Inverse design of near-field thermal radiation and nonlinear optics

Weiliang Jin

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Electrical Engineering
Advisor: Alejandro W. Rodriguez

September 2019
Abstract

Traditional photonic design relies on intuitive and established principles that perform extremely well when restricted to narrow-bandwidth applications. In this thesis, we show that recently developed inverse-design techniques can be tailored to systematically discover complex structures for multifrequency or multi-functional devices, including near-field radiative heat transfer (RHT) and nonlinear frequency conversion. We begin by drawing attention to far-field thermal radiation and spontaneous emission from wavelength-scale heterogeneous objects, captured by a fluctuating-volume current formulation of electromagnetic scattering, illustrating that even simple geometric designs can be used to tailor the emissive and scattering properties of compact bodies. Such tunability can be important for RHT at subwavelength separations, which as we show next, can transfer heat in competition with thermal conduction in nanostructures, resulting in modified temperature gradients in plate-plate geometries and more significantly gratings.

Next, we demonstrate that optimization techniques can be exploited to systematically tune and enhance near-field RHT at desirable (selective) frequencies. We begin by looking into multilayer structures with arbitrary dielectric profiles along the gap directions, demonstrating that appropriately optimized, multilayer structures can lead to larger RHT compared to the best possible homogeneous thin films. We then investigate RHT between two-dimensional gratings having an enlarged two-dimensional design space. We demonstrate that RHT between lossy metals at infrared wavelengths such as tungsten, when appropriately structured, can approach that of a pair of ideal metal plates supporting surface plasmonic resonances. Moreover, we provide evidence of a previously predicted material scaling law in those extended structures, up to a threshold enhancement ratio.

Lastly, we discuss the application of inverse design techniques to the discovery of three-dimensional metasurfaces and wavelength-scale cavities that exhibit highly
efficient nonlinear frequency-conversion. Large nonlinear enhancements are achieved
by the creation of modes with large nonlinear overlap factors. Furthermore, as effi-
cient nonlinear processes require simultaneous control of light coupling to and from
a waveguide at widely separate frequencies, we apply inverse design techniques to
demonstrate wavelength-scale waveguide–cavity couplers operating over tunable and
broad frequency bands in both two and three dimensions.
Acknowledgements

Being a graduate student in Princeton for the past five years has had a far more profound impact on me than I originally expected, thanks to the guidance and help from many talented and considerate people in this remote land. First and foremost, I am deeply thankful to my PhD advisor, Prof. Alejandro Rodriguez, for his immeasurable input and support on my research and all other aspects essential to academics. He is always available to discuss any issues happily, and has put an immense effort on teaching me the art of writing, oral presentations, and scientific exploration, including discovery of important research problems and systematic ways of tackling them. For instance, even before informal oral talks, he always spent lots of time on multiple rehearsals to improve my presentation. His passion, intuition, and insight into physics and numerical methods can hardly be fathomed, and can be readily appreciated from his ultra-fast and precise language. I will always cherish his constant encouragement, radiant energies (possibly violating some sum rule), and constructive advice.

Secondly, I would like to thank many wonderful colleagues for stimulating discussions and feedback. Chinmay Khandekar is always modest, joyful and full of new ideas in the areas of photonics and philosophy. Our many conversations over the past four years are a wonderful memory. Next, I am very grateful to Zin Lin for patiently teaching me his incredibly long inverse design code, and for his quick response to my questions. In addition to being a great researcher, capable of speaking almost as fast as Prof. Rodriguez, he is always cheerful, inspiring confidence in my research life. Prashanth Venkataram is knowledgeable in all fields, capable of memorizing long equations, and is also talented at conversations. He can always explain challenging physics using simple and plain language, which has inspired me greatly. Following up Chinmay’s thesis acknowledgment, we are still working hard to get Prashanth involved in our tea time, possibly occurring at the conjugate of imaginary frequency or relying on future breakthroughs in transformation optics. I thank Sean Molesky for
his invaluable help with my writing and oral talks, and for our close collaborations on many projects. Lastly, I am very fortunate to have Pengning Chao to point out bugs in my code.

Moreover, I am thankful to our collaborators Dr. Riccardo Messina, Prof. Athanassios Polimeridis, Dr. Adi Pick, Prof. Kai-Mei Fu, and Alan Logan. I greatly appreciate the EE department staff, including Colleen and Roelie, for smoothly processing everything. I am also thankful for the support of Princeton research computing, which graciously allowed us to boost priority multiple times. Beyond research, I always enjoyed outstanding concerts in Richardson Auditorium, nice plays in McCarter Theatre, the beautiful and peaceful campus, and many hiking sites that resembled green paradises during spring and summer.

Beyond academics, I thank the support of parents and all gurus. I am grateful to my language tutor M. Weiner, piano teacher M. Kogan, and my host family, M. Wu. I am also grateful to all of my friends, including H. X. Yang for her remote company during the first three years of my PhD, Y. M. Guo for her constant support and understanding, Y. S. Liu for our weekly shopping trips over the past five years, M. L. Liu and Y. J. Li for our wonderful trips, Y. X. Guo for her long-term friendship, and X. L. Dai for asking identical questions every time. Lastly, I thank all of the plants in my room for steadily growing and flourishing, inspiring me to become a patient and productive researcher.
To everyone in pursuit of wisdom.
Contents

Abstract ................................................................. iii
Acknowledgements ..................................................... v
List of Tables .......................................................... xi
List of Figures .......................................................... xii

1 Introduction ......................................................... 1
  1.1 Overview of inverse design ................................. 7
    1.1.1 Topology optimization .................................. 8
    1.1.2 Shape optimization ...................................... 9
  1.2 Overview of near-field radiative heat transfer ........... 10

2 Thermal radiation and spontaneous emission from heterogeneous objects ................................................. 14
  2.1 Fluctuating-volume current formulation ................. 15
  2.2 Temperature control of thermal radiation in composite bodies .............................................. 17
    2.2.1 Overview .................................................. 18
    2.2.2 Directional thermal emission .......................... 19
    2.2.3 Concluding remarks ..................................... 25
  2.3 Spontaneous emission from gain–composite structures ......................................................... 26
    2.3.1 Overview .................................................. 26
    2.3.2 Active medium: 4-level atoms ........................ 29
2.3.3 Self-consistent treatment of Purcell effect  . . . . . . . . . . . . 31
2.3.4 Radiation patterns from heterogeneous particles . . . . . . . . 34
2.3.5 Gain inhomogeneity arising from pumping . . . . . . . . . . . . 37
2.3.6 Gain inhomogeneity arising from Purcell effect . . . . . . . . . 39
2.3.7 Concluding Remarks . . . . . . . . . . . . . . . . . . . . . . . 43

3 Strongly coupled conduction and radiation at the nanoscale 45
3.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
3.1.1 General formulation . . . . . . . . . . . . . . . . . . . . . . . . 47
3.2 Planar objects . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48
3.2.1 Surface-sink approximation . . . . . . . . . . . . . . . . . . . . 50
3.2.2 Numerical predictions . . . . . . . . . . . . . . . . . . . . . . . 52
3.2.3 Concluding Remarks . . . . . . . . . . . . . . . . . . . . . . . . 56
3.3 Generic shapes: grating structures . . . . . . . . . . . . . . . . . . . 56
3.3.1 Extending fluctuating–volume current formulation . . . . . . . 57
3.3.2 Coupled conduction and radiation in nanorods . . . . . . . . . . 60
3.3.3 Concluding Remarks . . . . . . . . . . . . . . . . . . . . . . . . 65

4 Inverse design of radiative heat transfer 66
4.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.2 Inverse design in multilayer structures . . . . . . . . . . . . . . . . . 70
4.2.1 Multilayer formulation . . . . . . . . . . . . . . . . . . . . . . . 70
4.2.2 Optimizing heat transfer between plate–plate . . . . . . . . . . 73
4.2.3 Optimizing heat transfer between dipole–plate . . . . . . . . . . 77
4.2.4 Concluding Remarks . . . . . . . . . . . . . . . . . . . . . . . . 84
4.3 Inverse design in grating structures . . . . . . . . . . . . . . . . . . . 84
4.3.1 Low-rank formulation . . . . . . . . . . . . . . . . . . . . . . . 86
4.3.2 Lossy metal grating optimization . . . . . . . . . . . . . . . . . 88
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.3 Concluding Remarks</td>
<td>91</td>
</tr>
<tr>
<td>5 Nonlinear frequency conversion and multimode couplers</td>
<td>92</td>
</tr>
<tr>
<td>5.1 Optimizing nonlinear processes in three-dimensional structures</td>
<td>93</td>
</tr>
<tr>
<td>5.1.1 Metasurface</td>
<td>95</td>
</tr>
<tr>
<td>5.1.2 Microcavity</td>
<td>99</td>
</tr>
<tr>
<td>5.1.3 Concluding Remarks</td>
<td>101</td>
</tr>
<tr>
<td>5.2 Compact multimode cavity couplers</td>
<td>102</td>
</tr>
<tr>
<td>5.2.1 Overview</td>
<td>102</td>
</tr>
<tr>
<td>5.2.2 Formulation</td>
<td>103</td>
</tr>
<tr>
<td>5.2.3 Coupler in two dimensions</td>
<td>107</td>
</tr>
<tr>
<td>5.2.4 Coupler in three dimensions</td>
<td>111</td>
</tr>
<tr>
<td>5.2.5 Concluding Remarks</td>
<td>113</td>
</tr>
<tr>
<td>6 Conclusion and Outlook</td>
<td>115</td>
</tr>
<tr>
<td>A Material properties of GST alloy</td>
<td>118</td>
</tr>
<tr>
<td>A.1 Material parameters</td>
<td>118</td>
</tr>
<tr>
<td>A.2 Temperature gradients</td>
<td>119</td>
</tr>
<tr>
<td>A.3 Geometries</td>
<td>121</td>
</tr>
<tr>
<td>B Randomized singular value decomposition for near-field radiative heat transfer</td>
<td>123</td>
</tr>
<tr>
<td>B.1 Approximate Singular Value Decomposition</td>
<td>123</td>
</tr>
<tr>
<td>B.2 Fast Gradient Adjoints</td>
<td>125</td>
</tr>
<tr>
<td>Bibliography</td>
<td>127</td>
</tr>
</tbody>
</table>
List of Tables

5.1  Representative second-harmonic generation figures of merit for both hand- and inverse-designed metasurfaces, including \(\chi^{(2)}\), fundamental wavelength \(\lambda_1\), and conversion efficiency \(\eta\) per unit cell. Most of the literature only states the net conversion efficiency for metasurfaces with finite \(N\) unit cells, which we convert to the efficiency \(\eta\) per unit cell. 98

A.1  The thermal properties used in our COMSOL simulation . . . . . . . 121
List of Figures

2.1 Schematic of two composite bodies, involving GST (blue) hemi-
spheroids coated with Ti (green), AZO (red), or Si₃N₄ (orange) shells,
which rest on a low-index, transparent substrate in contact with a heat
reservoir at 300 K. The GST is heated from below by a conductive 2D
material (e.g. a carbon-nanotube wall or graphene sheet), leading to
temperature gradients within the structure, with temperature profiles
illustrated in the lower insets. The presence of boundary resistance at
material interfaces is captured by effective thermal resistances (see text).
2.2 Spectral emissivity $\epsilon(\omega) = \Phi(\omega)/\Phi_{BB}(\omega)$ (blue dots) from heterogeneous bodies comprising Ge$_2$Sb$_2$Te$_5$ (GST) hemispheroids coated with either (a) Si$_3$N$_4$ or (b) Ti shells and resting on a low-index substrate (object dimensions are specified in Fig. 2.1). The structures are heated from the GST–substrate interface by a 2D thin-film conductor under heating scenario (iii), described in the text, with an interface temperature $T_{\text{GST}} = 870$ K. The resulting temperature profiles $T(\vec{x})$ are shown in Fig. 2.1 and on the insets. $\epsilon$ is defined as the ratio of the thermal flux $\Phi(\omega)$ of each body normalized to the flux $\Phi_{BB} = \frac{A}{4\pi}(\omega/c)^2\Theta(\omega,T)$ from a corresponding black body of the same surface area $A$ but uniform $T = 870$ K (green lines, arb. units). Also shown are the partial directivities $\eta_\pm = \Phi_\pm/\Phi$ (red line), defined as the ratios of the outgoing flux into the upper/lower hemisphere $\Phi_\pm = 2\pi\int_{\mp\pi/2}^{\mp\pi/2} d\theta \Phi(\omega,\theta)$ to the total flux, where $\theta$ is defined in Fig. 2.1. (c) Angular radiation intensity $\Phi(\theta)$ normalized by the total flux $\Phi$, for the structures in (a) (solid red line) and (b) (solid blue line) under scenario (iii), compared to emission under uniform-temperature conditions (dashed lines).

2.3 Total (frequency integrated) partial directivity $P_- = H_-/H$ as a function of $T_{\text{GST}}$, with $H_- = \int d\omega \Phi_-\omega$, for the Ti structure under different heating conditions, corresponding to multiple degrees of temperature localization in the GST (see text).
2.4 (a,b) Peak directivity $\eta_-$ (dashed lines, right axes) and emissivity $\epsilon$ (solid lines, left axes) associated with the resonance closest to the peak of the BB spectrum of the Ti-shell hemispheroid of Fig. 2.1 as a function of various shape/size parameters, including aspect ratio $2h/D$ (magenta), overall size $D$ (black), metal thickness $t$ with a fixed GST size $L = L_0$ (red), and GST size $L$ with a fixed overall size $D = D_0$ (blue). Here, the subscripts “0” and the vertical dashed lines denote size parameters specified in Fig. 2.1. (c) Angular radiation intensity $\Phi(\theta)$ from various heterogeneous bodies—composite GST (blue), AZO (red), and Si$_3$N$_4$ (orange) mushrooms and cylinders—at selected frequencies $\omega$ and normalized by the total flux $\Phi$. All plots assume perfect $T = 870$ K localization in the GST while other materials are held at 300 K.

2.5 Schematic of a gain–loss composite body consisting of a dense (active) collection of 4-level atoms (left inset) optically pumped via monochromatic light at frequency $\omega_{30}$ and pump rate $P$, resulting in population inversion and spontaneous emission (SE) at $\omega_{21}$. The nonlinear feedback of Purcell effect on the SE rate of the object is captured via a system of nonlinear equations illustrated on the right schematic, in which: (a) the dielectric constant at the location of the $n$th emitter depends on the corresponding atomic decay rate $\gamma_{21,n}$ through Eq. 2.6 (b) the radiative emission $\Phi_n$ of a classical dipole at the location of the $n$th emitter, described by Eq. 2.5, is altered, e.g. enhanced or suppressed, by the surrounding $\epsilon(x)$; and (c) the radiative decay rate $\gamma_{21,n}^r$ of the emitter is dressed by the Purcell effect through Eq. 2.8 which in turn affects the dielectric response in (1).
2.6 (a) SE flux $\Phi$ from $\mathcal{PT}$–symmetric spheres consisting of $N = \{1, 2, 3, 4\}$ (black, green, blue, and red lines) regions of equal gain (orange) and loss (gray), with permittivities $\epsilon = 4 \pm i$ and gain temperature $T \to 0^\circ K$ (corresponding to complete population inversion), as a function of sphere radius $R$ (for a fixed vacuum wavelength $2\pi c/\omega$).

$\Phi(\omega)$ is normalized by the flux $|\Phi_{BB}(\omega)| = R^2(\omega/c)^2\hbar\omega$ from a “blackbody” of the same surface area and temperature. Insets show angular radiation intensities $\Phi(\theta, \phi)$ for different $N$ at selective radii $R = \{2.6, 3.2, 2.7, 2.1\}(c/\omega)$ (corresponding to increasing $N$), with orange/gray colormaps denoting regions of gain/loss. (b) Selected $\Phi(\theta, \phi)$ of various $\mathcal{PT}$–symmetric shapes, including “beach balls”, “magic” cubes, and cylinders, showing complex directivity patterns that depend strongly on the centro-symmetry of the objects. Here the color white (black) refers to the maximum (minimum) flux value. (c) Peak SE flux $\Phi$ (right axis, dashed lines) and directivity $\Phi_G/\Phi$ (left axis, solid lines), or the ratio of the flux emitted along gain direction $\Phi_G$ (see text) to the total flux $\Phi$, near the third resonance of the $N = 1$ sphere, as a function of the gain/loss tangent $\delta = |\text{Im}\,\epsilon|/\text{Re}\,\epsilon$ and for multiple values of Re $\epsilon = \{2, 4, 12\}$ (red, black, and blue lines), where the vertical dashed black line denotes $\delta$ for (a). The plots show increased directivity attained as the system approaches the lasing threshold (marked in the case of Re $\epsilon = 12$, where $\Phi$ diverges).
2.7 (a) SE flux $\Phi$ and (b) gain directivity $\Phi_G/\Phi$, or the ratio of the flux emitted along the gain directions $\Phi_G$ (see text) to the total flux $\Phi$, from the $N = 2$ sphere (inset) of Fig. 2.6 at frequency $\omega_{21}$, corresponding to the transition frequency of an active region consisting of Rhodamine 800 dye molecules. The gain medium is excited by planewaves propagating in opposite directions, $x \cos \phi + y \sin \phi$ and $-x \cos \phi - y \sin \phi$, for three different orientations, $\phi = 0$ (black line), $\pi/4$ (blue line), and $-\pi/4$ (red line), leading to significant spatial inhomogeneities. $\Phi$ is normalized by $\Phi_{BB}$ as in Fig. 2.6 and plotted as a function of radius $R$, in units of the dye $2 \rightarrow 1$ transition wavelength $c/\omega_{21}$ (see text). Insets in (a) show $z = 0$ cross-sections of the resulting gain profiles $-\Im \epsilon_g$ at $R = 3.4(c/\omega_{21})$ while those in (b) show angular radiation intensities $\Phi(\theta, \phi)$ normalized by the maximum intensity $\Phi_{max}$ at $R = 4(c/\omega_{21})$ for the different pump orientations.
2.8 Self-consistent treatment of Purcell effect (PE) based on solution of Eq. 2.6 and Eq. 2.8 using fixed-point iteration. (a) SE flux $\Phi$ normalized by $\Phi_{BB}$ as in Fig. 2.6 and (b) gain directivity $\Phi_G/\Phi$ from the $N = 1$ sphere of Fig. 2.6 as a function of radius $R$ (in units of $c/\omega_{21}$), with an active medium consisting of Er$^{3+}$ atoms and subject to a uniform pump rate $P/\gamma_0 = 10^{-4}$, where $\omega_{21}$ and $\gamma_0$ denote the bulk (bare) $2 \rightarrow 1$ transition frequency and decay rate of the atoms, respectively. Top/bottom insets in (a) show $z = 0$ cross-sections (contour plots) of the gain profiles $-\text{Im}\epsilon_g$ obtained during the first and final (fixed-point) iteration of the algorithm at two separate radii $R = \{2, 2.5\}(c/\omega_{21})$, with black/white denoting $\min_{\text{max}} -\text{Im}\epsilon_g \approx \{1.9, 1.3\}$ and $\{1.9, 1.5\}$, respectively. Insets in (b) show the angular radiation intensities $\Phi(\theta)$ at selected radii (black dots), normalized by the maximum intensity $\Phi_{\max}$. (c) $\Phi$ at a fixed radius $R = 2(c/\omega_{21})$ as a function of $P/\gamma_0$, with insets illustrating the evolution of $\Phi$ as a function of iteration at a fixed $P/\gamma_0 = 10^{-4}$ and for two different values of quantum yields (QY) [see text]. All plots compare quantities in the presence (solid blue lines) or absence (dashed black lines) of PE, and under a uniform-medium approximation (UMA) (solid red lines) described in the text; the solid green line in (a) denotes the average Purcell factor $\langle F \rangle = \frac{1}{V} \int V F$ at each radius.
2.9 Self-consistent treatment of Purcell effect (PE) in the metal–dielectric
composite shown on the inset, with orange and blue denoting Er³⁺–
doped polymer and Au regions, described via full solution of Eq. 2.6
and Eq. 2.8 (solid blue lines) or via the uniform-medium approximation
(dashed red line) described in the text. The solid line shows the SE flux
Φ from the particle as a function of radius R (in units of \(c/\omega_2\)). Insets
show \(z = 0\) cross-sections of the gain profile \(-\text{Im} \epsilon_g\) (top) and active
temperature (right), along with the angular radiation intensity \(\Phi(\theta, \phi)\)
(bottom) at a selected \(R = 0.95(c/\omega_2)\), normalized by the maximum
intensity \(\Phi_{\text{max}}\).

