we *could* find the Maxwell eigenfrequencies (Green's-function poles) by solving the nonlinear (in ω) eigenvalue problem $\hat{L}(\omega)\mathbf{E} = 0$. However, these make a poor basis because they diagonalize *different* operators $\hat{L}(\omega)$ with $\omega \neq \omega^+$. Instead, we fix $\omega = \omega^+$ and examine the set of eigenfunctions $\mathbf{E}_j^-$ that satisfy $\hat{L}(\omega^+)\mathbf{E}_j^- = \lambda_j\mathbf{E}_j^-$ and that transform as the $\exp\left(\frac{2\pi i m}{n}\right)$ irrep. Note that λ_j are *not* squared eigenfrequencies and $\mathbf{E}_j^-$ are *not* Maxwell solutions, except for $\lambda_j = 0$. Because this set is a complete basis for functions of this chirality, the function $(\mathbf{E}^+)^\star$ (which transforms in the same way as $\mathbf{E}_j^-$ because the rotation operator R_n is real) can be expanded in this basis:

$$\left(\mathbf{E}^{+}\right)^{\star}=\sum b_j\mathbf{E}_j^{-},$$

assuming that $\hat{L}(\omega^{+})$ is diagonalizable (which is generically true for matrices except at exceptional points; the situation for infinite-dimensional operators is more complicated, but diagonalizability is typically assumed there too in physics). We will now show that at least *one* of these $\mathbf{E}_j^{-}$ is exactly the $\mathbf{E}^{-}$ satisfying $\hat{L}(\omega^{+})\mathbf{E}^{-}=0$ that we are looking for.

First, we define the *unconjugated* inner product $(\mathbf{f},\mathbf{g})\equiv\int d^3x\mathbf{f}\cdot\mathbf{g}$. Then, for appropriate boundary conditions, $\hat{L}(\omega^{+})$ is complex symmetric, that is: $\left(\mathbf{f},\hat{L}\mathbf{g}\right)=\left(\hat{L}\mathbf{f},\mathbf{g}\right)$ for reciprocal materials $\varepsilon=\varepsilon^{T}$, $\mu=\mu^{T}$, and this is known as Lorentz reciprocity [34]. Because $\hat{L}(\omega^{+})$ is complex symmetric, its eigenfunctions with distinct eigenvalues are orthogonal; that is: $\left(\mathbf{E}_i^{-},\mathbf{E}_j^{-}\right)=0$ for $\lambda_i\neq\lambda_j$. Now write

$$\int\left|\mathbf{E}^{+}\right|^{2}=\sum b_j\left(\mathbf{E}^{+},\mathbf{E}_j^{-}\right).$$

If all $\mathbf{E}_j^{-}$ had $\lambda_j\neq0$, then $\left(\mathbf{E}^{+}\cdot\mathbf{E}_j^{-}\right)=0$ for all j. However, the left-hand side is obviously positive, so at least one term in the sum on the right-hand side must be non-vanishing. Hence, this term has the eigenvalue $\lambda_j=0$, and it is precisely the $\mathbf{E}^{-}$ that is degenerate to $\mathbf{E}^{+}$.

Chapter 5

Concluding remarks

In this thesis, we have developed new tools for nonlinear design in nanophotonics. On the topic of transformation optics (TO) [16–22], which is a design method based on coordinate invariance of Maxwell's equations, we have developed a framework that combines previously-used quasiconformal mapping with large-scale inverse design, which solves previously unaddressed issues with optimization constraints and interface continuity. For the equations of steady-state ab-initio lasing theory (SALT), we have developed a procedure, based on Newton's method [1, 3], that directly solves the nonlinear partial-differential equation. Our method circumvents the previously used method of constant-flux basis states, and unlike that method, is scalable to complex 3d geometries of practical interest to the laser community. Examples of work enabled by our numerical tools can be seen in some recent analytical results on laser linewidths [234, 236].

Using this new set of tools, we have shed new light on the fascinating topic of degenerate lasing modes. We have shown that in degenerate lasers, degeneracy and stability are two mutually-dependent qualities, and that just as in the linear problem, there must be a geometric symmetry in the nonlinear problem for there to be stability. We have extended the knowledge about degenerate modes from whispering-gallery modes in cylindrical geometries to chiral lasing modes in arbitrary regular polygons, and have formalized our arguments using a new threshold perturbation theory. We have also devised a numerical procedure that artificially restores broken degeneracies, allowing study of degenerate modes arising from arbitrary symmetry groups using generic grids (such as finite-difference frequency-domain [FDFD] [38, 110]) that do not have that symmetry.

While the tools and methods we have developed are general and open up paths for design and characterization of arbitrary transformation-optical devices and laser geometries, the work in these fields has only just begun. First, there are many improvements to our design method that can be applied, such as allowing interfaces to move freely as long as the Jacobian is continuous. Another area to explore would be to use bases other than Fourier/Chebyshev [95] to parametrize transformations, or algorithms other than COBYLA [93] to optimize these parameters; we made these choices because they seemed reasonable, but there may be others with better performance for transformation inverse design. One may also think about bringing transformation inverse design "to the masses" and creating an interface that can be used by optical physicists or engineers who are not experts in spectral methods or nonlinear optimization, so that they can design any desired integrated component with one click.

We can even go beyond further optimizations on current design method and think about new directions to take. One intriguing possibility is for our transformation inverse-design method to be applied to splitters, expanders, cylindrical cloaks, and another other conceivable 2d geometries. 3d geometries would be a logical next step, and a necessary task to tackle initially would be to extend the connection between 2d conformal transformations and isotropic scalar dielectric media to the third dimension. One can even consider abandoning the isotropic condition entirely, and exploring materials with anisotropy that is still easy to fabricate, which our inverse design method can certainly discover. Finally, we stress that the frequency-independent performance of transformation-optical devices is an idealization; real-world materials always have a non-zero amount of dispersion, so even with a completely isotropic (or even anisotropic) transformation with continuous Jacobian at interfaces, the performance would never be perfect for all frequencies. To this effect, we can conceive of using transformation optics not as a rule set in stone, but more as a guideline: transformation inverse design could be used in tandem with other design and optimization methods that are more frequency or polarization-dependent to create metamaterials and photonic devices with truly optimal performance.