3.1 Schematic illustration of two gratings of separated by a vacuum gap \(d\),
whose temperature distribution and energy exchange is mediated by
both conductive \(\nabla \cdot [\kappa(x)T(x)]\) and radiative \(H(x, x')\) heat transfer.

3.2 Schematic of two parallel slabs separated by a distance \(d\) along the
z direction (\(z = 0\) denotes the middle of the gap). The two slabs
exhibit a temperature profile \(T(z)\): the temperature is constant and
has values \(T_L\) and \(T_R\) in the regions \(z \leq -d/2 - t_a\) and \(z \geq d/2 + t_b\),
respectively, and variable in the regions \(-d/2 - t_a < z < -d/2\) and
\(d/2 < z < d/2 + t_b\), with \(T_a\) and \(T_b\) denoting the temperatures at the
slab–vacuum interfaces.
3.3 Temperature profile along the temperature-varying region of the hot slab in a configuration shown in Fig. 3.2, where the two silica slabs with \( t_a = t_b = 100 \mu m \) are held at external temperatures \( T_L = 600 \text{ K} \) and \( T_R = 300 \text{ K} \) at the outer ends. The points \( z = 0, 100 \mu m \) represent the boundary of the thermostat and the interface with the vacuum gap. The three lines correspond to \( d = 10 \text{ nm} \) (black), \( 20 \text{ nm} \) (red) and \( 50 \text{ nm} \) (blue). The inset shows the \( z \)-dependent radiative flux-rate \( \varphi(z_a) \) (see text) along slab \( a \) at \( d = 100 \text{ nm} \).

3.4 Total flux \( \varphi \) and temperature difference \( T_a - T_b \) (inset) as a function of distance \( d \) between two silica slabs (shown schematically in Fig. 3.2) that are being held at \( T_L = 600 \text{ K} \) and \( T_R = 300 \text{ K} \). The various solid lines correspond to different temperature-varying regions \( t \) (from top to bottom): \( 100 \text{ nm} \) (black), \( 1 \mu m \) (red), \( 10 \mu m \) (brown), \( 100 \mu m \) (blue) and \( 500 \mu m \) (green). The orange dashed line shows \( \varphi \) in the absence of temperature gradients.

3.5 Typical distance scale \( \tilde{d} \) relevant to conduction–radiation problems (see text) as a function of the material-dependent ratio \( h_0/\kappa \) for two slabs with \( t_a = t_b = 100 \mu m \). Solid circles denote corresponding values for specific material choices (abbreviations), with AZO[1.2] and AZO[0.05] denoting aluminum zinc oxides of different conductivities \( \kappa = 1.2 \text{ W/m·K} \) [284] and potential \( \kappa = 0.05 \text{ W/m·K} \) [92], respectively. The inset shows the dependence of the radiative-heat transfer coefficient \( h_0 \) on the external temperature gradient \( \Delta T = T_L - 300 \text{ K} \), for three cases, AZO (red), silica (blue) and SiC (black), with decreasing values.
3.6 (a) Total radiative heat transfer spectrum $\Phi(\omega)$ between two AZO nanorods (solid lines) of thickness $t = 500$ nm and cross-sectional area $A = l^2$, separated by $d = 20$ nm and held at temperatures $T_{a(b)} = 800(300)$ K, schematically illustrated in upper (b). The spectrum is shown for different cross-sections $l = \{10, 20\}$ nm (blue and red lines) and in the limit $l = \infty$, corresponding to two planar slabs. Lower (b) shows spatial radiative heat flux emitted from nanorod $a$ for the case $l = 20$ nm, corresponding to the (i) first, (ii) second, and (iii) SPP plasmon resonances, respectively, annotated in (a).

3.7 (a) Temperature profile along the $z$ coordinate of a nanorod (solid lines) when it is heated from one side to a temperature of 800 K, and is separated from an identical, constant- and uniform-temperature nanorod held at $T = 300$ K on the other side, by a gap size $d = 20$ nm. The nanorods have cross-sectional width $l = 20$ nm and thicknesses $t = 500$ nm, and are made up of AZO with results shown for multiple values of the doping concentration $\{2, 6, 11\}$ wt% (blue, red, and black lines). Also shown are the temperature profiles of slabs (dashed lines) of the same thickness (corresponding to the limit $l \to \infty$). (Inset:) Temperature distribution throughout the nanorod in the case of 11 wt%. (b) Temperature profiles of nanorods of width $l = 10$ nm under various separations $d = \{5, 10, 20, 30\}$ nm (black, blue, red, and green lines). (c) The flux value of radiative heat transfer for nanorods (red) and slabs (blue) in the presence (solid lines) or absence (dashed lines) of the temperature gradients induced by the interplay of conduction and radiation. (Inset:) The temperature $T_i$ at the interface $I$ of nanorod $a$ as a function of heat pumping rate at $I$, under various separations $d = \{5, 10, 20, 30\}$ nm (black, blue, red, and green lines).
4.1 (a) Schematic of two inhomogeneous $\varepsilon(z)$ slabs (A and B) separated by a vacuum gap of size $d$ along the $z$ direction. The radiative heat transfer (RHT) rate between them depends on their local temperatures $T_{A,B}$ and reflection coefficients $R_{A,B}$. Associated with each slab is a coordinate system centered at the slab–vacuum interface and pointing away from the gap. (b) Enhancement factor comparing RHT between optimized inhomogeneous (solid lines) or optimized uniform (dashed lines) slabs against that of semi-infinite uniform plates of $\text{Re}[\varepsilon] = -1$, at a fixed vacuum wavelength $\lambda = 8 \mu m$, as a function of material loss $\text{Im}[\varepsilon]$, and for two representative separations $d = 10$ nm (red) and 500 nm (blue). The black dotted line shows the theoretical bound described by Eq. 4.4. (c) Transmission coefficient $Z$ corresponding to either inhomogeneous (red), uniform (blue), or semi-infinite (black) slabs, as a function of the dimensionless wavevector $k_\beta d$ at fixed $d = 10$ nm and $\text{Im}[\varepsilon] = 10^{-2}$. The inset shows a typical dielectric profile $\text{Re}[\varepsilon(z)]$ needed to achieve $Z \approx 1$ over a broad range of $k_\beta$. To enhance the readability, two different $x$-scales are used in the ranges [0, 20] nm and [20, 100] nm, and the points resulting from numerical optimization are connected with segments.
4.2 (a) Schematic of an inhomogeneous \( \varepsilon(z) \) slab and a dipole separated by a vacuum gap of size \( d \) along the \( z \) direction. (b) Dielectric profile \( \text{Re}[\varepsilon(z)] \) corresponding to inhomogeneous slabs optimized to increase RHT from a dipole a distance \( d = 1 \, \mu m \) away from their \( z = 0 \) interface, at a fixed vacuum wavelength \( \lambda = 8 \, \mu m \) (frequency \( \omega \approx 0.785c/d \)). The profiles are obtained under different constraints on the maximum possible permittivity \( \varepsilon_{\text{max}} \equiv \max |\text{Re}\varepsilon| = \{5, 40\} \) (upper and lower figures) but correspond to the same uniform \( \text{Im}[\varepsilon] = 10^{-3} \). The imaginary part of the reflection coefficient \( \text{Im}[\mathcal{R}(k_\beta)] \) and LDOS \( \mathcal{L}(k_\beta) \) (in SI units) at the location of the dipole and at \( \omega \) are plotted in (b) and (c) as a function of \( k_\beta \), along with those of optimized uniform slabs (blue solid line). The black dotted line shows the theoretical bound described by Eq. 4.9. The inset in (c) shows the \( k_\beta \)-integrated spectrum \( \mathcal{L}(\omega') \) near \( \omega \) as a function of the dimensionless frequency \( (\omega' - \omega)/\omega \text{Im}[\varepsilon] \), indicating that contributions from smaller \( k_\beta \) are increasingly sensitive to the wavelength.

4.3 (a) Peak LDOS of optimized inhomogeneous slabs (solid lines) and optimized uniform slabs (dashed lines), as a function of the dimensionless separation \( \omega d/c \), for multiple material loss rates \( \text{Im}[\varepsilon] = \{10^{-1}, 10^{-2}, 10^{-3}\} \) (black, blue, and red, respectively). The green dotted line marks the largest possible LDOS in the far field, given by Eq. 4.11. (b) Enhancement factor comparing the peak LDOS of optimized inhomogeneous and uniform slabs.
4.4 Near-field RHT enhancement for inverse-designed tungsten gratings $F$, along with that of planar silicon carbide for comparison, relative to the enhancement achieved with ideal (lossless) planar metals, $F_{pl}^{2D}$ given in Eq. 4.16, with respect to frequency $\omega$ (left). Profiles of the structures are displayed as insets. Color plots depicting the $k$-dependent transfer functions $T(\omega, k)$ over the chosen frequency range are shown on the right. The depicted optimization proceeds from an unstructured planar system (0) to structured gratings (N) by successively introducing additional ellipsoidal degrees of freedom “N” to the design space. Performance is qualified by the magnitude of RHT at a single design frequency $\omega_{opt} = 2\pi c/3 \mu m$ ($\lambda_{opt} = 3 \mu m$), where tungsten behaves as a highly lossy metal far from the planar surface plasmon resonance. The gap separation and the period of gratings are $d(\Lambda) = 0.02 (0.04) \lambda_{opt}$, respectively.
4.5 The figure highlights the material trend in enhancement $F/F_{pl}^{2D}$ found by optimizing grating geometries over a wide range of material susceptibilities $\chi$. Quantities displayed in parenthesis correspond to the relative enhancement of a given grating design compared to a planar geometry of the same material. Various shape parameterizations such as polylines, Fourier curves, and ellipsoids marked as squares, triangles, and circles are considered. The susceptibility values are chosen to vary along either the real or imaginary axis: gratings $\{1, 2, 4\}$ correspond to $\chi = \{-76, -101, -151\} + 50i$, 5 to $\chi = -151 + 40i$, $\{3, 7\}$ to $\chi = \{-121, 169.5\} + 37.3i$, and $\{6 \rightarrow 9\}$ to $\chi = -169.5 + 50i, 37.3, 30, 25i$.

Results for tungsten are colored red. The upper inset illustrates the unit cell of correspondingly numbered grating. (There is no visible difference for gratings 6-9 and so only one of these is shown.) The lower inset depicts the RHT of tungsten grating 7 when the susceptibility is varied without altering the structure. The dashed line represents the susceptibility scaling predicted by recently derived limits on RHT [313].

5.1 (a) Schematic illustration of SHG in a square-lattice metasurface of finite thickness $t$ and period $\Lambda \times \Lambda$. Shown to the right are dielectric profiles and mode profiles $|E|^2$ corresponding to two inverse-designed metasurfaces, both over single unit cells and ($z = 0$) cross sections. The structures are optimized to ensure frequency- and phase-matching for light incident at an angle $\theta = 3^o$ (i) or normal incidence (ii). Dark (white) represents gallium phosphide (vacuum) regions. (b) Convergence of the objective function with respect to iteration number, leading to structure (ii).
5.2 (a) 3D view of an optimized microcavity on top of a silica substrate (gray). The cavity has a finite thickness $t = 250$ nm, and equal length along the other two dimension $L_x = L_y = 4.65 \mu m$. Shown in (b) is its dielectric profile over the $x - y$ cross section, where black (white) represents gallium phosphide (vacuum) regions. The mode profiles at both fundamental ($\omega_1$) and second harmonic ($\omega_s$) frequencies are shown in (c), in which the left depicts the E-field along the $x - y$ cross section, and the right the $y - z$ cross-section. Here we only show the $E_x (E_z)$ component at $\omega_1(s)$. 

5.3 Schematic of a general coupler (black region), which couples light between a wavelength-scale, multimode cavity and a multimode port (waveguide). The design freedom of the scatter enables controllable coupling between the two devices at several wavelengths (red, yellow, and green arrows).

5.4 Optimized coupler for SHG showing critical coupling between the double-ring resonators and the waveguide. All structures (black) are made of GaP, while the substrate (white) is assumed to be vacuum. The width of the waveguide is 150 nm, the diameter of the outer ring 2.6 $\mu m$, and the area of the designed coupling region 3.75 $\mu m \times 1.5 \mu m$. (c) shows the energy spectrum inside the resonator near the fundamental and second-harmonic wavelengths $\lambda_{1,s} = \{1500, 750\}$ nm, with matched azimuthal wavenumbers $m_1 = 8$ and $m_s = 2m_1$, (black, red) normalized by $U^0$. The middle figures show the TM-polarized electric fields at the respective wavelengths. The complete suppression of outgoing/transmitted power through the waveguide provides a visual confirmation of critical coupling.
5.5 Optimized coupler for SFG showing near critical coupling between the triple-ring resonators and the waveguide. The middle figures show the TM-polarized electric fields at the wavelengths \( \lambda_{(1,2,s)} = \{1500, 907, 565\} \) nm, \( m_{(1,2,s)} = \{9, 20, 28\} \). The width of the waveguide is 134 nm, the diameter of the outer ring 2.8 \( \mu\text{m} \), and the area of the coupling region is 5.4 \( \mu\text{m} \times 2 \mu\text{m} \).

5.6 Optimized coupler for comb generation showing near critical coupling over 6 frequencies. The width of the waveguide is 300 nm, the diameter of the ring 1.8 \( \mu\text{m} \), and the area of the designed coupling region 4.5 \( \mu\text{m} \times 4.5 \mu\text{m} \). The figures show the TM-polarized electric field profiles at the respective azimuthal number \( m = \{5 \rightarrow 10\} \), corresponding to frequencies \( f = \{0.667 \rightarrow 1.157\} \) c/1.5 \( \mu\text{m} \), with equal spacing \( \Delta f = 0.098 \text{ c/1.5 \( \mu\text{m} \).} \)

5.7 (a) On top of a silica substrate (gray) is a microcavity optimized for both large nonlinear overlap factors and critical coupling to the same waveguide at both frequencies. The cavity has a finite thickness \( t = 250 \text{ nm in the z-direction, and } L_x = 4.65 \mu\text{m}, L_y = 6.2 \mu\text{m} \) along the other two dimensions. The waveguide has the same thickness \( t = 250 \text{ nm, and a small width } w = 465 \text{ nm.} \) Shown in (b) is its dielectric profile over the \( x-y \) cross section, where black (white) represents gallium phosphide (vacuum) regions. The mode profiles along the \( x-y \) cross section at both fundamental (\( \omega_1 \)) and second harmonic (\( \omega_s \)) frequencies are shown in (c), where we are plotting the \( E_x (E_z) \) component at \( \omega_1(s) \).

A.1 Real (solid red) and imaginary (dashed blue) permittivity \( \varepsilon_{\text{GST}}(T, \lambda) \) of a bulk GST glass at \( \lambda = 5.8 \mu\text{m} \), obtained via simple linear interpolation of experimental data at multiple temperatures (circles).
A.2 Temperature distribution $T(\vec{x})$ of the Ti (left) and Si$_3$N$_4$ (right) hemi-

spheroid composites described in Fig. 1. Both structures rest on a SiO$_2$

substrate (thickness 0.3$\mu$m and radius $= 1.5 \times$ shell radius) whose bot-
tom surface is in contact with a heat reservoir at 300 K. All other

surfaces are exposed to vacuum and therefore satisfy adiabatic bound-

ary conditions; material interfaces on the other hand are subject to

thermal boundary resistance in accordance with operating condition

(iii) and (ii) described in the text, for the left and right body, respec-
tively.

A.3 Parameter descriptions for the bodies associated with Fig. 2.4 of the

main text.
Chapter 1

Introduction

The physics of light at the nanoscale is essential to many naturally occurring and technological processes, from plants’ evolved protein complexes for efficient light harvesting in photosynthesis [384] to the development of modern computing [97, 281, 209]. Earlier approaches of manipulating light relied most heavily on the simple physics of ray optics and related approximations, bearing important inventions such as optical fibers [177] and antennas [23] to guide light over long distances. Recent advances in nanotechnology have enabled us to further explore the wave-nature of light encoded in Maxwell’s equations [397], via photonic crystals [209], Anderson localization in disordered media [392], topological photonic crystals [286], metamaterials [80], and many others, shedding light on and providing new applications for light–matter interactions.

In recent years, the manipulation of light in highly integrated photonics circuits has become particularly relevant, at times setting up bottlenecks to emerging breakthroughs. For instance, photonic nanostructures offer a promising route to quantum computing and communication [11, 386, 281], since photons are robust carriers of quantum entanglement over long distances. However, photons suffer from weak interactions [281], demanding a significant boost to light–matter interactions in order to achieve single-photon-level nonlinearities [56], as well as large photon collection and
emission efficiencies [49]. Another example are artificial neural networks, which have recently received significant interest and success throughout academia and industry. Since it heavily relies on fixed matrix-vector multiplications, optical systems, in which the scattering processes of light naturally embed the multiplications, are shown to offer a promising platform to speed up such calculations while lowering power consumption [398, 408, 262]. The design of “optical neurons”, which aim to handle complicated light scattering and interference in the smallest volumes, can be crucial to the realization of appropriate convolution operations. Additionally, photonic design is important to many other applications, including solar energy conversion [145, 463], thermal energy manipulation [463, 206, 207], topological photonics [286], and frequency demultiplexers [352].

Photonics design can have dramatically distinct principles depending on the types of applications. Historically, design principles relied mainly on intuition-based approaches based on known physical effects such as index guiding [177], band-gap engineering [209], and resonances [431]. They have enabled many remarkable ways of tailoring light over narrow bandwidths, such as temporal confinement of light in micron-scale volumes with quality factors approaching $10^9$ [249], and tight spatial confinement of light [189, 190, 93] to atomic domains [375]. However, the pursuit of smaller footprint structures capable of operating over many discrete or continuous bands of wavelengths presents a challenge: traditional design principles based on repetitive combinations of highly symmetric shapes usually lead to suboptimal structure and suffer from multiple trade-offs. To automatically explore a much larger design space beyond such known paradigms, inverse design methods were introduced and developed, beginning in the late 1990s. In particular, large-scale optimization approaches capable of exploring millions of design variables are just becoming tractable in photonics, and have already been applied to discover novel structures, including
wavelength-division multiplexers [352, 136], nonlinear frequency converters [265, 267], and metalenses [266, 264].

In this thesis, we show that inverse design methods, briefly introduced in Sec. 1.1, can be tailored to both improve existing optical device performance and to reveal new physics. Our work focuses on two challenging photonic problems: photonic designs for near-field radiative heat transfer (NFRHT) and nonlinear frequency conversion. The former, reviewed in Sec. 1.2, can result in heat flux rates that are orders-of-magnitude larger than those predicted by Stefan–Boltzmann’s law [27, 444, 213]; it is therefore a likely candidate for many energy related applications such as thermophotovoltaic (TPV) energy conversion [27] and thermal lithography [410]. While NFRHT is known to depend strongly on both material and geometric properties, there is not yet a systematic framework that fully exploits their impact for tunability, partly because of the large density of participating states and the strong scattering of evanescent waves present in the near-field [207]. Nonlinear frequency conversion, on the other hand, plays a critical role in spectroscopy [179], the generation of novel optical states [240, 440, 235], long-range fiber transmission [340], ultrashort-pulse shaping [106, 12], and the aforementioned problems in the area of quantum computing. While recent algorithm-driven approaches have been proposed and applied to obtain resonators with improved conversion efficiencies [265, 267], there is still much room for improvement.

This thesis is organized as follows. In the first two chapters, as a warm-up to the inverse design of NFRHT detailed in Chapter 4, we begin by drawing attention to our recently developed numerical technique for capturing thermal radiation in structures with strong temperature and dielectric inhomogeneities. Enabled by this numerical tool, we discuss new emergent regimes of NFRHT and systematic methods of enhancing NFRHT. More specifically, at the beginning of Chapter 2 we introduce the aforementioned numerical tool, denoted as the fluctuating-volume current (FVC)
formulation of electromagnetic scattering. We then exploit the FVC formulation to demonstrate that when selectively heated, wavelength-scale composite bodies—complicated arrangements of phase-change materials and metals or semiconductors—can exhibit large temperature gradients that exhibit designable, directional emission at infrared wavelengths. Specifically, we show that micron-scale chalcogenide (GST) hemispheroids coated with titanium or silicon-nitride shells and resting on low-index, transparent substrates, can exhibit large emissivity and $\gtrsim 80\%$ partial directivity. This effect is shown to be a consequence of a dual radiation/conduction design strategy that achieves both resonance enhancement and temperature localization within the GST, the interplay of which enables these composite infrared thermal antennas to not only enhance but also selectively emit/absorb light in specific directions.

In the second half of Chapter 2, we extend the FVC formulation to the problem of modeling spontaneous emission and scattering from composite, wavelength-scale structures, e.g. metal–dielectric spasers [34, 414–415], subject to inhomogeneities in both material and noise properties. We begin by studying amplified spontaneous emission from piecewise-constant composite bodies, showing that their emissivity can exhibit a high degree of directionality. However, the gain profile can be far from homogeneous in realistic settings. Here, we consider two important sources of inhomogeneities affecting population inversion of atomically doped media: inhomogeneous pump profiles and modifications stemming from changes to the emitters’ local radiative environment. Below lasing threshold, the latter stems primarily from changes to atomic decay rates, which can be either enhanced or suppressed through the Purcell effect (PE) [197]. Because PE is sensitive to the gain and geometry of the objects, such a dependence manifests as a nonlinear and nonlocal interaction (or feedback) between the atomic medium and the optical environment [164], which we model within the stationary-phase approximation [71] via a self-consistent renormalization of the (dressed) atomic parameters.
While the above discussion assumes that the temperature profile of nanoscale devices is largely governed by thermal conduction alone, in Chapter 3, we study the impact of thermal radiation when RHT is magnified at near-field separations. We begin by presenting a general coupled conduction–radiation (CR) framework that captures the interplay of near-field RHT and thermal conduction along with the existence of large temperature gradients in arbitrary geometries. Starting from simple slab–slab geometries, in the first half of the chapter, we derived an exact, closed-form expression of the RHT in the presence of arbitrary temperature distributions. We find that CR interplay can greatly modify the behavior of both temperature gradients and RHT, causing the latter to asymptote to a constant, the conductive flux through a gap-less slab, at short separations. We show that such effects are well within the reach of current-generation experiments, leading for instance to a decrease in RHT between two silica slabs of more than an order of magnitude at relatively large gap sizes \( \sim \) several tens of nanometers. In the second half of Chapter 3, to generalize the above analysis to bodies of arbitrary shapes, we extend the FVC formulation to capture spatially resolved power emission and absorption rates throughout the system. As a proof of concept, we consider an idealized geometry involving aluminum-zinc oxide nanorods separated by vacuum gaps, where one of the nanorods is heated by a large-temperature reservoir while the other is kept at 300 K. We show that bulk, nanorod resonances enhance RHT compared to typical surface–plasmon resonances in planar geometries, leading to large and nonlinear temperature gradients.

In Chapter 4, we apply optimization techniques to tailor thermal radiation with the primary goal of enhancing NFRHT at selective frequencies. In the first half of the chapter, we limit our design variables to one dimension, namely, multilayer structures with arbitrary dielectric profiles along the separation direction. We demonstrate that appropriately optimized, multilayer structures can lead to larger NFRHT compared to the best possible homogeneous thin films, a consequence of careful interference or
“rate matching” of scattered and absorbed surface waves. In situations corresponding to a dipole–plate geometry, we show that for sufficient small dipoles, NFRHT is proportional to the local density of states (LDOS) and can be enhanced by orders of magnitude at mid-field separations.

In the second half of Chapter 4, we move on to investigate NFRHT between two-dimensional gratings, an enlarged two-dimensional design space. We present a fast computational approach to the evaluation of NFRHT which is enabled by the low-rank properties of elliptic operators for disjoint bodies. Focusing on lossy metals far from the SPP condition at infrared wavelengths, e.g. tungsten (W), we find a grating geometry that selectively achieves 50% of the RHT of a pair of ideal, lossless metal plates satisfying the SPP condition, for separations as small as 2% of a tunable design wavelength. These predictions represent RHT enhancements of nearly two orders of magnitude compared to corresponding planar objects, confirming the potential of even relatively simple structuring for selectively enhancing RHT. Moreover, we provide direct evidence of a recently predicted material scaling law [313].

In Chapter 5, we shift attention to nonlinear frequency conversion. We begin by applying optimization directly to three-dimensional structures, including metasurfaces and compact cavities for achieving high-efficiency second-harmonic frequency conversion. For metasurfaces, compared to hand-designed structures, enhancement factors as large as $10^7$ are observed. Such enhancements arise because these optimization tools ensure both frequency and phase matching, in addition to large nonlinear overlap factors. For wavelength-scale cavities, limiting our choices to radiative lifetimes to $Q \lesssim 10^3$ for the sake of robustness and fast operational timescales, the optimized designs can achieve conversion efficiencies as high as those corresponding to ultra-high $Q$ resonators, such as large microring cavities ($Q \sim 10^7$). To operate on-chip, each mode in these cavities must be coupled to a source or detector in a controlled way, which in a typical evanescent scheme, is realized by tuning the gap
separation between the waveguide and cavity [79] at each frequency. Consequently at least two waveguides are needed for efficient second harmonic generation. We instead apply optimization algorithms to design compact devices that efficiently couple light consisting of multiple, widely separated wavelengths from a single waveguide into a wavelength-scale multi-resonant cavity. Motivated by practical problems in nonlinear optics, we pursue three illustrative examples: compact multi-resonant cavities with resonant features mimicking those used for second-harmonic, sum-frequency, and frequency comb generation. In each situations, we demonstrate either total or near total critical coupling.

1.1 Overview of inverse design

In the context of nanophotonics, inverse design formulations are understandably much more recent. Earlier optimizations mostly explored genetic algorithms [17] or simple gradient-based algorithms [139] applicable to few variables. With few exceptions, most works focused on two classes of problems, involving either bandgap optimization in photonic crystal cavities [111, 115, 149] or mode coupling in waveguide geometries [109, 128, 204]. These problems commonly shared a high degree of symmetry, and typically only optimize a small selection of parameters within a predetermined family of periodic designs. Starting from the early 2000s, large-scale optimization methods began to reshape this landscape [260, 318], broadening its scope to a wider range of technologically relevant applications: photonic crystal cavity waveguide bends showing sub-1% transmission losses over a broad band of frequencies [201], few-wavelength-thick devices capable of acting as frequency demultiplexers [172], and more varied photonic crystal cavity configurations for the creation of wide bandgaps [359, 214]. Moreover, introduction of adjoint topology [53] and level-set [58, 57] optimization vastly broadened the generality and computational efficiency of inverse design.
In a typical inverse design problem in nanophotonics, one seeks to maximize or minimize an objective function $\mathcal{F}(x_\alpha)$, subject to certain nonlinear constraints $\mathcal{G}(x_\alpha) \leq 0$, over a set of free variables or degrees of freedom (DOFs) $\{x_\alpha\}$ [410]. For an electromagnetic problem, $\mathcal{F}$ and $\mathcal{G}$ are functions of the electric $\mathbf{E}$ or magnetic $\mathbf{H}$ fields, which are solutions of Maxwell’s equations given some external current or incident field. While one can solve Maxwell’s equations in any ways, in this thesis, we focus on the frequency-domain equation,

$$\left[ \nabla \times \frac{1}{\mu} \nabla \times \mathbf{E} - \varepsilon(r)\omega^2 \right] \mathbf{E} = i\omega \mathbf{J}$$

(1.1)

describing the steady-state solution $\mathbf{E}(r; \omega)$ in response to currents $\mathbf{J}(r, \omega)$ at frequency $\omega$, as a function of material dielectric profiles $\varepsilon(r)$. Starting from some random initial guess, optimization techniques aim to automatically discover $\varepsilon(r)$ for a large figure of merit. Based on choices of DOFs to parameterize $\varepsilon(r)$, there are two major classes of optimization strategies in photonic inverse design [105], known as density topology optimization and shape (or level-set) optimization.

### 1.1.1 Topology optimization

In topology optimization, given a computational grid consisting of finite discretization of the underlying physical problem via either finite-element or finite-difference method, $\varepsilon(r)$ at each voxel $r_\alpha$ is treated as a DOF, where $\alpha$ denotes the index of the voxel [33]. For instance, the permittivity at each node $\varepsilon_\alpha$ in design regions involving two materials can be expressed as,

$$\varepsilon_\alpha = \varepsilon_b + x_\alpha (\varepsilon_m - \varepsilon_b)$$

(1.2)

where $\varepsilon_b(m)$ denotes the permittivity of the background (structure) material, and $x_\alpha \in [0, 1]$ the DOF, with its value 0(1) referring to background (structure) material.
Topology optimization can offer an enormous design space, in which the dimension, or the number of voxels in the design region, can reach billions [1]. The key to making such a large-scale optimization problem tractable is to explore a fast-converging, computationally efficient, and gradient-based optimization algorithms, such as the method of moving asymptotes (MMA) [421]. To enable gradient-based algorithms, there are at least two important considerations. First, the DOFs \( \{x_\alpha\} \), or the permittivities must be continuous variables, which though give rise to intermediate “grey” structures, posing challenges to experimental realizations. To obtain fabricable binary structures, recently people have borrowed techniques from image processing to penalize the intermediate values \( x_\alpha \in (0, 1) \) via a variety of filter and regularization methods [200]. Similar regularization methods are also applied to increase minimal feature sizes [419]. Second, crucial to gradient-based algorithms, the derivatives \( \frac{\partial F}{\partial x_\alpha} \), \( \frac{\partial G}{\partial x_\alpha} \) at every pixel \( \mathbf{r}_\alpha \) must be computed very efficiently, which is made feasible by exploiting the adjoint-variable method [200], requiring only few additional solves of Eq. 1.1. While those optimization approaches don’t guarantee global optimal solution, it is nevertheless possible to find designs that perform remarkably better than existing designs, with the help of relaxation methods such as frequency averaging [260].

### 1.1.2 Shape optimization

While topology optimization is quite efficient in handling the enormity of an unconstrained design space, it often leads to geometries with irregular features that are difficult to fabricate, requiring many steps of post-processing to eliminate those small features. An alternative approach that is in principle more conducive to fabrication constraints is to exploit shape or level-set optimizations, where \( \varepsilon(\mathbf{r}) \) is parameterized by a finite set of shapes, independent of the numeric discretization. The shapes can be freeform contours represented implicitly by so-called level sets [452], evolved dur-
ing the optimization via a level-set partial differential equation characterized by a velocity field. Alternatively, $\varepsilon(r)$ can be expressed explicitly with parameters of basic geometric entities such as positions and sizes of polytopes \cite{176}, which is our focus in this thesis. More specifically, we express $\varepsilon(r)$ as a sum of basic shape functions, $\varepsilon_\alpha = \sum_\beta H_\beta(r_\alpha; \{p_\beta\})$, described by shape functions $H_\beta$ and a finite set of geometric parameters $\{p_\beta\}$, where $\beta$ denotes the shape index. Here, to deal with potential overlap of two or more shapes, one can implement a filter function that enforces the same maximum-permittivity constraint $\varepsilon_\alpha \leq \varepsilon_s$.

To enable gradient-based optimization algorithms, $H_\beta$ must yield a continuous derivative, which is not feasible a priori due to the finite computational discretization and must instead be enforced by the use of a “smoothing Kernel”. Typically, one can implement $H_\beta$ to vary exponentially near the shape boundaries, where the smoothness is determined by a few simple parameters that can, at various points along the optimization, be slowly adjusted to realize fully binary structures upon convergence. Such a “relaxation” process \cite{176} is analogous to the application of a binary filter in the objective function \cite{200}. The derivatives of a given objective function $f$ (and associated constraints) can then be obtained via the chain rule $\frac{\partial F}{\partial p_i} = \frac{\partial F}{\partial x_\alpha} \frac{\partial x_\alpha}{\partial p_i}$.

### 1.2 Overview of near-field radiative heat transfer

Heat transfer is a ubiquitous phenomenon associated with non-equilibrium systems undergoing temperature gradients, in which energy flows from hotter to colder regions via conduction, convection, and/or radiation \cite{186}. Thermal radiation is of particular interest because electromagnetic waves mediate energy exchange, enabling long-distance heat transfer (even across vacuum gaps). Until recently, the prevailing understanding of radiative heat transfer relied largely on Kirchhoff’s law \cite{223} and Planck’s law \cite{353}, enabling quantitative estimation of thermal emission power.
from optical absorption measurements. More importantly, these century-old principles imply that RHT between any two bodies of arbitrary shapes and materials is bounded by that between two blackbodies (which perfectly absorb all incoming light) of the same temperatures and surface areas. The upper bound on the net (frequency integrated) emission from blackbodies is known as the Stefan–Boltzmann law \[387\], derived nearly a century ago under the revolutionary assumption of energy quantization of electromagnetic waves, which subsequently came to be known as photons \[231\].

However, such a derivation is only valid at macroscopic separations (where ray optics is valid) much larger than the thermal wavelength \(\lambda_T \sim \hbar c/k_B T\) (several microns at room temperature). At subwavelength scales \(\ll \lambda_T\), where near field or evanescent waves can contribute flux, the physics is dramatically altered by the tunneling of photons. In particular, evanescent contributions absent in the far-field cause the rate of heat exchange to scale inversely with separations down to nanometer scales, leading to flux rates many orders of magnitude larger than those predicted by the Stefan–Boltzmann law \[225, 221, 406, 412\]. This increased flux can be enhanced and controlled by resonant electromagnetic surface modes \[342, 68, 442, 15\], allowing heat to be concentrated into narrow and designable spectral bandwidths. These properties, in principle, provide means of significantly improving the degree to which heat can be manipulated compared to thermal conduction, leading to the consideration of applications and devices based on NFRHT with proposals ranging from thermophotovoltaics energy capture \[324, 244, 338, 194, 215, 411\] to high-density heat management \[166, 468, 219\], and heat assisted magnetic recording \[78, 38\].

To handle heat exchange at near-field separations, in the 1950s, Rytov et al \[389, 355\] introduced a fluctuational electrodynamics theory in which thermal emission is attributed to thermal and quantum fluctuations of currents inside macroscopic media. In analogy with Brownian motion, their average fluctuating density can be related to
material dissipation via the fluctuation–dissipation theorem (FDT) \[242\],

\[ \langle j_i(x, \omega) j_j^*(y, \omega) \rangle = \frac{4}{\pi} \omega \operatorname{Im} \chi(x, \omega) \Theta(x, \omega) \delta(x - y) \delta_{ij} \]  

(1.3)

where “\(\langle \cdots \rangle\)” denotes the ensemble average, \(\chi\) the material susceptibility, and \(\Theta(x, \omega) = \hbar \omega / (e^{\hbar \omega / k_B T(x)} - 1)\) the Planck function depending on the local temperature \(T(x)\). The same principles can be extended to compute spontaneous emission by replacing \(T(x)\) with the effective temperature associated with population inversion \[302\].

The advantage of the above formulation is that it allows direct application of well-established techniques in electromagnetism \[333\], including Dyadic Green’s functions, scattering matrices, and integral equation methods. Until recently, the majority of theoretical progress focused on highly symmetric shapes where analytic methods can be straightforwardly applied, such as planar structures \[319\, 213\, 84\, 67\, 137\, 441\, 477\, 27\], spherical bodies \[326\, 323\], and metamaterials under effective medium approximations \[101\, 274\, 478\]. For more complicated geometries, the Dyadic Green’s function and hence NFRHT must be evaluated numerically. Since thermal radiation arises from incoherent fluctuating currents, its description can reduce to a series of electrodynamics scattering calculations involving fields due to currents \[210\, 333\]. Naively, this involves repeated calculations of Green’s functions throughout the bodies. To handle arbitrary shapes, the Green’s function can be computed numerically under general-purpose methods, including finite-element and finite-difference time-domain \[378\], and frequency–domain solvers.

While the above framework can be computationally intensive, more efficient methods do exist. Since many structures of interest exhibit certain symmetries, one possibility is to explore suitable delocalized bases \[333\] given the symmetry properties, such as Fourier bases for periodic gratings, known as rigorous coupled wave analysis.
Another general approach is to exploit integral equations, based on the well-known surface or volume equivalence principle. The former, being expressed in the form of traces of either scattering [44, 46, 306, 303, 165, 294, 237, 155] or boundary-element [379, 381, 382] matrices, is particularly suitable for systems consisting of piecewise constant dielectric and temperature profiles. The latter, as will be discussed in this thesis, is general and can efficiently handle inhomogeneous structures, i.e., having temperature gradients or varying permittivities. In situations of low refractive index contrast, the discrete-dipole approximation (DDA), a class of volume integral equations, has been suggested as an efficient tool for computing RHT [119].
Chapter 2

Thermal radiation and spontaneous emission from heterogeneous objects

Thermal radiation and amplified spontaneous emission in heterogeneous geometries can exhibit new properties, especially in scenarios where the inhomogeneity occurs at the scale of the electromagnetic wavelength. In this chapter, we consider radiation problems involving both strong temperature and dielectric inhomogeneities. We show that our recently developed fluctuating–volume current (FVC) formulation of electromagnetic scattering can be suitably employed to efficiently handle a wide range of emission problems.