As for our work on lasing modes and SALT, there are several further improvements and directions that immediately come to mind. First, as mentioned in chapter 3, we may consider using iterative solvers [112] (which have much lower memory and time requirements than direct solvers) combined with appropriate preconditioners for each linear Newton step, allowing even larger and complex 3d geometries to be explored. Second, another useful extension would be to interface our nonlinear solver with nonlinear optimization algorithms [94] in order to perform large-scale inverse design of laser geometries. While this type of lasing design would involve solving Maxwell's equations, unlike in transformation inverse design, the computational cost would be comparable to inverse design that involves solving linear Maxwell's equations, since each step the optimization algorithm takes will only change the lasing parameters by a small amount, so the Newton algorithm will always only require a few linear solves to update the Maxwell solution. Finally, just as in the case of transformation inverse design, it will be advantageous, once some of these optimization and improvements are implemented, to combine them into a software package that brings SALT-based laser design "to the masses" who are not experts in nonlinear eigenproblems or semiclassical laser theory.

Our work on degenerate lasing modes has raised some interesting physics questions as well. First, in Sec. 4.5, we briefly discussed the case of multimode lasing with many degenerate pairs of lasing modes. Theoretically, since a degenerate lasing pair can be thought of as a single mode with the correct stable superposition of the 2d subspace of threshold solutions, the case of degenerate multimode lasing should in principle be similar to the nondegenerate multimode case. However, because we never explicitly investigated this regime, there could in principle be behavior that is qualitatively different from the nondegenerate case, and it could be a subject of future of work. Second, our entire work here deals only with 2d geometries, which can only have symmetry groups with 2d irreps and hence two-fold degeneracies [117, 118]. For 3d lasing geometries with the symmetry group of regular polyhedrons, there will be 3d irreps and three-fold degeneracies. The 2x2 perturbation theory that we introduced in Sec. 4.1. would also need to be extended to the 3d case, and since the dimensionality of the degenerate subspace is now three, the resulting first-order equations would involve 3x3 matrices. It remains a fascinating open question whether any interesting physics can happen in such cases. Finally, in Sec. 4.4, we presented time-domain simulations of a slightly-broken degeneracy in a 1d ring laser. We saw the breakdown of SALT behavior, which was somewhat expected due to the stationary-inversion approximation clearly being violated in this regime. However, a comprehensive theory for the near-degenerate regime is an ongoing work, although headway has be made recently in Ref. [229]. An extreme example would be the case of vertical-cavity surface-emitting lasers [237], in which a continuum of modes near a photonic-crystal band edge all lase more or less simultaneously. While we predict exactly degenerate chiral lasing modes for n-fold symmetric systems in chapter 4, perfect symmetry never exists in real-world fabricated systems. There is always disorder, flaws, and geometric asymmetry. It remains an open question what effects such asymmetry will have on the dynamics of lasers.

Bibliography

[1] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: the Art of Scientific Computing.* Cambridge, third edition, 2007.

[2] Y. Su and Z. Bai. Solving rational eigenvalue problems via linearization. *SIAM J. Matrix. Anal. Appl.*, 32(1):201–216, 2011.

[3] G. Peters and J. H. Wilkinson. Inverse iteration, ill-conditioned equations and Newton's method. *SIAM Rev.*, 21:339–360, 1979.

[4] W. R. Frei, H. T. Johnson, and K. D. Choquette. Optimization of a single defect photonic crystal laser cavity. *J. Appl. Phys.*, 103(033102), 2008.

[5] C. Y. Kao and F. Santosa. Maximization of the quality factor of an optical resonator. *Wave Motion*, 45(4):412–427, 2008.

[6] D. C. Dobson and F. Santosa. Optimal localization of eigenfunctions in an inhomogeneous medium. *SIAM J. Appl. Math*, 64(3):762–774, 2004.

[7] J. Lu and J. Vuckovic. Inverse design of nanophotonic structures using complementary convex optimization. *Opt. Express*, 18(4):3793–3804, 2010.

[8] J. Vuckovic, M. Loncar, H. Mabuchi, and A. Scherer. Design of photonic crystal microcavities for cavity QED. *Phys. Rev. E*, 65(016608), 2002.

[9] L. Frandsen, A. Harpøth, P. Borel, M. Kristensen, J. Jensen, and O. Sigmund. Broadband photonic crystal waveguide 60° bend obtained utilizing topology optimization. *Opt. Express*, 12(24):5916–5921, 2004.

[10] W. R. Frei, H. T. Johnson, and D. A. Tortorelli. Optimization of photonic nanostructures. *Computer Methods in Applied Mechanics and Engineering*, 197(41):3410–3416, 2008.

[11] J. S. Jensen and O. Sigmund. Systematic design of photonic crystal structures using topology optimization: Low-loss waveguide bends. *Appl. Phys. Lett.*, 84(12):2022–2024, 2004.

[12] J. Andkjaer and O. Sigmund. Topology optimized low-contrast all-dielectric optical cloak. *Appl. Phys. Lett.*, 98(021112), 2011.

[13] P. Borel, A. Harpøth, L. Frandsen, M. Kristensen, P. Shi, J. Jensen, and O. Sigmund. Topology optimization and fabrication of photonic crystal structures. *Opt. Express*, 12(9):1996, 2004.

[14] S. J. Cox and D. C. Dobson. Maximizing band gaps in two-dimensional photonic crystals. *SIAM J. Appl. Math*, 59(6):2108–2120, 1999.

[15] J. S. Jensen and O. Sigmund. Topology optimization of photonic crystal structures: a high-bandwidth low-loss t-junction waveguide. *J. Opt. Soc. Am. B*, 22(6):1191–1198, 2005.

[16] A. Ward and J. Pendry. Refraction and geometry in maxwell's equations. *J. Mod. Opt.*, 43:773–793, 1996.

[17] J. B. Pendry, D. Schurig, and D. R. Smith. Controlling electromagnetic fields. *Science*, 312:1780–1782, 2006.

[18] H. Chen, C. T. Chan, and P. Sheng. Transformation optics and metamaterials. *Nat. Mat.*, 9:387–396, 2010.

[19] Y. Liu and X. Zhang. Recent advances in transformation optics. *Nanoscale*, 4(17):5277–5292, 2012.

[20] A. V. Novitsky. Inverse problem in transformation optics. *J. Opt.*, 13(035104), 2011.

[21] U. Leonhardt and T. G. Philbin. Transformation optics and the geometry of light. *Prog. Opt.*, 53:69–152, 2009.

[22] Q. Wu, J. P. Turpin, and D. H. Werner. Integrated photonic systems based on transformation optics enabled gradient index devices. *Light: Science and Applications*, 1(e38), 2012.

[23] L. Gabrielli, D. Liu, S. G. Johnson, and M. Lipson. On-chip transformation optics for multimode waveguide bends. *Nat. Comm.*, 3(1217), 2012.

[24] A. Cerjan, Y. Chong, L. Ge, and A. D. Stone. Steady-state ab initio laser theory for n-level lasers. *Opt. Express*, 20(1):474–488, 2012.