The organization of this chapter is as following. We begin by briefly reviewing the FVC formulation in Sec. 2.1. In Sec. 2.2, we demonstrate that recent advances in nanoscale thermal transport and temperature manipulation can be exploited to tailor thermal radiation from wavelength-scale composite bodies. Such objects can be designed to exhibit strong resonances and large temperature gradients, which in turn lead to large and highly directional emission at mid-infrared wavelengths. In
Sec. 2.3, we study amplified spontaneous emission from wavelength-scale composite bodies—complicated arrangements of active and passive media—demonstrating highly directional and tunable radiation patterns. We present a framework that self-consistently treats the renormalization of the (dressed) gain parameters in active media, capturing nonlinear feedback introduced by the Purcell effect on the active medium, and otherwise captured by the well-known Maxwell–Bloch equations.

2.1 Fluctuating–volume current formulation

In this section, we present a brief derivation of our FVC formulation for thermal radiation or spontaneous emission in inhomogeneous media, with validations and details of its numerical implementation described in our manuscript [357]. Our starting point is the volume–integral equations (VIE) formulation of EM scattering [356], describing scattering of an incident, 6-component electric (\(E\)) and magnetic (\(M\)) field \(\phi_{\text{inc}} = (E; H)\) from a body described by a spatially varying \(6 \times 6\) susceptibility tensor \(\chi(x)\). (For convenience, we omit the frequency \(\omega\) dependence of material properties, currents, fields, and operators.) Given a 6-component electric (\(J\)) and magnetic (\(M\)) dipole source \(\sigma = (J; M)\), the incident field is obtained via a convolution (\(*\)) with the \(6 \times 6\) homogeneous Green’s function (GF) of the ambient medium \(\Gamma(x, y)\), such that \(\phi_{\text{inc}} = \Gamma * \sigma = \int d^3y \Gamma(x, y)\sigma(y)\). Exploiting the volume equivalence principle [356], the unknown scattered fields \(\phi_{\text{sca}} = \Gamma * \xi\), can also be expressed via convolutions with \(\Gamma\), except that here \(\xi = -i\omega\chi\phi\) are the (unknown) bound currents in the body, related to the total field inside the body \(\phi = \phi_{\text{inc}} + \phi_{\text{sca}}\) through \(\chi\). Writing Maxwell’s equations in terms of the incident and induced currents, we arrive at the so-called JM–VIE equation [356]

\[
\xi + i\omega\chi(\Gamma * \xi) = -i\omega\chi(\Gamma * \sigma),
\]  

(2.1)
To numerically solve for $\xi$ in terms of the incident source $\sigma$, one typically applies the Galerkin discretization of Eq. 2.1 via expansions of the current sources $\sigma(x) = \sum_n s_n b_n(x)$ and $\xi(x) = \sum_n x_n b_n(x)$ in a convenient, orthonormal basis $\{b_n\}$ of $N$ 6-component vectors, with vector coefficients $s$ and $x$, respectively. The resulting matrix expression has the form $x + s = Ws$, where $(W^{-1})_{m,n} = \langle b_m, b_n + i\omega \chi(\Gamma \ast b_n) \rangle$ is known as the VIE matrix and $\langle , \rangle$ denotes the standard conjugated inner product.

Direct application of Poynting’s theorem then yields the far-field radiation flux $\Phi = \frac{1}{2} \text{Re} \int d^3x (E^\ast \times H) = -\frac{1}{2} \text{Re} (\xi + \sigma)^\ast \phi$ due to $\sigma$ in terms of the VIE operators [89]:

$$\Phi_\sigma = -\frac{1}{2} \text{Re} (\xi + \sigma)^\ast \Gamma \ast (\xi + \sigma) \quad (2.2)$$

$$= -\frac{1}{2} (x + s)^\ast \text{sym} G(x + s) \quad (2.3)$$

$$= -\frac{1}{2} s^\ast W^\ast \text{sym} GWs \quad (2.4)$$

$$= -\frac{1}{2} \text{Tr} [CW^\ast \text{sym} GW] \quad (2.5)$$

where we have defined the current–current correlation matrix $C$, whose elements are $C_{mn} = \langle s_m s_n^\ast \rangle = \int \int d^3x d^3y b_m^\ast(x) \langle \sigma(x) \sigma^\ast(y) \rangle b_n(y)$, in which the correlation functions $\langle \sigma(x) \sigma^\ast(y) \rangle$ satisfy FDT, defined in Eq. 1.3. Note FDT is only valid in situations where the temperature gradient is small compared to the material-dependent current–current correlation lengthscale (of the order of the atomic scale or phonon mean-free path), such that the charge distribution reaches local equilibrium [344, 357]. One can also obtain a similar trace expression for the angular radiation flux in a given direction, as shown in Ref. [357]. Another convenience is that if $\{b_n\}$ is chosen to be a localized basis of unit-amplitude, i.e. $\langle b_m, b_n \rangle = \delta_{mn}$, volume elements, then the flux contribution from a given volume (including different polarizations) $b_n$ is precisely the diagonal element $-\frac{1}{2} (CW^\ast \text{sym} GW)_{n,n}$.

Direct computation of Eq. 2.5 is expensive due to the large dimensionality $N$ of the problem, but it turns out that the Hermitian, negative-semidefinite, and low-rank
nature of $\text{sym } G$ (since it is associated with the smooth, imaginary part of the Green’s functions) enables re-expressing the trace as the Frobenius norm of a low-rank matrix. Specifically, decomposing $\text{sym } G = -U_r S_r U_r^*$ via a fast approximate SVD [185], where $r \ll N$ [75], and further decomposing $S_r = L_S L_S^*$, we find that the product $W^* \text{sym } GW$ can be written in the form $QQ^*$, with $Q = W^* U_r L_S$, reducing the calculation of the diagonal elements to a small series $r \ll N$ of scattering calculations (matrix-inverse operations) [356]. For example, while the calculations below require $N \geq 40^3$ basis functions to obtain accurate spatial resolution, we find that generally $r \lesssim 20$. Unlike scattering-matrix and surface-integral equation formulations of thermal radiation [381], the FVC scattering unknowns are volumetric currents, enabling to compute emission from arbitrarily shaped bodies with spatially varying dielectric and temperature properties.

2.2 Temperature control of thermal radiation in composite bodies

The ability to control thermal radiation over selective frequencies and angles through complex materials and nanostructured surfaces [161] has enabled unprecedented advances in important technological areas, including remote temperature sensing [301], incoherent sources [195 372], and energy-harvesting [123 36 131]. While unusual radiation patterns, e.g. spatial coherence and directivity, have been recently demonstrated in photonic crystals [161 104 455 131], metamaterials [110 248 138 277], and large-etalon structures composed of partially transparent materials [474], the design of temperature distributions in composite structures with features on the scale of the thermal wavelength (for enhanced emission) appears largely unexplored. Such an approach combines both nanophotonic and conductive design principles and is poised
to take advantage of recent progress in the areas of temperature management and heat transport at sub-micron scales \[64, 63\].

### 2.2.1 Overview

Attempts to obtain unusual thermal radiation patterns have primarily relied on Bragg scattering and related interference effects in nanostructured surfaces \[161\], including photonic gratings \[104, 455, 131\], metasurfaces \[160, 298, 213, 183, 325, 297, 226, 371\], multilayer structures \[228, 30, 117, 451\], and sub-wavelength metamaterials \[248, 138, 277, 36\]. Related ideas can also be found in the context of fluorescence emission, where directivity is often achieved by employing metallic objects (e.g. plasmonic antennas) to re-direct emission from individual dipolar emitters via gratings \[99, 230\] or by localizing fluorescent molecule(s) to within some region in the vicinity of an external scatterer \[425, 426, 251, 437, 436, 315\]. Matters become complicated when the emission is coming from random sources distributed within a wavelength-scale object, as is the case for thermal radiation, because the relative contribution of current sources to radiation in a particular direction is determined by both the shape and temperature distribution of the object. Optical antennas have recently been proposed as platforms for control and design of narrowband emitters \[390, 331\], though predictions of large directivity continue to be restricted to periodic structures. While there is increased focus on the study of light scattering from subwavelength particles and microwave antennas (useful for radar detection \[22\], sensing \[424\], and color routing \[3, 395\]), similar ideas have yet to be translated to the problem of thermal radiation from compact, wavelength-scale objects, whose radiation is typically quasi-isotropic \[161\]. Here, we show that temperature manipulation in composite particles could play an important role in the design of coherent thermal emitters.

Temperature gradients can arise near the interface of materials with highly disparate thermal conductivities \[63\]. Although often negligible at macroscopic
scales [21], recent experiments reveal that the presence of thermal boundary resistance [368, 295] (including intrinsic and contact resistance [413]) in nanostructures together with large dissipation can enable temperature localization over small distances [21]. Such temperature control has been recently investigated in the context of metallic nanospheres immersed in fluids [304], graphene transistors [196], nanowire resistive heaters [471], AFM tips [222], and magnetic contacts [348]. With the exception of a few high-symmetry structures, e.g. spheres [112] and planar films [451], however, calculations of thermal radiation from wavelength-scale bodies have been restricted to uniform-temperature operating conditions, exploiting Kirchoff’s law [287, 451] to obtain radiative emission via simple scattering calculations.

Figure 2.1: Schematic of two composite bodies, involving GST (blue) hemispheroids coated with Ti (green), AZO (red), or Si$_3$N$_4$ (orange) shells, which rest on a low-index, transparent substrate in contact with a heat reservoir at 300 K. The GST is heated from below by a conductive 2D material (e.g. a carbon-nanotube wall or graphene sheet), leading to temperature gradients within the structure, with temperature profiles illustrated in the lower insets. The presence of boundary resistance at material interfaces is captured by effective thermal resistances (see text).

### 2.2.2 Directional thermal emission

We explore radiation from composite bodies comprised of chalcogenide Ge$_2$Sb$_2$Te$_5$ (GST) alloys and metals or semiconductors. To begin with, we consider micron-scale GST hemispheroids coated with either titanium (Ti) or silicon-nitride (Si$_3$N$_4$) shells, depicted in Fig. 2.1. The structures rest on a low-index ($\varepsilon \approx 1$) transparent substrate which not only provides mechanical support but also a means of dissipating heat away
from the structure; the bottom of the substrate is assumed to be in contact with a 300 K heat reservoir while surfaces exposed to vacuum satisfy adiabatic boundary conditions ($\nabla T \cdot \vec{n} = 0$). When heated by a highly conductive 2D material (e.g. by applying current through a carbon nanotube wall [464] or graphene sheet) at the GST-substrate interface, such a structure can exhibit large temperature gradients within the core, a consequence of boundary resistance [368, 304, 413, 295] between the various interfaces and rapid heat dissipation in the highly conductive shells [368, 464, 259].

To model the corresponding steady-state temperature distribution $T(\vec{x})$, we solve the heat-conduction equation via COMSOL \textsuperscript{1} including the full temperature-dependent thermal conductivity $\kappa(T)$ of the GST [290]. Note that even at large temperatures, $\kappa_{\text{GST}}(800 \, \text{K}) \ll \kappa_{\text{Ti, Si}_3\text{N}_4}(300 \, \text{K}) \gtrsim 20 \, \text{W/mK}$. The existence of thermal boundary resistance [413, 186] at this scale is taken into account by the introduction of effective resistances $R_{\text{sh}c}, R_{\text{h}h}\text{su}$, and $R_{\text{sh}h}\text{su}$, at the interfaces between shell-GST, heater-substrate, and shell-substrate, respectively, detailed in Sec. A.2. Figure 2.1 shows $T(\vec{x})$ throughout the Ti structure when the GST-substrate interface is heated to $T_{\text{GST}} = 870 \, \text{K}$ (approaching the GST melting temperature [428] and corresponding to a thermal wavelength $\lambda_T \approx hc/2.8k_B T \approx 5.8 \mu\text{m}$), and under various operating conditions. We consider $R_{\text{sh}h}\text{su} = 10^{-8} \text{m}^2\text{W/K}$ [368] and $R_{\text{sh}c} = R_{\text{h}h}\text{su} = R_{\text{th}}$, with (i), (ii), and (iii) corresponding to typical values of $R_{\text{th}} = \{0.5, 1, 2\} \times 10^{-7} \text{m}^2\text{W/K}$ [368] while (iv) $R_{\text{th}} = \infty$ and (v) $R_{\text{th}} = 0$ describe either perfect temperature localization in the GST or uniform temperature throughout the structure, respectively. (Note that the values of boundary resistance associated with a given experimental realization of these composites will depend on the specific choice of fabrication technique and materials.) In what follows, we exploit $T(\vec{x})$, Eq. 2.5, and the dielectric properties (full material dispersion) of these materials [401, 256, 300, 224], to obtain the flux from these composites. Note that due to large temperature gradients and phase transitions

\textsuperscript{1} Note that at these temperatures convective and radiative effects are negligible compared to conductive transfer, allowing us to consider the radiation and conduction problems separately.
in the GST [290], its dielectric response \( \varepsilon_{\text{GST}}(T(\vec{x}),\omega) \) consists of continuously varying rather than piece-wise constant regions, detailed in Sec. A.1; our FVC method, however, can handle arbitrarily varying \( \varepsilon(\vec{x}) \) and \( T(\vec{x}) \). (Note here temperatures inside the Ti and Si\(_3\)N\(_4\) shells are roughly 300 K and thus we employ their permittivities at room temperature.)

Figure 2.2: Spectral emissivity \( \epsilon(\omega) = \Phi(\omega)/\Phi_{\text{BB}}(\omega) \) (blue dots) from heterogeneous bodies comprising Ge\(_2\)Sb\(_2\)Te\(_5\) (GST) hemispheroids coated with either (a) Si\(_3\)N\(_4\) or (b) Ti shells and resting on a low-index substrate (object dimensions are specified in Fig. 2.1). The structures are heated from the GST–substrate interface by a 2D thin-film conductor under heating scenario (iii), described in the text, with an interface temperature \( T_{\text{GST}} = 870 \) K. The resulting temperature profiles \( T(\vec{x}) \) are shown in Fig. 2.1 and on the insets. \( \epsilon \) is defined as the ratio of the thermal flux \( \Phi(\omega) \) of each body normalized to the flux \( \Phi_{\text{BB}} = \frac{A}{4\pi} \left( \frac{\omega}{c} \right)^2 \Theta(\omega,T) \) from a corresponding black body of the same surface area \( A \) but uniform \( T = 870 \) K (green lines, arb. units). Also shown are the partial directivities \( \eta_\pm = \Phi_\pm/\Phi \) (red line), defined as the ratio of the outgoing flux into the upper/lower hemisphere \( \Phi_\pm = 2\pi \int_{\pi/2}^{\pi/2} d\theta \Phi(\omega,\theta) \) to the total flux, where \( \theta \) is defined in Fig. 2.1 (c) Angular radiation intensity \( \Phi(\theta) \) normalized by the total flux \( \Phi \), for the structures in (a) (solid red line) and (b) (solid blue line) under scenario (iii), compared to emission under uniform-temperature conditions (dashed lines).

Figure 2.2 shows the emissivity (blue dots) and partial directivity (solid lines) of the (a) Si\(_3\)N\(_4\) and (b) Ti structures, along with the corresponding \( T(\vec{x}) \) (insets) under heating scenario (iii), assuming \( T_{\text{GST}} = 870 \) K. The emissivity \( \epsilon(\omega) = \Phi(\omega)/\Phi_{\text{BB}}(\omega) \) is defined as the ratio of the thermal flux \( \Phi(\omega) \) from each object to that of a blackbody \( \Phi_{\text{BB}}(\omega) = \frac{A}{4\pi} \left( \frac{\omega}{c} \right)^2 \Theta(\omega,T) \) of the same overall surface area \( A \) and \( T = 870 \) K (green lines); the partial directivity \( \eta_\pm = \Phi_\pm/\Phi \) is defined as the ratio of the flux into the
upper/lower hemisphere $\Phi_\pm(\omega) = 2\pi \int_{\mp \pi/2}^{\mp \pi/2} d\theta \Phi(\omega, \theta)$, to the total flux $\Phi$, where $\theta$ is defined with respect to the $\hat{z}$ axis [Fig. 2.1]. As expected, $\epsilon$ displays enhanced emission at geometric resonances, with peak magnitudes $\max \epsilon \lesssim 0.2$ limited by material losses ($\Im \epsilon \lesssim \Re \epsilon$) in this frequency range [50]. (Larger $\epsilon$ can likely be obtained with further design or other material combinations.) Also notable is the sharp increase in $\eta$ as the emission regime changes from quasistatic to wavelength scale: despite the temperature localization, at low $\omega L/c \ll 1$ the emission is quasi-isotropic (as expected from a randomly polarized dipolar emitter [160]). In contrast, one finds that with increasing $\omega L/c \gtrsim 1$ and with the help of the curvature [474], the Si$_3$N$_4$ and Ti shells increasingly redirect radiation upwards or downwards as a result of coherent interference between the radiated and scattered fields. Such strong directional scattering is absent from dipoles fluctuations originating in the shells, which tend to radiate isotropically and dominate emission, making the design of the temperature profile an essential ingredient for achieving large $\eta$. Figure 2.2(c) illustrates this point by showing the angular radiation intensity $\Phi(\theta)$ of the Si$_3$N$_4$ (red) and Ti (blue) structures at selected frequencies, under two of the heating conditions, corresponding to either (ii) partial temperature localization in the GST (solid lines) or (v) uniform temperature throughout the bodies (dashed lines). The dependence of radiation on $T(\vec{x})$ is further explored in Fig. 2.3 which shows the total (frequency-integrated) partial directivity $P_- = H_-/H$ of the Ti structure under different $T_{\text{GST}}$ and heating conditions, where $H_\pm = \int_0^\infty d\omega \Phi_\pm$. As shown, $P_-$ increases with increasing temperature localization (attained at larger $T_{\text{GST}}$), and remains almost constant, $P_- \approx 0.5$, under uniform-temperature conditions. Such an increase in partial directivity, however, comes at the expense of smaller $\epsilon$ (not shown), a consequence of the increasingly dominant role of higher frequencies which suffer from larger material losses.

Our choice of shapes and materials for these composites ensure large $\epsilon$ and $\eta$ at selective wavelengths $\omega L/c \sim 1$, dictated by our choice of operating temperature and
Figure 2.3: Total (frequency integrated) partial directivity $P_- = H_-/H$ as a function of $T_{GST}$, with $H_- = \int d\omega \Phi_-$, for the Ti structure under different heating conditions, corresponding to multiple degrees of temperature localization in the GST (see text).

satisfying the following general design criteria: (i) cold (scatterer) and hot (emitter) regions with large and small $\kappa$, respectively, in order to achieve and maintain a large temperature differential; (ii) hot regions with large $\text{Re}\, \epsilon$ and small losses supporting strong resonances near $\lambda_T$; (iii) large index contrasts and optimal shapes/sizes leading to high thermal extraction/reflections from hot regions and thus large directivity.