[25] L. Ge, R. J. Tandy, A. D. Stone, and H. E. Türeci. Quantitative verification of ab initio self-consistent laser theory. *Opt. Express*, 16(21):16895–16902, 2008.

[26] L. Ge, Y. D. Chong, and A. D. Stone. Steady-state ab initio laser theory: Generalizations and analytic results. *Phys. Rev. A*, 82(6):063824, 2010.

[27] H. Türeci, A. D. Stone, L. Ge, S. Rotter, and R. J. Tandy. Ab initio self-consistent laser theory and random lasers. *Nonlinearity*, 22(1):C1–C18, 2009.

[28] H. E. Türeci, A. D. Stone, and B. Collier. Self-consistent multimode lasing theory for complex or random lasing media. *Phys. Rev. A*, 74(4):043822, 2006.

[29] L. Ge. *Steady-state Ab Initio Laser Theory and its Applications in Random and Complex Media.* PhD thesis, Yale University, 2010.

[30] H. E. Türeci, L. Ge, S. Rotter, and A. D. Stone. Strong interactions in multimode random lasers. *Science*, 320(5876):643–646, 2008.

[31] L. Ge, D. Liu, S. G. Johnson, S. Rotter, H. E. Tureci, H. Cao, A. Cerjan, and A. D. Stone. Interaction-induced mode switching in microlasers. arXiv 1504.08018, 2015.

[32] V. S. Ilchenko and A. B. Matsko. Optical resonators with whispering-gallery modes-part i: Basics. *IEEE J. Sel. Topics Quantum Electron.*, 12(1):3–14, 2006.

[33] H. Haken. *Light: Laser light dynamics.* North-Holland Pub. Co., 1985.

[34] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade. *Photonic Crystals: Molding the Flow of Light.* Princeton, second edition, 2008.

[35] T. Ito and S. Okazaki. Pushing the limits of lithography. *Nature*, 406:1027–1031, 2000.

[36] W. C. Chew, J. M. Jin, E. Michielssen, and J. M. Song, editors. *Fast and Efficient Algorithms in Computational Electromagnetics.* Artech House, 2001.

[37] A. Taflove and S. C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method.* Artech House, third edition, 2005.

[38] N. J. Champagne, J. G. Berryman, and H. M. Buettner. FDFD: A 3d finite-difference frequency-domain code for electromagnetic induction tomography. *J. Comp. Phys.*, 170:830–848, 2001.

[39] A. Logg, K. A. Mardal, and G. N. Wells. *Automated Solution of Differential Equations by the Finite Element Method.* Springer, 2012.

[40] J. S. Jensen and O. Sigmund. Topology optimization for nano-photonics. *Laser and Photonics Reviews*, 5(2):308–321, 2011.

[41] Y. Watanabe, N. Ikeda, Y. Sugimoto, Y. Takata, Y. Kitagawa, A. Mizutani, N. Ozaki, and K. Asakawa. Topology optimization of waveguide bends with wide, flat bandwidth in air-bridge-type photonic crystal slabs topology optimization of waveguide bends with wide, flat bandwidth in air-bridge-type photonic crystal slabs topology optimization of waveguide bends with wide, flat bandwidth in air-bridge-type photonic crystal slabs. *J. Appl. Phys.*, 101(113108), 2007.

[42] Y. Tsuji and K. Hirayama. Design of optical circuit devices using topology optimization method with function-expansionbased refractive index distribution. *IEEE Phot. Tech. Lett.*, 20(12):982–984, 2008.

[43] C. Y. Yao, S. Osher, and E. Yablonovitch. Maximizing band gaps in two-dimensional photonic crystals by using level set methods. *Appl. Phys. B*, 81(2–3):235–244, 2005.

[44] W. R. Frei, D. A. Tortorelli, and H. T. Johnson. Topology optimization of a photonic crystal waveguide termination to maximize directional emission. *Appl. Phys. Lett.*, 86(111114), 2005.

[45] D. Schurig, J. B. Pendry, and D. R. Smith. Transformation-designed optical elements. *Opt. Express*, 15(22):14772–14782, 2007.

[46] M. Lax and D. F. Nelson. Maxwell's equations in material form. *Phys. Rev. B*, 13(1777), 1976.

[47] J.-P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 114(2):185–200, 1994.

[48] W. Shin and S. Fan. Choice of the perfectly matched layer boundary condition for frequency domain Maxwell's equations solvers. *J. Comp. Phys.*, 231:3406–3431, 2012.

[49] D. Schurig, J. B. Pendry, and D. R. Smith. Calculation of material properties and ray tracing in transformation media. *Opt. Express*, 14(21):9794–9804, 2006.

[50] U. Leonhardt and T. Tyc. Broadband invisibility by non-euclidian cloaking. *Science*, 323(5910):110–112, 2009.

[51] J. Li and J. B. Pendry. Hiding under the carpet: A new strategy for cloaking. *Phys. Rev. Lett.*, 101(203901), 2008.

[52] R. Liu, R. C. Ji, J. J. Mock, J. Y. Chin, T. J. Cui, and D. R. Smith. Broadband ground-plane cloak. *Science*, 323(5912):366–369, 2009.

[53] J. Valentine, J. Li, T. Zentgraf, G. Bartal, and X. Zhang. An optical cloak made of dielectrics. *Nat. Mat.*, 8:568–571, 2009.

[54] T. Ergin, N. Stenger, P. Brenner, J. B. Pendry, and M. Wegener. Three-dimensional invisibility cloak at optical wavelengths. *Science*, 328(5976):337–339, 2010.

[55] J. H. Lee, J. Blair, V. A. Tamma, Q. Wu, S. J. Rhee, C. J. Summers, and W. Park. Direct visualization of optical frequency invisibility cloak based on silicon nanorod array. *Opt. Express*, 17(15):12922–12928, 2009.

[56] H. Gao, B. Zhang, S. G. Johnson, and G. Barbastathis. Design of thin-film photonic metamaterial Lüneburg lens using analytical approach. *Opt. Express*, 20:1617–1628, 2012.

[57] N. Kundtz and D. R. Smith. Extreme-angle broadband metamaterial lens. *Nat. Mat.*, 9:129–132, 2010.

[58] H. F. Ma and T. J. Cui. Three-dimensional broadband and broad-angle transformation-optics lens. *Nat. Comm.*, 1(124), 2010.

[59] B. U. Chen, E. Marom, and A. Lee. Geodesic lenses in single-mode $LiNbO_3$ waveguides. *Appl. Phys. Lett.*, 31(4):263, 1977.

[60] F. Zernike. Luneburg lens for optical waveguide use. *Opt. Commun.*, 12(4):379–381, 1974.

[61] A. Chutinan, M. Okano, and S. Noda. Wider bandwidth with high transmission through waveguide bends in two-dimensional photonic crystal slabs. *Appl. Phys. Lett.*, 80(10):1698–1699, 2002.

[62] Z. L. Mei and T. J. Cui. Arbitrary bending of electromagnetic waves using isotropic materials. *J. Appl. Phys.*, 105(104913), 2009.

[63] H. Xu, B. Zhang, Y. Yu, G. Barbastathis, and H. Sun. Dielectric waveguide bending adapter with ideal transmission. *Opt. Express*, 29(6):1287–1290, 2012.

[64] V. Liu and S. Fan. Compact bends for multi-mode photonic crystal waveguides with high transmission and suppressed modal crosstalk. *Opt. Express*, 21(7):8069–8075, 2013.

[65] Z. L. Mei and T. T. Cui. Experimental realization of a broadband bend structure using gradient index metamaterials. *Opt. Express*, 17(20):18354–18363, 2009.

[66] G. P. Agrawal. *Fiber-Optic Communication Systems.* Wiley, fourth edition, 2010.

[67] L. A. Coldren, S. W. Corzine, and M. L. Mashanovitch. *Diode Lasers and Photonic Integrated Circuits.* Wiley, second edition, 2012.

[68] J. Huangfu, S. Xi, F. Kong, J. Zhang, H. Chen, D. Wang, B. Wu, L. Ran, and J. A. Kong. Application of coordinate transformation in bent waveguides. *J. Appl. Phys.*, 104(014502), 2008.

[69] M. Yin, X. Y. Tian, H. X. Han, and D. C. Li. Free-space carpet-cloak based on gradient index photonic crystals in metamaterial regime. *Appl. Phys. Lett.*, 100(124101), 2012.

[70] J. H. Bramble and J. E. Pasciak. Analysis of a finite element PML approximation for the three dimensional time-harmonic Maxwell problem. *Math. Comp.*, 77(261):1–10 (electronic), 2008.

[71] D. Smith, J. Mock, A. Starr, and D. Schurig. Gradient index metamaterials. *Phys. Rev. E*, 71(036609), 2005.

[72] D. R. Smith, J. B. Pendry, and M. C. K. Wiltshire. Metamaterials and negative refractive index. *Science*, 305(5685):788–792, 2004.

[73] F. Xu, R. C. Tyan, P. C. Sun, Y. Fainman, C. C. Cheng, and A. Scherer. Fabrication, modeling, and characterization of form-birefringent nanostructures. *Opt. Lett.*, 20(24):2457–2459, 1995.

[74] D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith. Metamaterial electromagnetic cloak at microwave frequencies. *Science*, 314(5801):977–980, 2006.

[75] L. H. Gabrielli and M. Lipson. Transformation optics on a silicon platform. *J. Opt.*, 13(024010), 2011.

[76] R. Murali, D. K. Brown, K. P. Martin, and J. D. Meindl. Process optimization and proximity effect correction for gray scale e-beam lithography. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 24:2936–2939, 2006.

[77] L. Mosher, C. M. Waits, B. Morgan, and R. Ghodssi. Double-Exposure Grayscale Photolithography. *Journal of Microelectromechanical Systems*, 18:308–315, 2009.

[78] J. Kim, D. Joy, and S.-Y. Lee. Controlling resist thickness and etch depth for fabrication of 3D structures in electron-beam grayscale lithography. *Microelectronic Engineering*, 84:2859–2864, 2007.

[79] L. Gabrielli and M. Lipson. Integrated luneburg lens via ultra-strong index gradient on silicon. *Opt. Express*, 19(21):20122–20127, 2011.

[80] H. Kurt and D. S. Citrin. Graded index photonic crystals. *Opt. Express*, 15(3):1240–1253, 2007.

[81] P. Knupp and S. Steinberg. *Fundamentals of Grid Generation.* CRC Press, 1999.

[82] J. F. Thompson, B. K. Soni, and N. P. Weatherill. *Handbook of Grid Generation.* CRC Press, 1999.

[83] M. Rahm, D. A. Roberts, J. B. Pendry, and D. R. Smith. Transformation-optical design of adaptive beam bends and beam expanders. *Opt. Express*, 16(15):11555–11567, 2008.

[84] B. Vasić, G. Isić, R. Gajić, and K. Hingerl. Coordinate transformation based design of confined metamaterial structures. *Phys. Rev. B*, 79:085103, 2009.

[85] W. Rudin. *Real and Complex Analysis.* McGraw–Hill, third edition, 1986.

[86] G. E. Shilov. *Elementary Real and Complex Analysis.* MIT Press, 1973.

[87] M. Schmiele, V. S. Varma, C. Rockstuhl, and F. Lederer. Designing optical elements from isotropic materials by using transformation optics. *Phys. Rev. A*, 81(033837), 2010.

[88] T. Han, C. Qiu, J. Dong, X. Tang, and S. Zouhdi. Homogeneous and isotropic bends to tunnel waves through multiple different/equal waveguides along arbitrary directions. *Opt. Express*, 19(14):13020–13030, 2011.

[89] Z. Chang, X. Zhou, J. Hu, and G. Hu. Design method for quasi-isotropic transformation materials based on inverse Laplace's equation with sliding boundaries. *Opt. Express*, 18(6):6089–6096, 2010.

[90] W. Yan, M. Yan, and M. Qiu. Necessary and sufficient conditions for reflectionless transformation media in an isotropic and homogenous background. arXiv:0806.3231, 2008.

[91] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, 2004.

[92] J. Lu and J. Vuckovic. Inverse design of nanophotonic structures using complementary convex optimization. *Opt. Express*, 18(4):3793–3804, 2010.

[93] M. J. D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In S. Gomez and J. P. Hennart, editors, *Proceedings of the Sixth workshop on Optimization and Numerical Analysis*, pages 51–67. Kluwer Academic Publishers, 1994.

[94] S. G. Johnson. The NLopt nonlinear-optimization package (http://ab-initio.mit.edu/nlopt), 2007.

[95] J. P. Boyd. *Chebyshev and Fourier Spectral Methods.* Dover, second edition, 2001.

[96] M. Sargent, M. O. Scully, and W. E. Lamb. *Laser Physics.* Westview Press, 1977.

[97] H. Haken. *Laser Theory.* Springer, 1970.

[98] H. Haken and H. Sauermann. Nonlinear interactions of laser modes. *Z. Phys.*, 173(261), 1963.

[99] W. E. Lamb. Theory of optical maser. *Phys. Rev.*, 134(A1429), 1964.

[100] H. Fu and H. Haken. Multifrequency operations in a short-cavity standing-wave laser. *Phys. Rev. A*, 43(5):2446–2454, 1991.