GST/Ti/Si$_3$N$_4$ material combinations partially satisfy these conditions, since $\varepsilon_{GST} \approx 30 + 5i$ $^{[256]}$ $^{[101]}$, $\varepsilon_{Ti} \approx -100 + 80i$ $^{[300]}$, and $\varepsilon_{SiN4} \approx 5 + 0.1i$ $^{[224]}$ at mid-infrared wavelengths $\lambda_T \approx 5.8\mu m$. Optimal shapes for achieving directional emission depend strongly on the choice of hot/cold material and operating temperature: we find that GST–Si$_3$N$_4$ composites favour large-curvature prolate structures while GST–Ti/AZO composites favour oblate structures, ensuring increased extraction and reflections from the shells, respectively. In Fig. 2.2 the geometric parameters are optimized to meet the above criteria, corresponding to $L_0 = 1.7\mu m$, $t_0 = 0.3\mu m$, $h_0 = 0.5\mu m$ for GST–Ti, and $L_0 = 1.3\mu m$, $t_0 = 1.3\mu m$, $h_0 = 2.6\mu m$ for GST–Si$_3$N$_4$.

Figure 2.4(a)(b) illustrate the sensitive relationship between shape and emission for the GST–Ti structure under heating scenario (iv), with the GST and Ti regions held at $870 \text{ K}$ and $300 \text{ K}$, respectively. Shown are the peak $\eta_-$ (dashed lines, right axes) and $\epsilon$ (solid lines, left axes) corresponding to the resonance closest to the peak.
of the BB spectrum, as a function of various geometric parameters, including: (i) aspect ratio $2h/D$ (magenta line), demonstrating that directivity is maximized at a specific aspect ratio $\approx 0.4$ and thus validating our previous observation that oblate objects ($2h/D < 1$) tend to maximize directivity in this configuration; (ii) overall object size $D$ with all other parameters fixed (black line), showing that there exists an optimal $D_0$, due mainly to material dispersion; (iii) Ti thickness $t$ with fixed GST size $L = L_0$ (red line), showing that there is an optimal shell thickness ($t_0$) which maximizes reflections (note that as $t$ increases, $\epsilon$ decreases due to larger metal losses); (iv) GST size $L$ with a fixed overall size $D = D_0$ (blue line), showing two optimal values of $L = 0.4L_0$ and $1.1L_0$, with the former exhibiting slightly larger $\eta_-$ but leading to lower...
emissivity: $\epsilon \to 0$ as $L \to 0$ due to increased radiative losses. These results illustrate the dramatically different design criteria associated with wavelength- versus large-scale bodies (where Kirchoff’s law is valid \[456\]). For instance, while larger $\eta_-$ can be attained in the ray-optics limit by increasing the shell thicknesses relative to the core dimensions (thereby enhancing extraction/reflections from the GST), this would also result in lower $\epsilon$. Wavelength-scale structures not only provide a high degree of temperature and emission control, but enable simultaneous enhancement in $\eta$ and $\epsilon$ with the latter potentially exceeding the blackbody limit \[50\]. Figure 2.4(c) shows radiation patterns from other composite shapes, with geometric parameters detailed in Sec. A.3 including mushroom-like particles, and cylinders (blue/red/orange denote GST/AZO/Si$_3$N$_4$), at selective $\omega$ and under heating scenario (iv), demonstrating a high degree of emission tunability.

2.2.3 Concluding remarks

Our predictions above provide just a glimpse of the interesting phenomena that can arise in structures that combine conductive and radiative design principles. Moreover, they offer proof of the need for rigorous models that account for the presence of temperature and dielectric gradients in phase-change and thermal-interface materials, along with their impact on emission. While our focus here is on compact objects, extensions to periodic structures are also feasible within the same framework, requiring minor modifications \[159\]. Similar ideas and techniques apply to problems involving fluorescence or spontaneous emission in which case rather than controlling the temperature profile, it is possible to localize and control the sources of emission by doping \[25\] or judicious choice of incident laser light \[247\].
2.3 Spontaneous emission from gain–composite structures

Noise in structures comprising passive and active materials can lead to important radiative effects [360], e.g. spontaneous emission (SE) [4], superluminescence [107], and fluorescence [241]. Although large-etalon gain amplifiers and related devices have been studied for decades [5, 153, 171, 453], there is increased interest in the design of wavelength-scale composites for tunable sources of scattering and incoherent emission [472, 148], or for realizing perfect absorbers [94] at mid-infrared and visible wavelengths.

2.3.1 Overview

Gain–composite structures are the subject of recent theoretical and experimental works [360], and have been studied in a variety of contexts, including spasers (combinations of metallic and gain media) with low-threshold characteristics [34, 414, 415, 337, 14], random structures with special absorption properties [475, 9, 16, 261, 137], and nano-scale particles with highly tunable emission and scattering properties [125, 363]. \(\mathcal{PT}\)–symmetric structures have received special attention recently as they shed insights into important non-Hermitian physics, such as design criteria for realizing exceptional points [180], symmetry–breaking [180], uni-directional scattering [268, 146, 2], and lasing thresholds [239]. Until recently, most studies of radiation/scattering from \(\mathcal{PT}\) structures remained confined to 1d and 2d geometries [95, 283, 146, 2, 314, 157, 429, 472]. In such low-dimensional systems, it is common to employ scattering matrix formulations [283, 146, 95] to solve for the complex eigenmodes and scattering properties of bodies, leading to many analytical insights. For instance, while the introduction of gain violates energy conservation, a generalized optical theorem can be obtained in 1d, establishing conditions for unidirectional trans-
mission of light [268 146 2]. Other studies focus on 2d high–symmetry objects such as cylindrical or spherical bodies [314 157] or particle lattices [429], demonstrating strong asymmetric and gain-dependent scattering cross-sections, while 3d structures such as ring resonators have been studied within the framework of coupled-mode theory [129 343]. With few exceptions [472], however, most studies of gain–composite bodies have focused on their scattering rather than emission properties.

Furthermore, while these systems are typically studied under the assumption of piecewise-constant [472 95 146] or linearly varying [148] gain profiles, in realistic situations, inhomogeneities in the pump or material parameters (e.g. arising from PE [164], hole burning [173 402], and gain saturation [402]) result in spatially varying dielectric profiles which alter SE. For example, the highly-localized nature of plasmonic resonances in spasers result in strongly inhomogeneous pumping rates [414] and orders-of-magnitude enhancements in atomic radiative decay rates [479]. In random lasers [475], partial pumping plays an important role in determining the lasing threshold [10 16] and directionality [261 147]. Rigorous descriptions of lasing effects in these systems commonly resort to solution of the full Maxwell–Bloch (MB) equations, in which the electric field $E$ and induced (atomic) polarization field $P$ couple to affect the atomic population decay rates [72]. However, the MB equations are a set of coupled, time-dependent, nonlinear partial differential equations [122] which, not surprisingly, prove challenging to solve except in simple situations involving high-symmetry [479 184] or low-dimensional structures [360]. Despite their overhead, general FDTD methods have been employed to study the transition from amplified spontaneous emission (ASE) to lasing in 1d random media [9 60], 2d metamaterials [364 184], photonic crystals [37], and more recently, nano-spasers [479].

A more recent, general-purpose method that is applicable to arbitrary structures is the steady-state ab-initio laser theory (SALT), an eigenmode formulation that exploits the stationary-inversion approximation to remove the time dependence and
internal atomic dynamics of the MB equations \cite{73, 71, 74}, yet captures important nonlinear effects such as hole burning and gain saturation \cite{71} through effective two-level polarization and population equations \cite{432}. The resulting nonlinear eigenvalue equation can be solved via a combination of Newton-Raphson \cite{122, 59}, sparse-matrix solver \cite{182}, and nonlinear eigenproblem \cite{167} techniques, exploiting either spectral “CF” basis expansions (especially suited for structures with special symmetries) \cite{430} or brute-force methods \cite{8} that can handle a wider range of shapes and conditions. Although this formulation can describe many situations of interest, it nevertheless poses computational challenges in 3d or when applied to structures supporting a large number of modes \cite{122}. Furthermore, the impact of noise below or near threshold has yet to be addressed, though recent progress is being made along these directions \cite{122, 59}.

Below and near the lasing threshold, stimulated emission is often negligible, enabling linearized descriptions of the gain medium \cite{402, 65, 72, 96}. Such approximations, however, ignore nonlinearities stemming from the induced radiation rate $\sim \mathbf{E} \cdot \mathbf{P}$ present in the MB equations, which captures feedback on the atomic medium due to amplification or suppression of noise from changes in the local density of states (also known as Purcell effect) \cite{479, 470, 245}. Here, we show that PE can be introduced into the linearized framework via a self-consistent renormalization or dressing of the gain parameters, an approach that was recently suggested \cite{164} but which has yet to be demonstrated. In particular, working within the scope of the linearized MB equations and stationary-inversion approximation, we capture the nonlinear feedback of ASE on gain, i.e. the steady-state enhancement/suppression of gain and atomic decay rates due to PE, via a series of nonlinear equations involving many coupled, linear, classical scattering calculations—local density of states or far-field emission due to electric dipole currents. Since the gain profile can become highly spatially inhomogeneous, it is advantageous to tackle this problem using brute-force methods, e.g. finite
differences \cite{423}, finite elements \cite{90}, or via the scattering VIE framework described below. Our FVC method described in Sec. 2.1 is particularly advantageous in that it is general and especially suited for handling scattering problems with large numbers of degrees of freedom (defined only within the volumes of the objects), in contrast to eigenmode expansions \cite{430} which become inefficient in situations involving many resonances \cite{422, 430} or near-field effects \cite{274, 218}.

### 2.3.2 Active medium: 4-level atoms

![Figure 2.5: Schematic of a gain–loss composite body consisting of a dense (active) collection of 4-level atoms (left inset) optically pumped via monochromatic light at frequency $\omega_{30}$ and pump rate $P$, resulting in population inversion and spontaneous emission (SE) at $\omega_{21}$. The nonlinear feedback of Purcell effect on the SE rate of the object is captured via a system of nonlinear equations illustrated on the right schematic, in which: (a) the dielectric constant at the location of the $n$th emitter depends on the corresponding atomic decay rate $\gamma_{21,n}$ through Eq. 2.6; (b) the radiative emission $\Phi_n$ of a classical dipole at the location of the $n$th emitter, described by Eq. 2.5, is altered, e.g. enhanced or suppressed, by the surrounding $\epsilon(x)$; and (c) the radiative decay rate $\gamma_{21,n}$ of the emitter is dressed by the Purcell effect through Eq. 2.8 which in turn affects the dielectric response in (1).](image)

To begin with, we review the linearized description of a gain medium consisting of optically pumped 4-level atoms, as shown in Fig. 2.5: a dense collection of active emitters (e.g. dye molecules \cite{402} or quantum dots \cite{417, 162}) embedded in a passive (background) dielectric medium $\epsilon_r$. (Note that our choice of 4-level system here is merely illustrative since the same approach described below also applies to other
active media.) Below threshold where stimulated emission (and effects such as hole burning) can be safely ignored \[402\], the effective permittivity of such a medium can be well approximated by a simple 2-level Lorentzian gain profile \[402, 65, 96\],

\[
\epsilon(\omega, x) = \epsilon_r(\omega) + \frac{4\pi g(x)^2}{\hbar \gamma_{\perp}} \frac{\gamma_{\perp} D_0(x)}{\omega - \omega_{21} + i \gamma_{\perp}} \frac{1}{\epsilon_g(x, \omega)},
\] (2.6)

where \(\epsilon_g\) depends explicitly on the frequency \(\omega_{12}\), polarization decay rate \(\gamma_{\perp}\) (or gain bandwidth), coupling strength \(g^2 = \frac{3\hbar c^3}{2\sqrt{\epsilon_r} \gamma_{21} \omega_{21}}\) \[96, 65\], and inversion factor \(D_0 = n_2 - n_1\) associated with the \(2 \rightarrow 1\) transition. Under the adiabatic or stationary-inversion approximation \[71\] and assuming that the system is pumped at \(\omega_{30}\), the steady-state population inversion is given by:

\[
D_0 = \frac{\left(1 - \frac{\gamma_{21}}{\gamma_{10}}\right) P / \gamma_{21}}{1 + \left(A \gamma_{32}^{\perp} + \frac{\gamma_{21}}{\gamma_{10}} + 1\right) P / \gamma_{21}} n.
\] (2.7)

(Note that in a dense medium, \(\gamma_{\perp} \gg \gamma_{21}\) is dominated by collisional and dephasing effects \[402\].) Here, \(A = 1(2)\) for incoherent (coherent) pump source; \(c\) denotes the speed of light, \(n_i\) the population (per unit volume) of level \(i\), \(n = \sum_i n_i\) the overall atomic population, \(\gamma_{ij} = \gamma_{ij}^{r} + \gamma_{ij}^{nr}\) the decay rate from level \(i \rightarrow j\), consisting of radiative and non-radiative terms, respectively, and \(P(x) = \frac{\sigma c}{\sqrt{\epsilon_r(\omega_{30}) \hbar \omega_{30}}} |E(x, \omega_{30})|^2\) is the position-dependent pump rate from \(0 \rightarrow 3\), which is often the main source of spatial dispersion.

The SE properties of such a medium are described by the fluctuation–dissipation theorem (FDT) \[302\ 199\ 158\]. In particular, local thermodynamic considerations imply that the presence of absorption or amplification must be accompanied by fluctuating polarization currents \(\sigma\) whose correlation functions, \(\langle \sigma_i(x, \omega) \sigma_j^*(y, \omega) \rangle = \frac{4}{\pi} \text{Im} \epsilon_g(x, \omega) n_2/(n_1 - n_2) \delta(x - y) \delta_{ij}\), depend on the corresponding macroscopic gain profile \(\text{Im} \epsilon_g\) and population inversion, \(n_2/(n_1 - n_2)\) \[351\]. The latter is often de-
scribed in terms of an effective (local) temperature $T$ defined with respect to the Planck spectrum $\left(e^{\hbar \omega/k_B T} - 1\right)^{-1}$, whose value turns negative for systems undergoing population inversion, $n_2 > n_1$, where $T \to 0^-$ in the limit of complete inversion and $T \to -\infty$ when the system is in the ground state. In the particular case of a steady-state four-level system, the relative populations $n_2/n_1$ depend only on the relative decay rates and one finds that $k_B T = \hbar \omega \ln\left(n_1/n_2\right) = \hbar \omega \ln\left(\gamma_{10}/\gamma_{21}\right)$ depends only implicitly on the pumping rate only through changes in the atomic decay rates (Purcell effect); in contrast, the effective temperature of three-level systems depends explicitly on the pumping rate. Of course, in addition to the fluctuations of gain atoms, the passive host medium will also exhibit thermal fluctuations, depending on its thermodynamic temperature and loss rate, which is described by the standard FDT. The overall emission will be the summation of the two independent radiations. In this chapter, our focus is to study SE at transition frequencies $\omega_{21}$ in the range of mid-infrared and visible wavelengths, where the flux rate from the active medium tends to dominate. The presence of inhomogeneities in the dielectric function and fluctuation statistics can be a hurdle for calculations of SE that rely on scattering-matrix formulations, but the FVC formulation described in Sec. 2.1 can capture all of the relevant physics, where the only modification is to replace the thermal temperature with the effective temperature in the current–current correlation matrix $C$.

2.3.3 Self-consistent treatment of Purcell effect

Although often assumed to be uniform below threshold, the atomic radiative decay rates $\gamma_{ij}$ entering Eq. 2.6 are in fact position dependent due to PE, leading to changes in the dipole coupling $g$ and population inversion factor $D_0$, either enhancing or suppressing (quenching) gain. In what follows, we only consider modifications to the radiative decay rate at the lasing transition $\gamma_{21}$; unlike the pump which is
incident at a non-resonant frequency, changes in $\gamma_{21}^r$ can have a significant impact on SE and must therefore be treated self-consistently.

The impact of PE on the radiative decay rate $\gamma_{21}^r$ of an atom at some position $b_n$ is captured by the coupling of the atomic polarization and electric fields $\sim \mathbf{P} \cdot \mathbf{E}$, or the induced radiation term in the MB equations, in the presence of the noise and surrounding dielectric environment \cite{447}. While technically this requires abandoning the linear model above, the weak nature of noise (ignoring stimulated emission) implies that the latter can also be obtained (perturbatively) from a linear, classical calculation: the radiative flux $\Phi_{bn}$ from a classical dipole at $b_n$. Specifically, the renormalized or dressed decay rate of an atom at position $x$ can be expressed as \cite{470}:

$$\gamma_{21}^r(x) = \mathcal{F}(x)\gamma_{21}^{r,0}, \quad (2.8)$$

where $\mathcal{F}(x)$ denotes the Purcell factor of a dipole at $x$, and the supper-script “0” denotes the decay rate of the atomic population in the lossy (background) medium.

It follows that the decay rate associated with a given $b_n$ and entering Eq. 2.6 is given by $\gamma_{21}(b_n) = \gamma_{21}^{nr} + \mathcal{F}_{bn} \gamma_{21}^{r,0}$, where the Purcell factor,

$$\mathcal{F}_{bn} = \Phi_{bn}/\Phi^0. \quad (2.9)$$

Here $\Phi_{bn} \sim (W^* \text{sym} GW)_{n,n}$ can be computed very fast via FVC formulation\footnote{Note $\mathcal{F}_{bn}$ does not depend on the amplitude of the fluctuations since it is measuring the relative radiation rate of a classical emitter at this position.}. Here, we assume that the bulk (background) medium $\epsilon_r$ only has a significant impact on $\gamma_{21}^{nr,0}$ (obtained either experimentally or theoretically by accounting for atomic interactions within the bulk) \cite{402} but not on the radiative decay rate $\gamma_{21}^{r,0}$, in which case $\Phi^0 = \omega^4/12\pi\epsilon_0 c^3$ is the emission rate of the atom in vacuum (assuming a unit-amplitude dipole). Note that in a lossy medium, e.g. in metals, the bare $\gamma_{21}^0$ will
be dominated by non-radiative processes \[462\] [126], leading to small quantum yields (QY) \(\gamma_{21}^r/\gamma_{21}^0 \ll 1\). Technically, the calculation of the Purcell factor requires integration over the gain bandwidth, \(F_b = \int d\omega \frac{\gamma_{\perp}}{\gamma_{\perp}^2 + (\omega - \omega_{21})^2} \Phi_{bn}(\omega) / \Phi_0(\omega)\), but here we make the often-employed and simplifying assumption that \(\gamma_{\perp} \gg \gamma_{21}\) and \(\gamma_{\perp} \lesssim\) spectral radiative features \[285\] [236], so as to only consider radiation at \(\omega_{21}\).