[101] R. W. Ziolkowski, J. M. Arnold, and D. M. Gogny. Ultrafast pulse interactions with two-level atoms. *Phys. Rev. A*, 52:3082–3094, 1995.

[102] A. S. Nagra and R. A. York. FDTD analysis of wave propagation in nonlinear absorbing and gain media. *Antennas Propag.*, 46:334–340, 1998.

[103] P. Bermel, E. Lidorikis, and J. D. Joannopoulos. Active materials embedded in photonic crystals and coupled to electromagnetic radiation. *Phys. Rev. B*, 73(165125), 2006.

[104] P. Guillaume. Nonlinear eigenproblems. *SIAM J. Matrix. Anal. Appl.*, 20:575–595, 1999.

[105] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules.* Oxford University Press, first edition, 1989.

[106] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B), 1964.

[107] K. Burke, J. Werschnik, and E. K. U. Gross. Time-dependent density functional theory: Past, present, and future. *J. Chem. Phys.*, 123(6), 1964.

[108] T. J. Kippenberg, J. Kalkman, A. Polman, and K. J. Valhala. Demonstration of an erbium-doped microdisk laser on a silicon chip. *Phys. Rev. A*, 74(051802), 2006.

[109] H. Altug, D. Englund, and J. Vuckovic. Ultrafast photonic crystal nanocavity laser. *Nat. Phys.*, 2:484–488, 2006.

[110] A. Christ and H. L. Hartnagel. Three-dimensional finite-difference method for the analysis of microwave-device embedding. *IEEE Trans. Microwave. Theory and Tech.*, 35:688–696, 1987.

[111] P. Solin. *Partial Differential Equations and the Finite Element Method.* Wiley, 2005.

[112] L. N. Trefethen and D. Bau. *Numerical Linear Algebra.* SIAM, 1997.

[113] P. Hénon, P. Ramet, and J. Roman. Pastix: a high-performance parallel direct solver for sparse symmetric positive definite systems. *Parallel Computing*, 28(2):301 – 321, 2002.

[114] T. A. Davis. *Direct Methods for Sparse Linear Systems.* SIAM, 2006.

[115] C. P. Dettmann, G. V. Morozov, M. Sieber, and H. Waalkens. Internal and external resonances of dielectric disks. *Europhys. Lett.*, 87:34003, 2009.

[116] K. J. Vahala. Optical microcavities. *Nature*, 424(6950):839–846, 2003.

[117] T. Inui, Y. Tanabe, and Y. Onodera. *Group Theory and Its Applications in Physics.* Springer, 1996.

[118] M. Tinkham. *Group Theory and Quantum Mechanics.* Dover, second edition, 2003.

[119] B. Hopkins, Y. Hwang, A. N. Poddubny, A. E. Miroshnichenko, T. J. Davis, and Y. S. Kivshar. Circular dichroism induced by fano resonances in planar chiral oligomers. arXiv 1412.1120v2, 2015.

[120] D. Liu, L. Gabrielli, M. Lipson, and S. G. Johnson. Transformation inverse design. *Opt. Express*, 21(12):14223–14243, 2013.

[121] N. I. Landy and W. J. Padilla. Guiding light with conformal transformations. *Opt. Express*, 17(17):14872–14879, 2009.

[122] M. Heiblum and J. H. Harris. Analysis of curved optical waveguides by conformal transformation. *J. Quantum Electronics*, QE-11(2):75–83, 1975.

[123] Y. Ma, N. Wang, and C. K. Ong. Application of inverse, strict conformal transformation to design waveguide devices. *J. Opt. Soc. Am. A*, 27:968–972, 2010.

[124] S. Han, Y. Xiong, D. Genov, Z. Liu, G. Bartel, and X. Zhang. Ray optics at a deep-subwavelength scale: A transformation optics approach. *Nano Lett.*, 8(12):4243–4247, 2008.

[125] K. Yao and X. Jiang. Designing feasible optical devices via conformal mapping. *J. Opt. Soc. Am. B*, 28(5):1037–1042, 2011.

[126] D. A. Roberts, M. Rahm, J. B. Pendry, and D. R. Smith. Transformation-optical design of sharp waveguide bends and corners. *Appl. Phys. Lett.*, 93(251111), 2008.

[127] C. García-Meca, M. M. Tung, J. V. Galán, R. Ortuño, F. J. Rodríguez-Fortuño, J. Martí, and A. Martínez. Squeezing and expanding light without reflections via transformation optics. *Opt. Express*, 19(4):3562–3757, 2011.

[128] X. Zhang, H. Chen, X. Luo, and H. Ma. Transformation media that turn a narrow slit into a large window. *Opt. Express*, 16(16):11764–11768, 2008.

[129] O. Ozgun and M. Kuzuoglu. Utilization of anisotropic metamaterial layers in waveguide miniaturization and transitions. *IEEE Microw. Wirel. Compon. Lett.*, 17:754, 2007.

[130] B. Zhang, Y. Luo, X. Liu, and G. Barbastathis. Macroscopic invisibility cloak for visible light. *Phys. Rev. Lett.*, 106(033901), 2011.

[131] Z. Liang, X. Jiang, F. Miao, S. Guenneau, and J. Li. Transformation media with variable optical axes. *New J. Phys.*, 14(103042), 2012.

[132] Z. L. Mei, Y. S. Liu, F. Yang, and T. J. Cui. A DC carpet cloak based on resistor networks. *Opt. Express*, 20(23):25758–25764, 2012.

[133] S. Wang and S. Liu. Controlling electromagnetic scattering of a cavity by transformation media. *Opt. Express*, 20(6):6777–6785, 2012.

[134] X. Chen, Y. Luo, J. Zhang, K. Jiang, J. B. Pendry, and S. Zhang. Macroscopic invisibility cloaking of visible light. *Nat. Comm.*, 2(176), 2011.

[135] J. Mei, Q. Wu, and K. Zhang. Multimultifunctional complementary cloak with homogeneous anisotropic material parameters. *J. Opt. Soc. Am. A*, 29(10):2067–2073, 2012.

[136] C. Garcia-Meca, A. Martinez, and U. Leonhardt. Engineering antenna radiation patterns via quasi-conformal mappings. *Opt. Express*, 19(24):23743–23750, 2011.

[137] N. I. Landy, N. Kundtz, and D. R. Smith. Designing three-dimensional transformation optical media using quasiconformal coordinate transformations. *Phys. Rev. Lett.*, 105(193902), 2010.

[138] Z. L. Mei, J. Bai, and T. J. Cui. Illusion devices with quasi-conformal mapping. *J. Electromagnetic Waves and Applications*, 24(17):2561–2573, 2010.

[139] L. Gabrielli, J. Cardenas, C. B. Poitras, and M. Lipson. Silicon nanostructure cloak operating at optical frequencies. *Nat. Phot.*, 3:461–463, 2009.

[140] P. Markov, J. G. Valentine, and S. M. Weiss. Fiber-to-chip coupler designed using an optical transformation. *Opt. Express*, 20(13):14705–14712, 2012.

[141] Z. L. Mei, J. Bai, and T. J. Cui. Experimental verification of a broadband planar focusing antenna based on transformation optics. *New J. Phys.*, 13(063028), 2011.

[142] L. Tang, J. Yin, G. Yuan, J. Du, H. Gao, X. Dong, Y. Lu, and C. Du. General conformal transformation method based on Schwarz–Christoffel approach. *Opt. Express*, 19(16):15119–15126, 2011.

[143] J. P. Turpin, A. T. Massoud, Z. H. Jiang, P. L. Werner, and D. H. Wener. Conformal mappings to achieve simple material parameters for transformation optics devices. *Opt. Express*, 18(1):244–252, 2010.

[144] D. R. Smith, Y. Urzhumov, N. B. Kundtz, and N. I. Landy. Enhancing imaging systems using transformation optics. *Opt. Express*, 18(20):21238–21251, 2010.

[145] J. Riishede and O. Sigmund. Inverse design of dispersion compensating optical fiber using topology optimization. *J. Opt. Soc. Am. B*, 25(1):88–97, 2008.

[146] C. Y. Kao and S. Osher. Incorporating topological derivatives into shape derivatives based level set methods. *J. Comp. Phys.*, 225(1):891–909, 2007.

[147] P. Seliger, M. Mahvash, C. Wang, and A. F. J. Levi. Optimization of aperiodic dielectric structures. *J. Appl. Phys.*, 100(034310), 2006.

[148] A. Mekis, J. C. Chen, I. Kurland, S. Fan, P. R. Villeneuve, and J. D. Joannopoulos. High transmission through sharp bends in photonic crystal waveguides. *Phys. Rev. Lett.*, 77:3787–3790, 1996.

[149] C. Ma, Q. Zhang, and E. V. Keuren. Right-angle slot waveguide bends with high bending efficiency. *Opt. Express*, 16(19):14330, 2008.

[150] V. Liu and S. Fan. Compact bends for multi-mode photonic crystal waveguides with high transmission and suppressed modal crosstalk. *Opt. Express*, 21(7):8069–8075, 2013.

[151] A. Chutinan and S. Noda. Highly confined waveguides and waveguide bends in three-dimensional photonic crystal. *Appl. Phys. Lett.*, 75(24):3739–3741, 1999.

[152] B. Chen, T. Tang, and H. Chen. Study on a compact flexible photonic crystal waveguide and its bends. *Opt. Express*, 17(7):5033–5038, 2009.

[153] Y. Zhang and B. Li. Photonic crystal-based bending waveguides for optical interconnections. *Opt. Express*, 14(12):5723–5732, 2006.

[154] J. Smajic, C. Hafner, and D. Erni. Design and optimization of an achromatic photonic crystal bend. *Opt. Express*, 11(12):1378–1384, 2003.

[155] M. Rahm, S. A. Cummer, D. Schurig, J. B. Pendry, and D. R. Smith. Optical design of reflectionless complex media by finite embedded coordinate transformations. *Phys. Rev. Lett.*, 100(063903), 2008.

[156] B. Vasić, R. Gajić, and K. Hingerl. Graded photonic crystals for implementation of gradient refractive index media. *J. Nanophotonics*, 5(051806), 2011.

[157] U. Levy, M. Abashin, K. Ikeda, A. Krishnamoorthy, J. Cunningham, and Y. Fainman. Inhomogenous dielectric metamaterials with space-variant polarizability. *Phys. Rev. Lett.*, 98(243901), 2007.

[158] Y. Wang, C. Sheng, H. Liu, Y. Zheng, C. Zhu, S. M. Wang, and S. N. Zhu. Transformation bending device emulated by graded-index waveguide. *Opt. Express*, 20(12):13006–13013, 2012.

[159] R. Ulrich and R. J. Martin. Geometric optics in thin film light guides. *Appl. Opt.*, 10(9):2077–2085, 1971.

[160] J. Brazas, G. Kohnke, and J. McMullen. Mode-index waveguide lens with novel gradient boundaries developed for application to optical recording. *Appl. Opt.*, 31(18):3420–3428, 1992.

[161] J. Li and J. B. Pendry. Private communication, 2013.

[162] L. Bergamin. Electromagnetic fields and boundary conditions at the interface of generalized transformation media. *Phys. Rev. A*, 80(063835), 2009.

[163] W. Yan, M. Yan, Z. Ruan, and M. Qiu. Coordinate transformations make perfect invisibility cloaks with arbitrary shape. *New J. Phys.*, 10(043040), 2008.

[164] O. Weber, A. Myles, and D. Zorin. Computing extremal quasiconformal maps. *Europgraphics Symposium on Geometry Processing*, 31(5):1679–1689, 2012.

[165] K. Astala, T. Iwaniec, and G. Martin. *Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane.* Princeton University Press, first edition, 2008.

[166] A. Papadopoulos, editor. *Handbook of Teichmüller Theory*, volume 1. European Mathematical Society, 2007.

[167] R. Kühnau, editor. *Handbook of Complex Analysis: Geometric Function Theory*, volume 2. Elsevier B.V., 2005.

[168] L. Bers. An extremal problem for quasi-conformal mappings and a problem of Thurston. *Acta Math.*, pages 73–98, 1978.

[169] L. V. Ahlfors. Lectures on quasiconformal mappings. *University Lecture Series*, 38, 1966.

[170] Z. Balogh, K. Fässler, and I. Platis. Modulus of curve families and extremality of spiral-stretch maps. *J. d'Analyse Mathématique*, 113(1):265–291, 2011.

[171] K. Astala, T. Iwaniec, and G. Martin. Deformations of annuli with smallest mean distortion. *Arch. Rational Mech. Anal.*, 195:899—921, 2010.

[172] Z. Balogh, K. Fässler, and I. Platis. Modulus method and radial stretch map in the heisenberg group. *Annales Academiae Scientiarum Fennicae*, 38(1):149–180, 2013.

[173] J. D. Jackson. *Classical Electrodynamics.* Wiley, third edition, 1998.

[174] L. Ahlfors. On quasiconformal mappings. *J. d'Analyse Mathématique*, 3(1):1–58, 1953.

[175] W. Zeng, F. Luo, S. T. Yau, and X. D. Gu. Surface quasi-conformal mapping by solving Beltrami equations. In *Proc. 13th International Conference on Mathematics of Surfaces*, pages 391–408, 2009.