The gain profile of a body subject to an incident pump rate \(P\) can be obtained by enforcing that Eq. 2.6 and Eq. 2.8 be satisfied simultaneously. Such systems of nonlinear equations are most often solved iteratively using one or a combination of algorithms, ranging from simple fixed-point iteration \[420\] to more sophisticated approaches like Newton–Raphson and nonlinear Arnoldi methods \[122\] [446]. Essentially, as illustrated in Fig. 2.5 starting with the bare parameters, dressed decay rates are computed via Eq. 2.8 from the radiation equation Eq. 2.5 after which, having updated the gain–medium equation Eq. 2.6, the entire process is repeated until one arrives at a fixed-point of the system. In principle, this requires hundreds of thousands \(N\) of scattering calculations (flux from each dipole source in the active region) to be solved \textit{per iteration}, which becomes prohibitive in large systems, but the key here is that the entire spatially varying flux \(\{\Phi_{bn}\}\) throughout the body can be computed extremely fast, requiring far fewer \((\ll N)\) scattering calculations (as described above). Note that these large systems of nonlinear equations have many fixed points and hence convergence to the correct solution is never guaranteed, depending largely on the initial guess and algorithm employed \[448\]. However, a convenient and effective approach is to begin by first solving the system in the fast-converging (passive) regime \(Pn/\gamma_{12} \ll 1\), and then employing this solution as an initial guess at larger pumps.

In what follows, we begin by showing that wavelength-scale composite bodies can exhibit highly complex, tunable, and directional radiation patterns. Although it is not surprising that objects undergoing ASE (once known as “mirrorless lasers”) exhibit highly directed radiation patterns \[366\], few studies have gone beyond large-etalon
Fabry-Perot cavities [480] or fiber waveguides [175], often modelled via ray-optical or scalar-wave equations [360], which miss important effects present in wavelength-scale systems [122]. Further below, we show that dielectric inhomogeneities arising from the pump and/or radiation process can also introduce important changes to the ASE patterns. In particular, we apply the renormalization approach described above to consider the nonlinear impact of Purcell effect on the gain medium, and show that while in many cases a homogeneous approximation leads to accurate results, there are situations where these can fail dramatically. Our calculations are only meant to serve as proof of principle and revolve around highly doped (Er$^{3+}$ and Rhodamine) but simple dielectric objects, allowing faster computations but requiring very large values of Im$\epsilon_g$ to achieve significant gain. Similar results follow, however, in systems subject to smaller $\epsilon_g$ or smaller doping densities, at the expense larger pump powers or by exploiting resonances with greater confinement or smaller radiative loss rates (e.g. compact bodies of larger dimensions and/or refractive indices, or more complex structures such as photonic-crystal resonators).

2.3.4 Radiation patterns from heterogeneous particles

We begin by exploring SE from piecewise-constant $\mathcal{PT}$-symmetric spheres [Fig. 2.6 insets] consisting of a background dielectric medium, e.g. nano-composite polymers [476, 292] or semiconductors [121, 129], doped with active materials to realize different ($N$) regions of equal gain (orange) or loss (gray). Figure 2.6(a) shows the SE flux $\Phi$ from spheres of varying $N = \{1, 2, 3, 4\}$ (black, green, blue, red lines) and gain/loss permittivities $\epsilon = 4 \pm i$, at a fixed frequency $\omega$ and gain temperature $T \to 0^\circ K$ (corresponding to complete inversion), as a function of radius $R$ (in units of the vacuum wavelength $c/\omega$). $\Phi(\omega)$ is normalized by the flux $|\Phi_{BB}(\omega)| = R^2(\omega/c)^2\hbar\omega$ from a “blackbody” of the same surface area and temperature. As expected, $\Phi$ exhibits peaks at selected $R \gtrsim (c/\omega)$ corresponding to enhanced emission at Mie res-
Figure 2.6: (a) SE flux $\Phi$ from $\mathcal{PT}$–symmetric spheres consisting of $N = \{1, 2, 3, 4\}$ (black, green, blue, and red lines) regions of equal gain (orange) and loss (gray), with permittivities $\epsilon = 4 \pm i$ and gain temperature $T \to 0^-\text{K}$ (corresponding to complete population inversion), as a function of sphere radius $R$ (for a fixed vacuum wavelength $2\pi c/\omega$). $\Phi(\omega)$ is normalized by the flux $|\Phi_{BB}(\omega)| = R^2(\omega/c)^2\hbar\omega$ from a “blackbody” of the same surface area and temperature. Insets show angular radiation intensities $\Phi(\theta, \phi)$ for different $N$ at selective radii $R = \{2.6, 3, 2.7, 2.1\}(c/\omega)$ (corresponding to increasing $N$), with orange/gray colormaps denoting regions of gain/loss. (b) Selected $\Phi(\theta, \phi)$ of various $\mathcal{PT}$–symmetric shapes, including “beach balls”, “magic” cubes, and cylinders, showing complex directivity patterns that depend strongly on the centro-symmetry of the objects. Here the color white (black) refers to the maximum (minimum) flux value. (c) Peak SE flux $\Phi$ (right axis, dashed lines) and directivity $\Phi_G/\Phi$ (left axis, solid lines), or the ratio of the flux emitted along gain direction $\Phi_G$ (see text) to the total flux $\Phi$, near the third resonance of the $N = 1$ sphere, as a function of the gain/loss tangent $\delta = |\text{Im}\epsilon|/\text{Re}\epsilon$ and for multiple values of $\text{Re}\epsilon = \{2, 4, 12\}$ (red, black, and blue lines), where the vertical dashed black line denotes $\delta$ for (a). The plots show increased directivity attained as the system approaches the lasing threshold (marked in the case of $\text{Re}\epsilon = 12$, where $\Phi$ diverges ).
respectively (insets). In particular, the ratio $\Phi_G/\Phi$ of the flux emitted from the gain surfaces,

$$\Phi_G = \sum_{k=0}^{N-1} \int d\theta \sin^2(\theta) \int_{\pi(2k+1)/N}^{\pi} d\phi \Phi(\theta, \phi),$$

to the total flux $\Phi$, is generally $\lesssim 0.5$ for odd $N$ (centrosymmetric) and $\gtrsim 0.5$ otherwise. For instance, $N = 1$ spheres exhibit $\Phi_G/\Phi \approx 0.1$ at $R \approx 3.3(c/\omega)$ whereas $N = 2$ spheres exhibit $\Phi_G/\Phi \approx 0.75$ at $R \approx 3(c/\omega)$. The sensitive dependence of emission pattern on geometry and gain profile is not unique to spherical structures, as illustrated in Fig. 2.6(b), which shows $\Phi(\theta, \phi)$ for various shapes, including “magic” cubes, “beach balls”, and cylinders—as before, the presence/absence of centro-symmetry results in high/low gain directivity.

To understand the features and origin of these emission patterns, Fig. 2.6(c) explores the dependence of the peak $\Phi$ (dashed lines) and $\Phi_G/\Phi$ (solid lines) on the gain/loss tangent $\delta = |\text{Im } \epsilon|/\text{Re } \epsilon$ of the $N = 1$ sphere, near the third resonance and for multiple values of $\text{Re } \epsilon = \{2, 4, 12\}$. As shown, there is negligible ASE in the limit $\text{Im } \epsilon \to 0$, yet the localization of fluctuating dipoles to the gain-half of the (increasingly uniform) sphere leads to a small (though observable) amount of directionality, favoring emission toward the loss direction. The tendency of dipoles within a sphere to emit in a preferred direction has been studied in the context of fluorescence [280] in the ray-optical limit $R \gg c/\omega$, which as shown here is exacerbated in the presence of gain [88]: essentially, dipoles within a sphere tend to emit in the direction opposite the nearest surface, which explains why spheres having/lacking centro-symmetry tend to emit along directions of gain/loss. Moreover, in order to achieve large directivity, there needs to be a significant amount of mode confinement and gain, as illustrated by the negligible ASE and directivity of the $\text{Re } \epsilon = 2$ sphere. Finally, we find that for large enough $\text{Re } \epsilon$, the directivity increases with increasing $\text{Im } \epsilon$, peaking at a critical $\delta$, corresponding to the onset of lasing. Such a transition is marked by a diverging $\Phi$ near the threshold along with a corresponding narrowing of the resonance linewidth.
(not shown). (Note that our predictions close to and above this critical gain are no longer accurate since they neglect important effects stemming from stimulated emission \textsuperscript{163}. For instance, at and above the critical gain, the resonance linewidth goes to zero and then broadens with increasing $\text{Im } \epsilon$, while it is well known that nonlinear gain saturation results in a finite laser linewidth \textsuperscript{73} \textsuperscript{74}) Nevertheless, our results demonstrate that a significant amount of directivity can be obtained below the onset of lasing, where the linear approximation is still valid.

### 2.3.5 Gain inhomogeneity arising from pumping

Figure 2.7: (a) SE flux $\Phi$ and (b) gain directivity $\Phi_G/\Phi$, or the ratio of the flux emitted along the gain directions $\Phi_G$ (see text) to the total flux $\Phi$, from the $N = 2$ sphere (inset) of Fig. 2.6 at frequency $\omega_{21}$, corresponding to the transition frequency of an active region consisting of Rhodamine 800 dye molecules. The gain medium is excited by planewaves propagating in opposite directions, $x \cos \phi + y \sin \phi$ and $-x \cos \phi - y \sin \phi$, for three different orientations, $\phi = 0$ (black line), $\pi/4$ (blue line), and $-\pi/4$ (red line), leading to significant spatial inhomogeneities. $\Phi$ is normalized by $\Phi_{BB}$ as in Fig. 2.6 and plotted as a function of radius $R$, in units of the dye $2 \to 1$ transition wavelength $c/\omega_{21}$ (see text). Insets in (a) show $z = 0$ cross-sections of the resulting gain profiles $-\text{Im } \epsilon_g$ at $R = 3.4(c/\omega_{21})$ while those in (b) show angular radiation intensities $\Phi(\theta, \phi)$ normalized by the maximum intensity $\Phi_{\text{max}}$ at $R = 4(c/\omega_{21})$ for the different pump orientations.

Next, we employ the 4-level gain model of Eq. 2.6 to illustrate the impact of gain inhomogeneities on ASE. While both $\mathcal{P}$ and PE are simultaneous sources of spatial
dispersion, for the sake of comparison we consider each independently of one another. We begin by studying the impact of pump on the $N = 2$ sphere (above) for an active region consisting of a background medium $\epsilon_r = 4$ that is doped with Rhodamine 800 dye molecules with atomic parameters: $\omega_{21} = 2.65 \times 10^{15} \text{s}^{-1}$ ($\lambda \approx 711 \text{ nm}$), $\gamma_1/\omega_{21} = 0.04$, $\gamma_{21}/\omega_{21} = 7.5 \times 10^{-7}$, $\gamma_{32}/\omega_{21} = \gamma_{10}/\omega_{21} = 10^{-3}$, QY of 20%, and concentration $n = 40 \text{ mM}$ ($2.4 \times 10^{19} \text{ cm}^{-3}$) [65]. Note that since $\omega_{30} \gg \omega_{21}$, it is safe to neglect feedback due to PE and hence $P$ is determined from a single scattering calculation [357]. For these parameters [65], a pump rate $P/\gamma_{21} \approx 3$ results in $-\text{Im} \epsilon_g(\omega_{21}) \approx 1$. We consider illumination with $z$-polarized planewaves incident from two opposite directions along the $x$-$y$ plane, shown schematically in Fig. 2.7. The insets depict $z = 0$ cross-sections of the resulting $\epsilon_g$ profile [Fig. 2.7(a)] at $R = 3.4(c/\omega_{21})$ along with emission patterns $\Phi(\theta, \phi)$ [Fig. 2.7(b)] at $R = 4(c/\omega_{21})$, under three incident conditions, corresponding to different directions of incidence, $\hat{x} \cos \phi + \hat{y} \sin \phi$ and $-\hat{x} \cos \phi - \hat{y} \sin \phi$, with $\phi = \{0, \pm \pi/4\}$; in each case, the incident power is chosen such that $\max\{-\text{Im}[\epsilon_g]\} = 1$. As shown, the gain profiles vary dramatically with respect to position and incident angle, with $\text{Im} \epsilon_g$ changing from $0 \rightarrow 1$ on the scale of the wavelength. These spatial variations lead to correspondingly large changes in the overall ASE [Fig. 2.7(a)] and directivity [Fig. 2.7(b)]. More importantly, we find that these features cannot be explained by naive, uniform-medium approximations (UMA). For instance, replacing $\epsilon_g$ with the average gain $\langle \epsilon_g \rangle = \frac{1}{V} \int_V \epsilon_g$ in the case $\phi = -\pi/4$, we find that UMA predicts an emission rate $\Phi/\Phi_{BB} \approx 30$ that is three times larger than that predicted by exact calculations. Differences in illumination angle also result in different angular radiation patterns $\Phi(\theta, \phi)$. For instance, we find that $\phi = -\pi/4$ leads to much more isotropic radiation than $\phi = \{0, \pi/4\}$, a consequence of the larger $\text{Im} \epsilon_g$ near the center of the sphere and the fact that dipoles near the center tend to radiate more isotropically and efficiently than those which are farther laying farther away.
2.3.6 Gain inhomogeneity arising from Purcell effect

We now consider inhomogeneities arising from PE, assuming a uniform pump and doping concentration. In particular, we apply the self-consistent framework described in Sec. 2.3.3 to study ASE from the $N = 1$ sphere above, but with an active region consisting of a background medium $\epsilon_r = 4$ that is doped with Er$^{3+}$ atoms [154, 72], with parameters: $\omega_{21} = 6.28 \times 10^{14} \text{s}^{-1} (\lambda \approx 2.8 \mu\text{m}), \gamma_{\perp}/\omega_{21} = 0.03, \gamma_{21}/\omega_{21} = 5 \times 10^{-5}$, $\gamma_{32}/\omega_{21} = \gamma_{10}/\omega_{21} = 1$, bare QY of 50%, and concentration $n = 10^{19} \text{cm}^{-3}$. (Further below we also consider a different geometry, a metal-dielectric spaser consisting of similar gain parameters but passive metallic regions.) Here, a pump rate $\mathcal{P}/\gamma_{21} = 10^{-4}$ results in $-\text{Im} \epsilon_g(\omega_{21}) \approx 1$ in the absence of PE. As discussed above, we employ fixed-point iteration to solve Eq. 2.5 and Eq. 2.8 and hence obtain consistent values of $\epsilon_g$ and $\Phi$, starting with the bare ($\mathcal{F} = 1$) atomic parameters and iterating until the gain parameters converge to the nearest fixed point. Generally, the convergence rate of the fixed-point algorithm depends sensitively on the chosen parameter regime, requiring larger number of iterations with decreasing $\frac{\partial \epsilon_g}{\partial \gamma_{21}}$ (decreasing local slope) [420]. The convergence also depends on the degree of nonlinearity in the system, which in the case of our 4-level system can be significant under small QY ($\gamma_{21}^*/\gamma_{21} \ll 1$), large $\mathcal{P}$, or $\gamma_{21}/\gamma_{10} \lesssim 1$ (in which case there is significant gain saturation). Nevertheless, in practice we find that for a wide range of parameters, a judicious combination of fixed-point iteration and Anderson acceleration [448] ensures convergence within dozens of iterations. The bottom/top insets of Fig. 2.8(c) demonstrate the iterative process at a fixed $R = 2(c/\omega_{21})$ and for two different sets of concentrations $n = \{1, 5\} \times 10^{19} \text{cm}^{-3}$ and quantum yields $\approx \{10, 50\}\%$.

Figure 2.8 illustrates the impact of PE on the emission of the sphere, showing variation in (a) SE flux $\Phi(\omega_{21})$ and (b) gain directivity $\Phi_g/\Phi$ of the sphere with respect to radius $R$ at a fixed $\mathcal{P}/\gamma_{21}^0 = 10^{-4}$, or with respect to (c) pump rate $\mathcal{P}$ at a fixed $R = 2(c/\omega_{21})$, both including (solid blue lines) and excluding (dashed black
Figure 2.8: Self-consistent treatment of Purcell effect (PE) based on solution of Eq. 2.6 and Eq. 2.8 using fixed-point iteration. (a) SE flux $\Phi$ normalized by $\Phi_{BB}$ as in Fig. 2.6 and (b) gain directivity $\Phi_G/\Phi$ from the $N = 1$ sphere of Fig. 2.6 as a function of radius $R$ (in units of $c/\omega_{21}$), with an active medium consisting of Er$^{3+}$ atoms and subject to a uniform pump rate $P/\gamma_{21}^0 = 10^{-4}$, where $\omega_{21}$ and $\gamma_{21}^0$ denote the bulk (bare) 2 $\rightarrow$ 1 transition frequency and decay rate of the atoms, respectively. Top/bottom insets in (a) show $z = 0$ cross-sections (contour plots) of the gain profiles $-\text{Im } \epsilon_g$ obtained during the first and final (fixed-point) iteration of the algorithm at two separate radii $R = \{2, 2.5\}(c/\omega_{21})$, with black/white denoting min, max $-\text{Im } \epsilon_g \approx \{1.9, 1.3\}$ and $\{1.9, 1.5\}$, respectively. Insets in (b) show the angular radiation intensities $\Phi(\theta)$ at selected radii (black dots), normalized by the maximum intensity $\Phi_{max}$. (c) $\Phi$ at a fixed radius $R = 2(c/\omega_{21})$ as a function of $P/\gamma_{21}^0$, with insets illustrating the evolution of $\Phi$ as a function of iteration at a fixed $P/\gamma_{21}^0 = 10^{-4}$ and for two different values of quantum yields (QY) [see text]. All plots compare quantities in the presence (solid blue lines) or absence (dashed black lines) of PE, and under a uniform–medium approximation (UMA) (solid red lines) described in the text; the solid green line in (a) denotes the average Purcell factor $\langle F \rangle = \frac{1}{V} \int_V F$ at each radius.

lines) PE. As before, $\Phi$ is normalized by $\Phi_{BB}$. Shown as insets in Fig. 2.8(a) are $z = 0$ cross sections of $-\text{Im } \epsilon_g$ for the first and final (fixed-point) iteration of the algorithm, at two different radii $R = \{2, 2.5\}(c/\omega_{21})$ (black dots), demonstrating large gain enhancement and spatial variations. As expected, $\Phi$ is either enhanced or suppressed depending on the average PE (green line) which we have defined as $\langle F \rangle \neq 1$, where for convenience we have defined:

$$\langle F \rangle = \frac{1}{V} \int_V F = \frac{1}{N} \sum_n \Phi_{bn}/\Phi^0 = -\frac{1}{2N\Phi^0} \text{Tr } W^* \text{ sym } GW.$$ 

(As discussed above, $\langle F \rangle \sim \Phi$ turns out to be the Frobenius norm of a low-rank matrix and is therefore susceptible to fast computations.) As shown, at small $R \lesssim$
In the absence of resonances, $\langle F \rangle < 1$ and hence $\Phi$ is suppressed with respect to the predictions of the bare. Conversely, $\langle F \rangle > 1$ near resonances and hence $\Phi$ is enhances. Note that for our choice of parameters, the gain profile scales linearly with the quantum yield, i.e. $\epsilon_g \propto \text{QY} = \gamma_{r21}^r/\gamma_{21}$, such that in the limit as $\langle F \rangle \to \infty$ (ignoring quenching occurring as $\gamma_{21}^r \to \gamma_{10}$), $-\text{Im} \epsilon_g \to 2$. (For smaller QY $\ll 1$, $\epsilon_g$ can be many times larger than the bare permittivity with increasing $\langle F \rangle$, saturating at much larger values of PE.) In addition to changing the overall SE rate, PE also modifies the sphere’s directivity. This is illustrated in Fig. 2.8(b), which shows enhancements in $\Phi_G/\Phi$ and correspondingly changes in emission patterns (insets) at selective $R = \{2, 2.5\}(c/\omega_{21})$.

Figure 2.8(c) also explores the dependence of $\Phi$ on $P$ at a fixed $R = 2(c/\omega_{21})$, showing that $\Phi$ peaks at a finite value of $P/\gamma_{12}^0 \gtrsim 10^{-4}$ and then decreases with increasing $P$; the same is true for $\langle F \rangle$ and $\langle \epsilon_g \rangle$ (not shown). Such a non-monotonicity stems from the fact that near the critical pump rate, $\text{Im} \epsilon_g \approx \text{Re} \epsilon_g$, causing the resonance frequency to shift to smaller radii, a trend that is observed both in the presence and absence of PE. Surprisingly, however, we find that while PE causes large inhomogeneities in $\epsilon_g$, in both scenarios the peak emission is approximately the same, suggesting the possibility that one could explain the impact of PE by a simple UMA. In what follows, we exploit a UMA that not only greatly simplifies the calculation of PE but also leads to accurate results over a wider range of parameters. In particular, we consider a UMA in which the otherwise inhomogeneous gain profile of the object is replaced with that of a uniform medium $\epsilon_g(x) \to \langle \epsilon_g \rangle$ (assuming a uniform pump rate) given by Eq. 2.6 but with $\gamma_{21}^r \to \langle \gamma_{21}^r \rangle = \langle F \rangle \gamma_{21}^r \gamma_{21}^0$, corresponding to a homogeneous broadening/narrowing of the gain atoms throughout the sphere. Within this approximation, the system of nonlinear equations above is described by a single (as opposed to $N \gg 1$) degree of freedom $\langle \gamma_{21}^r \rangle$, enabling faster convergence along with application of algorithms that are especially suited for handling low-dimensional
systems of equations [98]. Ignoring other sources of inhomogeneity (e.g. induced by density or pump variations), such an approximation allows calculation of $\Phi$ via scattering formulations best-suited for handling piecewise-constant dielectrics, including SIE [379] and related scattering matrix [46] methods. The solid red lines in Fig. 2.8 are obtained by employing the UMA, demonstrating its validity over a wide range of parameters. Surprisingly, we find that this holds even in regimes marked by strong gain saturation (e.g. $\gamma_{10} \lesssim \gamma_{21}$). It follows that in this geometry, the effect of PE on radiation can be attributed primarily to the presence of a larger average gain or pump rate in the sphere, whereas the actual spatial variation in $\epsilon_g$ is largely unimportant.

Figure 2.9: Self-consistent treatment of Purcell effect (PE) in the metal–dielectric composite shown on the inset, with orange and blue denoting Er$^{3+}$–doped polymer and Au regions, described via full solution of Eq. 2.6 and Eq. 2.8 (solid blue lines) or via the uniform-medium approximation (dashed red line) described in the text. The solid line shows the SE flux $\Phi$ from the particle as a function of radius $R$ (in units of $c/\omega_{21}$). Insets show $z = 0$ cross-sections of the gain profile $-\text{Im} \epsilon_g$ (top) and active temperature (right), along with the angular radiation intensity $\Phi(\theta, \phi)$ (bottom) at a selected $R = 0.95(c/\omega_{21})$, normalized by the maximum intensity $\Phi_{\text{max}}$.

There are geometries and situations where such a UMA is expected to fail, e.g. structures subject to even large dielectric inhomogeneities (as in Fig. 2.7). Such conditions arise in large objects (supporting higher-order resonances) or in metal–dielectric composites (supporting highly localized surface waves). Figure 2.9 shows $\Phi$ for one
such structure [bottom inset]: a dielectric sphere with the same gain medium (orange) of Fig. 2.8 but partitioned into three metallic (red) regions along the azimuthal direction, given by $\epsilon(\phi) = -2 + i$ for $\phi \in [2n\pi/3, 2n\pi/3 + \pi/8]$, where $n = 0, 1, 2$. (Note that our choice of $\epsilon$ for the metal does not lead to a strong plasmonic resonance, but still yields significant sub-wavelength confinement.) A $z = 0$ cross-section of the steady-state $\epsilon_g$ and temperature distributions at a fixed $R = 0.95c/\omega_{21}$ is shown on the top inset, demonstrating significant and complex variations in both quantities. In particular, while population inversion is suppressed near the metal (black regions), $\epsilon_g$ attains its maximum value (white regions) close to the metal surface, decaying rapidly within the dielectric. Comparing the exact (solid blue lines) and UMA (dashed red line) predictions, one finds that the presence of multiple nodes in $\epsilon_g$ leads to a dramatic failure for UMA (with a peak error of $\approx 50\%$). Despite the different radiation rates, however, we find that UMA effectively captures the main features of the far-field radiation pattern (inset).

2.3.7 Concluding Remarks

We have shown that wavelength-scale, active-composite bodies can lead to complex radiative effects, depending sensitively on the arrangement of gain and loss. By exploiting a general-purpose formulation of EM fluctuations, we quantified the non-negligible impact that dielectric and noise inhomogeneities can have on emission in these systems. Furthermore, we introduced an approach that captures feedback from Purcell effect (i.e., the optical environment) on the gain medium. We note that in situations where ASE is dominated by relatively few leaky resonances, it is possible and practical to perform a similar procedure by expanding the fields in terms of eigenmodes, in which case the problem boils down to solution of a linear generalized eigenvalue problem for the leaky modes. (In VIE as in FDFD or related brute-force methods, leaky modes can be computed via the solution of a generalized eigenvalue
problem of the form $Z(\omega)\xi = 0$, for a complex frequency $[296]$. However, the FVC approach above is advantageous in that it casts the problem in the context of solutions of relatively few ($\ll$ degrees of freedom) scattering calculations. More importantly, FVC can handle structures supporting many modes or situations where near-field effects are of interest and contribute to PE $[218]$. The latter is especially important when the relevant quantity is the energy exchange between two nearby objects, a regime that motivated initial development of these and related scattering methods $[274]$. Note that above we mainly explored structures with small $\text{Re} \epsilon \approx 4$ and large gain concentration $n$, leading to large $\text{Im} \epsilon_g \lesssim \text{Re} \epsilon$ even for relatively weakly confined resonances. However, similar effects can be obtained with smaller $n$ and $\text{Im} \epsilon_g$ in structures with larger $\text{Re} \epsilon$ and dimensions, or supporting highly localized fields (e.g. spacers), where there exist larger Purcell enhancement. Finally, we note that micron-scale particles like the spheres explored above lie within the reach of current experiments $[216]$. 


Chapter 3

Strongly coupled conduction and radiation at the nanoscale

In Sec. 2.2, we showed that thermal radiation patterns can be greatly modified by the presence of large temperature gradients arising from thermal conduction. One underlying assumption is that the temperature profile in nanoscale devices is largely governed by thermal conduction alone, while the feedback from thermal radiation can be ignored. Indeed, conductive heat transfer is typically orders of magnitude larger than radiation. However, when objects of different temperatures are held at sub-wavelength separations, radiative heat exchange can be enhanced by many orders of magnitude due to additional evanescent contributions [27, 444, 334, 367, 405]. In this chapter, we show that NFRHT can have a noticeable impact on conduction, and the interplay of these two effects can dramatically modify the temperature and thermal exchange rate of bodies at sub-micron separations [309, 305].

The organization of the chapter is as following. In Sec. 3.1, we briefly review the history of coupled photonic and phononic diffusion processes in nanostructures, and provide a generic formulation applicable to arbitrary geometries. The formulation is first applied to a simpler planar geometry of two parallel planar objects in Sec. 3.2.
where we derived a closed-form expression for the heat flux between plates subject to arbitrary temperature profiles. We demonstrate that temperature gradients induced by radiation can change the asymptotic behavior of RHT with respect to gap sizes. Moving on to arbitrarily shaped bodies in Sec. 3.3 where RHT is computed by extending the FVC formulation, larger (even nonlinear) temperature gradients are observed.

### 3.1 Overview

Coupled photonic and phononic diffusion processes in nanostructures are becoming increasingly important [212, 91]. While recent works have primarily focused on the interplay between thermal diffusion and external optical illumination, e.g. laser-induced, localized heating of plasmonic structures [19, 20, 291, 24, 47], the thermal radiation emitted by a heated body and absorbed by nearby objects can also be a great source of heating or cooling. To date, however, the impact of RHT on conduction remains largely unexplored, with the consensus view being that radiation is insufficiently large to result in appreciable temperature gradients [460, 459, 246]. On the other hand, modern experiments measuring RHT between planar surfaces are beginning to probe the nanometer regime [225, 188, 326, 383, 396, 232, 334, 434, 435, 233, 221, 412, 406, 227], and in certain cases offer evidence of deviations from the typical $1/d^2$ behavior associated with near-field enhancement [81, 319] at nanometric distances, often attributed to nonlocal [181, 212] or phonon-tunneling [91] effects. Here, we find that depending on geometric configuration and materials, radiation-induced temperature gradients can play a significant role on heat transport at relatively large separations $\sim$ tens of nanometers, requiring a full treatment of the CR interplay.
3.1.1 General formulation

Here we describe a general formulation of coupled CR applicable to arbitrary geometries. Consider a situation involving two bodies (the same framework can be extended to multiple bodies), subject to arbitrary temperature profiles and exchanging heat among one other, shown schematically in Fig. 3.1. Neglecting convection and considering bodies with lengthscales larger than their phonon mean-free path, in which case Fourier conduction is valid, the stationary temperature distribution satisfies:

$$\nabla \cdot \left[ \kappa(x) \nabla T(x) \right] + \int d^3x' \phi(x, x') = Q(x) \quad (3.1)$$

where $\kappa(x)$ and $Q(x)$ describe the bulk Fourier conductivity and presence of external heat sources at $x$, respectively, and $\phi(x, x')$ is the radiative power per unit volume from $x'$ to $x$.

Previous studies of Eq. 3.1 considered only radiative energy escaping into vacuum through the surfaces of the objects, exploiting simple, albeit inaccurate ray-optical approximations that are inapplicable for sub-wavelength objects or in the near field [100, 186]. The novelty of our approach to Eq. 3.1 is that $\phi(x, x')$, as written above and computed below, is fundamentally tied to accurate and modern descriptions of RHT based on macroscopic fluctuational EM [388, 27], allowing us to...
explore regimes (e.g. distances \( \ll \lambda T \)) where near-field effects dominate RHT among different objects. In particular, as we show below, in some regimes RHT can lead to observable temperature distributions. Although we only consider the impact of external radiation on the temperature profile and vice versa, under large temperature gradients, RHT could potentially modify the intrinsic thermal conductivity of these objects [83, 212, 415], a situation that we leave to future work. We also ignore far-field radiation since it is negligible compared to conduction or RHT at the distances considered in this work.

### 3.2 Planar objects

![Figure 3.2: Schematic of two parallel slabs separated by a distance \( d \) along the \( z \) direction (\( z = 0 \) denotes the middle of the gap). The two slabs exhibit a temperature profile \( T(z) \): the temperature is constant and has values \( T_L \) and \( T_R \) in the regions \( z \leq -d/2 - t_a \) and \( z \geq d/2 + t_b \), respectively, and variable in the regions \( -d/2 - t_a < z < -d/2 \) and \( d/2 < z < d/2 + t_b \), with \( T_a \) and \( T_b \) denoting the temperatures at the slab–vacuum interfaces.](image)

In this section, we study the coupling of conduction and radiation that can arise in the context of planar objects separated by nanoscale gaps. Consider a situation involving two planar bodies, labelled \( a \) and \( b \), illustrated in Fig. 3.2 subject to arbitrary temperature profiles and exchanging heat among one other, Eq. 3.1 can be
simplified to a one-dimensional coupled CR equation:

\[
\frac{\partial}{\partial z} \left[ \kappa(z) \frac{\partial T(z)}{\partial z} \right] + \int dz' \varphi(z, z') = Q(z), \quad (3.2)
\]

where \(\kappa(z)\) and \(Q(z)\) represent the bulk Fourier conductivity and external heat-flux rate at \(z\), respectively, while \(\varphi(z, z')\) denotes the radiative power per unit volume from a point \(z'\) to \(z\).

We focus on the scenario illustrated on the inset of Fig. 3.2 in which the temperature of slab \(a\) (\(b\)) slab is fixed at \(T_L\) (\(T_R\)) by means of a thermostat, except for a region of thickness \(t_a\) (\(t_b\)). To study this problem, we calculate the RHT via a Fourier expansion of the slabs’ scattering matrices [133]. Such techniques were recently employed to obtain close-formed analytical expressions of RHT between plates of uniform temperature [32, 306, 308, 307]. Here, we extend these results to consider the more general problem of slabs under arbitrary temperature distributions. Toward this aim, we generalize prior methods [306, 307] by dividing each slab in films of infinitesimal thicknesses, each having a fixed temperature. Describing the associated EM fields at each point by means of the fluctuation-dissipation theorem (from which the RHT can be deduced), and considering multiple reflections associated with the various interfaces, we find that the evanescent RHT per unit volume from a point \(z_a\) in slab \(a\) to a point \(z_b\) in slab \(b\), \(\varphi(z_b, z_a) = \int_0^\infty d\omega \int_{\omega/c}^\infty d\beta H_a(\omega, \beta; z_b, z_a)\), can be expressed analytically in the closed form:

\[
\varphi(\omega, \beta; z_b, z_a) = \frac{4\beta}{\pi^2} (r''k''_{zm})^2 \frac{e^{-2k''_{zm}d}}{|1 - r^2e^{-2k''_zd}|^2} \times e^{-2k''_{zm}(z_b - z_a - d)} \left( N[\omega, T(z_a)] - N[\omega, T(z_b)] \right), \quad (3.3)
\]

where \(\beta\) denotes the conserved, parallel \((x-y)\) wavevector \(k_z = \sqrt{\omega^2/c^2 - \beta^2}\) and \(k_{zm} = \sqrt{\varepsilon\omega^2/c^2 - \beta^2}\) the perpendicular wavevectors in vacuum and the interior of the slabs, respectively, and we also introduced the transverse-magnetic (dominant)
polarization Fresnel reflection coefficient of a planar ε–vacuum interface, \( r = (\varepsilon k_z - k_{zm})/(\varepsilon k_z + k_{zm}) \). Note that we restricted our attention to the transverse-magnetic polarization, since it supports a surface phonon-polariton resonance. Equation 3.3 permits fast solutions of coupled CR problems in this geometry for a wide range of parameters and materials.

### 3.2.1 Surface-sink approximation

![Figure 3.3: Temperature profile along the temperature-varying region of the hot slab in a configuration shown in Fig. 3.2 where the two silica slabs with \( t_a = t_b = 100 \mu m \) are held at external temperatures \( T_L = 600 \text{ K} \) and \( T_R = 300 \text{ K} \) at the outer ends. The points \( z = 0, 100 \mu m \) represent the boundary of the thermostat and the interface with the vacuum gap. The three lines correspond to \( d = 10 \text{ nm} \) (black), 20 nm (red) and 50 nm (blue). The inset shows the \( z \)-dependent radiative flux-rate \( \varphi(z_a) \) (see text) along slab \( a \) at \( d = 100 \text{ nm} \).](image)

Figure 3.3: Temperature profile along the temperature-varying region of the hot slab in a configuration shown in Fig. 3.2 where the two silica slabs with \( t_a = t_b = 100 \mu m \) are held at external temperatures \( T_L = 600 \text{ K} \) and \( T_R = 300 \text{ K} \) at the outer ends. The points \( z = 0, 100 \mu m \) represent the boundary of the thermostat and the interface with the vacuum gap. The three lines correspond to \( d = 10 \text{ nm} \) (black), 20 nm (red) and 50 nm (blue). The inset shows the \( z \)-dependent radiative flux-rate \( \varphi(z_a) \) (see text) along slab \( a \) at \( d = 100 \text{ nm} \).

At small separations, near-field RHT is dominated by large-\( \beta \) surface modes that are exponentially confined to the slab–vacuum interfaces [213]. Hence, it is sufficient (as confirmed below, see Fig. 3.3) to treat its impact on conduction as a purely surface effect, in which case its influence can be described purely in terms of the surface temperatures. In particular, under this assumption, given a distance \( d \), identical conductivities \( \kappa \) and temperature-varying regions \( t = t_a = t_b \), and external temperatures \( T_L \) and \( T_R \), the only unknowns are the interface temperatures \( T_a \) and \( T_b \), which satisfy...
the following boundary conditions:

\[
- \kappa \frac{T_a - T_L}{t} = -\kappa \frac{T_R - T_b}{t} = \varphi, \tag{3.4}
\]

where \( \varphi \) denotes the net heat exchanged between the two slabs. This can be obtained using the same approach discussed above and its evanescent component at frequency \( \omega \) and wavevector \( \beta \) reads,

\[
\varphi(\omega, \beta) = \frac{2\beta}{\pi^2} (r'')^2 k'' e^{-2k''d} \int_0^{+\infty} dz e^{-2k''d} \left( N[\omega, T(-d/2 - z)] - N[\omega, T(d/2 + z)] \right).
\tag{3.5}
\]

Despite the complex dependence of the heat flux on separation and temperature profile, we find that it is possible to approximate the former using a simple, power-law expression of the form, \( \varphi \sim h_0(T_a - T_b)/d^2 \) (valid as long as the radiation is primarily coming from the surface of the slabs), with the coefficient \( h_0 \) calculated as the near-field heat flux between two uniform-temperature slabs held at \( T_L \) and \( T_R \) divided by \( T_L - T_R \). Essentially, while the dependence of RHT on absolute temperature is generally nonlinear, the fact that conduction through the interior of the slabs scales linearly with \( T_a - T_b \) in conjunction with energy conservation (i.e. changes in conductive transfer must be balanced by corresponding changes in RHT), implies that \( \varphi \) must also scale linearly with \( T_a - T_b \), with the precise value of the coefficient \( h_0 \) determined from the radiative conductivity at values of where radiation cannot impact conduction. Given these simplifications, Eq. 4.4 can be solved to yield:

\[
\frac{T_a - T_b}{T_L - T_R} = \left( 1 + \frac{2th_0}{\kappa d^2} \right)^{-1}, \quad \frac{\varphi}{T_L - T_R} = \frac{h_0}{d^2} \left( \frac{T_a - T_b}{T_L - T_R} \right).
\tag{3.6}
\]

These formulas reveal that the interplay of conduction and radiation causes \( T_a - T_b \rightarrow 0 \) quadratically with \( d \), producing a continuous temperature profile and leading to a
finite value of \( \varphi \to \kappa (T_L - T_R)/2t \) as \( d \to 0 \), asymptoting to the expected value of conductive flux through a gap-less slab of thickness 2t subject to a linear temperature gradient \( T_L - T_R \) (as it must, from energy conservation). Below, we show that the existence of such temperature gradients along with deviations from the typical 1/d^2 RHT power law are within the reach of present experimental detection.