[176] S. G. Johnson, M. L. Povinelli, M. Soljačić, A. Karalis, S. Jacobs, and J. D. Joannopoulos. Roughness losses and volume-current methods in photonic-crystal waveguides. *Appl. Phys. B*, 81(283–293), 2005.

[177] A. W. Snyder and J. D. Love. *Optical Waveguide Theory.* Chapman and Hall, London, 1983.

[178] W. C. Chew. *Waves and Fields in Inhomogeneous Media.* IEEE Press, New York, NY, 1995.

[179] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numerica*, 7:287–336, 1998.

[180] A. Oskooi, A. Mutapcic, S. Noda, J. D. Joannopoulos, S. P. Boyd, and S. G. Johnson. Robust optimization of adiabatic tapers for coupling to slow-light photonic-crystal waveguides. *Opt. Express*, 20(19):21558–21575, 2012.

[181] A. Mutapcica, S. Boyd, A. Farjadpour, S. G. Johnson, and Y. Avniel. Robust design of slow-light tapers in periodic waveguides. *Engineering Optimization*, 41(4):365–384, 2009.

[182] T. F. Chan, J. Cong, T. Kong, and J. R. Shinnerl. Multilevel optimization for large-scale circuit placement. In *International Conference on Computer Aided Design*, pages 171–176, 2000.

[183] K. W. Chun and J. Ra. Fast block-matching algorithm by successive refinement of matching criterion. In P. Maragos, editor, *Proc. SPIE, Visual Communications and Image Processing*, volume 1818, pages 552–560, 1992.

[184] S. K. Yao, D. B. Anderson, R. R. August, B. R. Youmans, and C. M. Oania. Guided-wave optical thin-film Luneburg lenses: fabrication technique and properties. *Appl. Opt.*, 18:4067–4079, 1979.

[185] J. Brazas, G. Kohnke, and J. McMullen. Mode-Index Waveguide Lens With Novel Gradient Boundaries Developed For Application To Optical Recording. *Appl. Opt.*, 31:3420–3428, 1992.

[186] A. Di Falco, S. C. Kehr, and U. Leonhardt. Luneburg lens in silicon photonics. *Opt. Express*, 19:5156–5162, 2011.

[187] S. Randel, R. Ryf, A. Sierra, P. J. Winzer, A. H. Gnauck, C. A. Bolle, R.-J. Essiambre, D. W. Peckham, A. McCurdy, and R. Lingle. 6×56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization. *Opt. Express*, 19:16697–16707, 2011.

[188] N. Bai, E. Ip, Y.-K. Huang, E. Mateo, F. Yaman, M.-J. Li, S. Bickham, S. Ten, J. Liñares, C. Montero, V. Moreno, X. Prieto, V. Tse, K. Man Chung, A. P. T. Lau, H.-Y. Tam, C. Lu, Y. Luo, G.-D. Peng, G. Li, and T. Wang. Mode-division multiplexed transmission with inline few-mode fiber amplifier. *Opt. Express*, 20:2668–2680, 2012.

[189] N. Riesen, J. D. Love, and J. W. Arkwright. Few-Mode Elliptical-Core Fiber Data Transmission. *IEEE Photonics Technology Letters*, 24:344–346, 2012.

[190] R. Ryf, R. Essiambre, A. Gnauck, S. Randel, M. A. Mestre, C. Schmidt, P. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, T. Hayashi, T. Taru, and T. Sasaki. Space-Division Multiplexed Transmission over 4200 km 3-Core Microstructured Fiber. In *National Fiber Optic Engineers Conference, OSA Technical Digest*, page PDP5C.2. Optical Society of America, 2012.

[191] J. Wang, J.-y. Yang, I. M. Fazal, N. Ahmed, Y. Yan, H. Huang, Y. Ren, Y. Yue, S. Dolinar, M. Tur, and A. E. Willner. Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nat. Phot.*, 6:488–496, 2012.

[192] S. G. Johnson, P. Bienstman, M. A. Skorobogatiy, M. Ibanescu, E. Lidorikis, and J. D. Joannopoulos. Adiabatic theorem and continuous coupled-mode theory for efficient taper transitions in photonic crystals. *Phys. Rev. E*, 66(066608), 2002.

[193] C. Manolatou, S. G. Johnson, S. Fan, P. R. Villeneuve, H. A. Haus, and J. D. Joannopoulos. High-density integrated optics. *J. Lightwave Technology*, 17(9):1682–1692, 1999.

[194] O. Painter, R. K. Lee, A. Scherer, A. Yariv, J. D. O'Brien, P. D. Dapkus, and I. Kim. Two-dimensional photonic band-gap defect mode laser. *Science*, 284:1819–1821, 1999.

[195] S. Noda, A. Chutinan, and M. Imada. Trapping and emission of photons by a single defect in a photonic bandgap structure. *Nature*, 407:608–610, 2000.

[196] B. Ellis, M. A. Mayer, G. Shambat, T. Sarmiento, J. Harris, E. E. Haller, and J. Vuckovic. Ultralow-threshold electrically pumped quantum-dot photonic-crystal nanocavity laser. *Nat. Phot.*, 5:297–300, 2011.

[197] S. Esterhazy, D. Liu, M. Liertzer, A. Cerjan, K. G. Makris, A. D. Stone, J. M. Melenk, S. G. Johnson, and S. Rotter. Scalable numerical approach for the steady-state *ab-initio* laser theory. *Phys. Rev. A*, 90:023816, 2014.

[198] P. Stano and P. Jacquod. Suppression of interactions in multimode random lasers in the anderson localized regime. *Nat. Phot.*, 7(1):66–71, 2013.

[199] T. Hisch, M. Liertzer, D. Pogany, F. Mintert, and S. Rotter. Pump-controlled directional light emission from random lasers. *Phys. Rev. Lett.*, 111(2):023902, 2013.

[200] M. Liertzer, L. Ge, A. Cerjan, A. D. Stone, H. E. Türeci, and S. Rotter. Pump-induced exceptional points in lasers. *Phys. Rev. Lett.*, 108(17):173901, 2012.

[201] S.-L. Chua, Y. Chong, A. D. Stone, M. Soljačić, and J. Bravo-Abad. Low-threshold lasing action in photonic crystal slabs enabled by fano resonances. *Opt. Express*, 19(2):1539–1562, 2011.

[202] Y. D. Chong and A. D. Stone. General linewidth formula for steady-state multimode lasing in arbitrary cavities. *Phys. Rev. Lett.*, 109(6):063902, 2012.

[203] A. Yariv. *Quantum Electronics*. Wiley, 3 edition, 1989.

[204] M. Brandstetter, M. Liertzer, C. Deutsch, P. Klang, J. Schöberl, H. E. Türeci, G. Strasser, K. Unterrainer, and S. Rotter. Reversing the pump dependence of a laser at an exceptional point. *Nat. Comm.*, 5, 2014.

[205] X. Liang and S. G. Johnson. Formulation for scalable optimization of microcavities via the frequency-averaged local density of states. *Opt. Express*, 21(25):30812–30841, 2013.

[206] X. Liang. *Modeling of fluids and waves with analytics and numerics.* PhD thesis, MIT, 2012.

[207] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhäuser, 1997.

[208] G. L. Sleijpen, A. G. Booten, D. R. Fokkema, and H. A. van der Vorst. Jacobi–Davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT Numer. Math.*, 36:595–633, 1996.

[209] A. Ruhe. The rational Krylov algorithm for nonlinear matrix eigenvalue problems. *J. Math. Sci.*, 114:1854–1856, 2003.

[210] H. Voss. An Arnoldi method for nonlinear eigenvalue problems. *BIT Numer. Math.*, 44:387–401, 2004.

[211] J. Asakura, T. Sakurai, H. Tadano, T. Ikegami, and K. Kimura. A numerical method for polynomial eigenvalue problems using contour integral. *Japan J. Appl. Math.*, 27:73–90, 2010.

[212] W.-J. Beyn. An integral method for solving nonlinear eigenvalue problems. *Lin. Alg. Appl.*, 436(10):3839–3863, 2012.

[213] O. I. Tolstikhin, V. N. Ostrovsky, and H. Nakamura. Siegert pseudostate formulation of scattering theory: One-channel case. *Phys. Rev. A*, 58(3):2077–2096, 1998.

[214] L. Ge, Y. D. Chong, S. Rotter, H. E. Türeci, and A. D. Stone. Unconventional modes in lasers with spatially varying gain and loss. *Phys. Rev. A*, 84(2):023820, 2011.

[215] S. Y. Lin, E. Chow, S. G. Johnson, and J. D. Joannopoulos. Direct measurement of the quality factor in a two-dimensional photonic-crystal microcavity. *Opt. Lett.*, 26:1903–1905, 2001.

[216] W. R. Frei, H. T. Johnson, and K. D. Choquette. Optimization of a single defect photonic crystal laser cavity. *J. App. Phys.*, 103(3):033102, 2008.

[217] C.-Y. Kao and F. Santosa. Maximization of the quality factor of an optical resonator. *Wave Motion*, 45(4):412–427, 2008.

[218] S.-L. Chua, C. A. Caccamise, D. J. Phillips, J. D. Joannopoulos, M. Soljačić, H. O. Everitt, and J. Bravo-Abad. Spatio-temporal theory of lasing action in optically-pumped rotationally excited molecular gases. *Opt. Express*, 19(8):7513–7529, 2011.

[219] A. Cerjan, Y. D. Chong, and A. D. Stone. Steady-state ab initio laser theory for complex gain media. arXiv 1406.6659, 2014.

[220] H. Cao and J. Wiersig. Dielectric microcavities: Model systems for wave chaos and non-hermitian physics. *Rev. Mod. Phys.*, 87(1), 2015.

[221] Q. J. Wang, C. Yan, N. Yu, ulia Unterhinninghofen, J. Wiersig, C. Pflugl, L. Diehl, T. Edamurac, M. Yamanishic, H. Kanc, and F. Capassoa. Whispering-gallery mode resonators for highly unidirectional laser action. *Proc. Natl. Acad. Sci. USA*, 107(52):22407–22412, 2005.

[222] H. Risken and K. Nummedal. Instability of off resonance modes in lasers. *Phys. Lett.*, 26A(7), 1968.

[223] H. Zeghlache, P. Mandel, N. B. Abraham, L. M. Hoffer, G. L. Lippi, and T. Mello. Bidirectional ring laser: Stability analysis and time-dependent solutions. *Phys. Rev. A*, 37(2), 1987.

[224] L. A. Lugiato, L. M. Narducci, and M. F. Squicciarini. Exact linear stability analysis of the plane-wave maxwell–bloch equations for a ring laser. *Phys. Rev. A*, 34(4), 1986.

[225] L. A. Lugiato, F. Prati, L. M. Narducci, and G. L. Oppo. Spontaneous breaking of the cylindrical symmetry in lasers. *Opt. Commun.*, 69(5), 1988.

[226] C. Tamm. Frequency locking of two transverse optical modes of a laser. *Phys. Rev. A*, 38(11), 1988.

[227] I. Braun, G. Ihlein, F. Laeri, J. U. Nöckel, G. Schulz-Ekloff, F. Schüth, U. Vietze, Ö. Weiss, and D. Wöhrle. Hexagonal microlasers based on organic dyes in nanoporous crystals. *Appl. Phys. B*, 70(1):335, 2000.

[228] S. V. Boriskina. Symmetry, degeneracy and optical confinement of modes in coupled microdisk resonators and photonic crystal cavities. arXiv 1504.08018, 2005.

[229] S. Burkhardt, M. Liertzer, D. O. Krimer, and S. Rotter. Steady-state ab initio laser theory for fully or nearly degenerate cavity modes. *Phys. Rev. A*, 92(013847), 2015.

[230] D. Vogan. Private communication, 2015.

[231] A. Farjadpour, D. Roundy, A. Rodriguez, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson. Improving accuracy by subpixel smoothing in the finite-difference time domain. *Opt. Lett.*, 31(20):2972–2974, 2006.

[232] A. Raman and S. Fan. Perturbation theory for plasmonic modulation and sensing. *Phys. Rev. B*, 83(205131), 2011.

[233] M. M. Soljačić, E. Lidorikis, L. V. Hau, and J. D. Joannopoulos. Enhancement of microcavity lifetimes using highly dispersive materials. *Phys. Rev. E*, 71(026602), 2005.

[234] A. Pick, A. Cerjan, D. Liu, A. W. Rodriguez, A. D. Stone, Y. D. Chong, and S. G. Johnson. Ab initio multimode linewidth theory for arbitrary inhomogeneous laser cavities. *Phys. Rev. A*, 91(063806), 2015.

[235] A. Cerjan and A. D. Stone. Steady-state ab initio theory of lasers with injected signals. *Phys. Rev. A*, 90(013840), 2014.

[236] A. Cerjan, A. Pick, Y. D. Chong, S. G. Johnson, and A. D. Stone. Quantitative test of general theories of the intrinsic laser linewidth. arXiv 1505.01884, 2015.

[237] K. Iga. Surface-emitting laser–its birth and generation of new optoelectronics field. *IEEE J. Sel. Topics Quantum Electron.*, 6(6):1201–1215, 2000.