### 3.2.2 Numerical predictions

To begin with, we first address the validity of the surface–sink approximation above. In order to do so, we of course need to consider the full coupled CR problem described by Eq. 3.2, requiring numerical evaluation of the spatial heat transfer \( \varphi(z_a, z_b) \) in Eq. 3.3. For concreteness, we consider a practical situation typical of RHT experiments, involving two silica (SiO\(_2\)) slabs subject to external temperatures \( T_L = 600 \) K and \( T_R = 300 \) K by a thermostat at distance \( t = t_a = t_b = 100 \) \( \mu \)m away from the slab–vacuum interfaces. Silica not only has relatively low \( \kappa \approx 1.4 \) W/m-K but also supports polaritonic resonances at mid-infrared wavelengths and has well-tabulated optical properties [336]. Figure 3.3 illustrates the increasing, linear temperature gradient present in slab \( a \) with decreasing separations \( d \), a consequence of the exponential decay of the spatial heat transfer, \( \varphi(z_a) = \int dz_b \varphi(z_a, z_b) \), illustrated on the inset at a fixed \( d = 100 \) nm. Results obtained through Eq. 3.6 with \( h_0 = 5.53 \times 10^{-12} \) W/K, are in almost perfect (essentially indistinguishable) agreement with those of the full CR treatment and are therefore not shown. The same is true at smaller values of \( t \), down to tens of nanometers, below which the surface–sink approximation begins to fail.

Figure 3.4 shows \( \varphi \) and \( T_a - T_b \) (inset), normalized by the external temperature difference \( T_L - T_R \), as a function of \( d \) and for the same slab configuration but considering multiple \( t = \{0, 0.1, 1, 10, 100, 500\} \) \( \mu \)m, with decreasing values of \( t \) leading to smaller temperature gradients and larger \( \varphi \). Here, \( t = 0 \) (dashed line) corre-
responds to the typical scenario where conduction dominates and hence there are no
temperature gradients, in which case $\varphi = h_0(T_L - T_R)/d^2$ exhibits the expected di-
vergence. Quite interestingly, we find that at typical values of $t = 100 \mu$m, the flux
decreases by $\approx 50\%$ at distances $d \approx 30 \text{ nm}$, well within the reach of current exper-
iments [225, 396, 227, 406, 221]. In particular, this result may be relevant to recent
experiments [396] investigating RHT between large silica objects over distances rang-
ing from tens of nanometers to several microns, and which indicate deviations from
the $1/d^2$ scaling behavior (along with flux saturation) at short distances.

![Figure 3.4: Total flux $\varphi$ and temperature difference $T_a - T_b$ (inset) as a function of
distance $d$ between two silica slabs (shown schematically in Fig. 3.2) that are being
held at $T_L = 600 \text{ K}$ and $T_R = 300 \text{ K}$. The various solid lines correspond to different
temperature-varying regions $t$ (from top to bottom): 100 nm (black), 1 $\mu$m (red),
10 $\mu$m (brown), 100 $\mu$m (blue) and 500 $\mu$m (green). The orange dashed line shows $\varphi$
in the absence of temperature gradients.](image)

Motivated by the promising results for silica slabs, we now explore the degree
to which these saturation effects depend on the choice of material and operating
conditions, using as a useful figure of merit the separation regime at which they be-
come significant. In particular, inspection of Eq. 3.6 allows us to define the distance
$\tilde{d} = \sqrt{2t_h_0/\kappa}$ at which $T_a - T_b = \frac{1}{2}(T_L - T_R)$ and $\varphi = \frac{1}{2}h_0(T_L - T_R)/\tilde{d}^2$, correspond-
ing to half the value of the RHT obtained when CR coupling is ignored. Figure 3.5 shows \( \tilde{d} \) as a function of the material-dependent ratio \( h_0/\kappa \) for the particular choice of \( T_L = 600 \text{ K}, T_R = 300 \text{ K}, \) and \( t = 100 \mu\text{m}, \) highlighting the square-root dependence of the former on the latter. Superimposed are the expected \( \tilde{d} \) associated with various materials of possible experimental interest (solid circles), obtained by employing appropriate values of \( \kappa \) and \( h_0, \) which depend primarily on the choice of external temperature. Within the surface–sink approximation (valid here), the latter do not influence the scaling of either \( \varphi \) or \( T_a - T_b \) with respect to separation, as evident from Eq. 3.6. The inset of Fig. 3.5 shows \( h_0 \) as a function \( T_L \) for SiC, SiO\(_2\), and AZO (aluminium zinc oxide), identified by their increasing values of \( h_0, \) respectively, illustrating the relatively constant character of the coefficient over a wide range of acceptable temperature differences. (Note that we consider unrealistically large values of \( T_L \) only to illustrate the onset of asymptotic behavior.)

Noticeably, despite relatively small differences in the value of \( h_0 \) between various materials, there are striking variations in \( \tilde{d}, \) which can range anywhere from a few nanometers in the case of SiC and GaAs, up to several tens of nanometers for SiO\(_2\) and AZO, respectively. Such variations are almost entirely due to differences in thermal conductivities, which naturally play a major role in this problem, with the conductivities of SiC, SiO\(_2\), and AZO taken to be \( \kappa \simeq 120 \text{ W/m-K}, 1.4 \text{ W/m-K}, \) and 1.2 W/m-K, respectively. Note that the relatively low value of \( \kappa = 1.2 \) used for AZO has been realized \[284\] by way of novel techniques used to engineer the conductive properties of such composite materials \[211\], which have in fact been applied to realize ultra-low \( \kappa \lesssim 0.05 \text{ W/m-K} \) \[92, 63\]. The open circle in Fig. 3.5 indicates the expected \( \tilde{d} \) associated with such a potentially structured, ultra-insulating AZO, which takes on hundreds of nanometers. We stress that our predictions support the lack of gradients or deviations observed in recent experiments involving materials such as silicon and gold, which exhibit low and high values of \( h_0 \) and \( \kappa, \) respectively. As noted above,
Figure 3.5: Typical distance scale $\tilde{d}$ relevant to conduction–radiation problems (see text) as a function of the material-dependent ratio $h_0/\kappa$ for two slabs with $t_a = t_b = 100 \mu$m. Solid circles denote corresponding values for specific material choices (abbreviations), with AZO[1.2] and AZO[0.05] denoting aluminum zinc oxides of different conductivities $\kappa = 1.2$ W/m-K [284] and potential $\kappa = 0.05$ W/m-K [92], respectively. The inset shows the dependence of the radiative-heat transfer coefficient $h_0$ on the external temperature gradient $\Delta T = T_L - 300$ K, for three cases, AZO (red), silica (blue) and SiC (black), with decreasing values.

However, the case of silica is particularly interesting since it is typically used in RHT experiments, yet the possibility of temperature gradients has never been considered. These results along with Eq. 3.6 can serve as a reference for future experiments, allowing estimates of the regimes under which these effects become relevant.

While our analysis above is based on the assumption of vacuum gaps, it is straightforward to generalize Eq. 3.6 to include the possibility of finite intervening conductivities, $\kappa_0 > 0$, requiring only that $h_0$ be replaced with $h_0 + \kappa_0 d$ in the first expression of Eq. 3.6. We find, however, that similar conclusions follow for small but finite $\kappa_0 \lesssim 10^{-5}$ W/m-K (typical of RHT experiments).
3.2.3 Concluding Remarks

We have presented a study of coupled conduction-radiation heat transfer between planar objects at short distances. We have expressed the resulting temperature gradients and radiative-flux modifications in terms of simple, analytical expressions involving geometric and material parameters, showing that in systems well within experimental reach or already considered in experiments [396], both temperature gradients and flux saturation should be observed. A similar saturation phenomenon has been predicted to occur due to nonlocal damping [212, 181] and/or phonon tunneling below the nanometer scale [91] (note that at atomistic scales where continuum electrodynamics fails, the boundary between phonon and radiative conduction is blurred). Our work suggests that even at and above nanometer gaps, and depending on material and geometric conditions, CR interplay could instead become the dominant mechanism limiting RHT. Furthermore, there are significant efforts underway aimed at exploring regimes, e.g., smaller gap sizes or materials and structures leading to larger RHT (for applications in nanoscale cooling [166] and other thermal devices [31]), where these effects may be observed at even larger separations. Moreover, a material implying a thin film of high conductivity polaritonic material (e.g. SiC) on top of a low conductivity slab could be also a promising candidate for the observation of the effect at large separations. Arguably, advances in either or both directions will make such analyses necessary.

3.3 Generic shapes: grating structures

In this section, we extend the results of Sec. 3.2 to objects of generic shapes. Structured materials can further enhance [313, 269, 350, 169, 43, 101] and tune [382, 380, 120, 378] RHT, indicating that coupled conduction–radiation (CR) phenomena are bound to play a larger role in structured media.
3.3.1 Extending fluctuating–volume current formulation

Our ability to compute \( \varphi(x, x') \) in full generality hinges on an extension of the FVC method introduced in Sec. 2.1 which exploits powerful EM scattering techniques \[357\] to enable fast calculations of RHT under arbitrary geometries and temperature distributions. Previously in Ref. \[357\], we exploited this formalism to efficiently compute the total heat transfer. The solution of Eq. 3.1 requires an extension of the FVC method to include the spatially resolved heat transfer between any two voxels, which we describe below.

Following the notation defined in Sec. 2.1, here we start by considering a fluctuating current-source \( \sigma^a \) located at \( x_a = b_\alpha \) in body \( a \); such a source induces polarization–currents and EM fields in both objects, with the expression for heat flux absorbed at \( x_b = b_\beta \) given by:

\[
\Phi(\omega; x_a \rightarrow x_b) = \frac{1}{2} \langle \text{Re} (\xi^*_\beta \phi_\beta) \rangle
\]

where the subscript \( \beta \) denotes the amplitude of the corresponding quantity at \( b_\beta \), and “\( \langle \ldots \rangle \)” the thermodynamic ensemble average. Expressing the polarization–currents and fields in the localized basis \( \{ b_n(x) \} \), in which the current source \( \sigma^a(x) = \sum_n s^\alpha b_n(x) \) involves a sparse vector \( s^\alpha \) that is only nonzero at the \( \alpha \)-th term, and the induced currents are decomposed as \( \xi(x) = \sum_n x_n b_n(x) \). Exploiting the volume equivalence principle to express the field as a convolution of the incident and induced currents with the vacuum Green’s function (GF), \( \phi = \Gamma * (\xi + \sigma) \), one finds that
Eq. 3.7 can be expressed compactly in terms of the VIE matrices:

\[
\Phi(\omega; x_a \rightarrow x_b) = \frac{1}{2} \langle \text{Re} \left\{ x^*_\beta [G(x + s^\alpha)]_\beta \right\} \rangle 
\]

\[
= \frac{1}{2} \langle \text{Re} \left\{ (x + s^\alpha)_\beta^* [G(x + s^\alpha)]_\beta \right\} \rangle 
\]

\[
= \frac{1}{2} \langle \text{Re} \left\{ (W s^\alpha)_\beta^* (GW s^\alpha)_\beta \right\} \rangle 
\]

\[
= \frac{1}{2} \text{Re} \left[ C_{\alpha,\alpha} W_{\alpha,\beta}^* (GW)_{\beta,\alpha} \right], 
\]

where the matrices \( C, G \) and \( W \) are already specified in Sec. 2.1. Step 3.8 follows from the fact that \( s^\alpha \) is only non-zero inside body \( a \), and step 3.9 makes use of the VIE equation \( W s = x + s \) [357] for a given dipole distribution profile \( s \).

As shown in the coupled conduction and radiation equation Eq. 3.1 to solve for temperature profiles in body \( a \), the relevant quantity is \( \Phi(\omega, x) \), requiring integration over the volume of body \( b \). Consequently, it is more convenient to write Eq. 3.10 in terms of subspaces spanned by each object, where the matrices and vectors will be expanded in the block form. Taking the matrix \( C \) and the vector \( x \) as examples, the expansion is,

\[
C = \begin{pmatrix} C^{aa} & C^{ab} \\ C^{ba} & C^{bb} \end{pmatrix} ; \quad x = \begin{pmatrix} x^a \\ x^b \end{pmatrix} 
\]

where the superscripts \( a \) and \( b \) denote blocks associated with bodies \( a \) and \( b \), respectively, and the diagonal and off-diagonal components correspond to self- and intra-body interactions among the objects. Straightforward algebraic manipulation allows the total emission \( \Phi_e \) from point \( b_\alpha \) in body \( a \) to body \( b \), absorption \( \Phi_a \) from body \( b \) to point \( b_\alpha \) in body \( a \), and net power exchange between point \( b_\alpha \) and body \( b \)
to be written as:

$$\Phi_a(\omega; x_a) = \frac{1}{2} \text{Re} \left[ G^{ab} W_{bb}^\dagger D^{bb} \right],$$

$$\Phi_b(\omega; x_b) = -\frac{1}{2} \left[ \text{Re}(D^{aa} W_{ba}^\dagger G^{ba} W_{aa}) \right] + \frac{1}{2} \left[ \text{Re}(G^{aa} W_{ab}^\dagger D_{bb} W_{bb} \dagger) \right],$$

where $\text{sym}(X) = \frac{1}{2}(X + X^\dagger)$ denotes the symmetric part of $X$. Equation \ref{eq:3.13} describing emission, can be evaluated efficiently because the matrices $G^{ba}$ and $\text{sym} G^{bb}$ are both low rank ($r \ll N$) \cite{357}, in which case they can be SVD factorized to allow fast matrix multiplications. Unfortunately, the second term of Eq. \ref{eq:3.12} involves both the symmetric and anti-symmetric parts of $G^{aa}$, the latter of which is full rank. More conveniently, detailed balance dictates that $\Phi(\omega; x_b \rightarrow x_a) = \Phi(\omega; x_a \rightarrow x_b)$ whenever $T(x_a) = T(x_b)$, which implies that

$$\text{Re} \left[ M_{\beta,\beta} W_{\beta,\alpha}^\dagger (GW)_{\alpha,\beta} \right] = \text{Re} \left[ M_{\alpha,\alpha} W_{\alpha,\beta}^\dagger (GW)_{\beta,\alpha} \right],$$

where $M_{\alpha,\alpha} = \text{Im} \varepsilon(x_\alpha, \omega)$ is a real, diagonal matrix encoding the dissipative properties of the bodies, leading to the modified expression for the absorption rate,

$$\Phi_a(\omega; x_a) = \frac{1}{2} \left[ \text{Re}(M^{aa} W_{ba}^\dagger K^{bb} G^{ba} W_{aa}) \right] + M^{aa} W_{ba}^\dagger \text{sym}(K^{bb} G^{bb}) W_{ba} \right]_{\alpha,\alpha},$$

where $K_{\alpha,\alpha} = D_{\alpha,\alpha} / M_{\alpha,\alpha}$ is also a real, diagonal matrices encoding the Planck function. By inspection, one finds that the symmetrized operator in the second term is full rank except whenever the temperature of object $b$ is close to uniform, in which case $K^{bb}$ is nearly diagonal and hence $\text{sym}(K^{bb} G^{bb}) = K^{bb} \text{sym} G^{bb} + (G^{bb} K^{bb} - K^{bb} G^{bb} W_{ba} \dagger) / 2 \approx K^{bb} \text{sym} G^{bb}$. While solution of Eq. \ref{eq:3.14} is in principle feasible, it becomes prohibitive when the number of degrees of freedom is large. It therefore remains an open problem to establish a formulation allowing fast evaluations of the
spatially resolved absorbed power under arbitrary temperature distributions. The net heat exchange between point \( x_a \) and body \( b \) is then straightforward,

\[
\Phi(\omega, x_a) = \Phi_a(\omega; x_a) - \Phi_e(\omega, x_a)
\]  

(3.15)

Given Eq. 3.15, one can solve the coupled CR equation in any number of ways [361]. Here, we exploit a fixed-point iteration procedure based on repeated and independent evaluations of Eq. 3.10 and Eq. 3.1 converging once both quantities approach a set of self-consistent steady-state values. Equation Eq. 3.1 is solved via a commercial, finite-element heat solver whereas Eq. 3.10 is solved through a free, in-house implementation of our FVC method [357]. While the above formulation is general, for computational convenience we consider situations in which object \( b \) is kept at a constant, uniform temperature by means of a thermal reservoir, such that the absorbed power in object \( a \) can be computed efficiently via Eq. 3.14. The power emitted by \( a \) (the heated object), turns out to be much more convenient to compute, since the time-consuming part of the scattering calculation can be precomputed independently from the temperature distribution and stored for repeated and subsequent evaluations of Eq. 3.1 under different temperature profiles.

### 3.3.2 Coupled conduction and radiation in nanorods

As a proof of principle, we consider CR effects in a simple geometry consisting of two metallic nanorods of cross-sectional widths \( l \) and thickness \( t \); in practice, to obtain even larger RHT [412], such a structure could be realized as a lattice or grating on a substrate, shown schematically in Fig. 3.1. However, for computational convenience, we restrict calculations to a two-body configuration involving two nanorods separated by a gap, in which only one of the nanorods experiences a large temperature gradient while the other is held at 300K, shown in Fig. 3.6(b). Such a situation, while
highly restrictive, could for instance be realized by requiring large grating periods (in which case multiple-scattering can be safely ignored), low-index substrates (in which case the nanorods are effectively suspended in air), and that one of the nanorods has much higher effective conductivity than the other (e.g. through nanostructuring [211]). The strongest CR effects generally will arise in materials that exhibit large RHT, e.g. supporting surface–plasmon polaritons (SPP), and low thermal conductivities, including silica, sapphire, and AZO, whose typical thermal conductivities $\sim 1 \text{ W/m-K}$ [284, 321]. In the following, we take AZO as an illustrative example and assume a temperature-independent dielectric constant to illustrate the main effects stemming from CR coupling, leaving a full description, which is more relevant in the presence of large temperature gradients, to future work. Note that we recently considered the full temperature-dependent dielectric response in the context of far-field emission [208, 48], which can also be handled by the FVC framework. To begin with, we show that even in the absence of CR interplay, the RHT spectrum and spatial
distribution inside the nanorods differs significantly from those of AZO slabs of the same thickness.

As an illustrative example, Fig. 3.6(a) shows the RHT spectrum $\Phi(\omega)$ per unit area $A = l^2$ between two AZO nanorods (with doping concentration 11wt% [321]) of length $t = 500$ nm and varying widths $l = \{10, 20, \infty\}$ nm (blue solid, red solid, and black dashed lines), held at a large temperature difference $T_a(b) = 800(300)$ K and small vacuum gap $d = 20$ nm to thermally excite all resonant modes. Our assumptions here can soon fall within the scope of experimental realizations, as currently a temperature difference of $\sim 10^2$ K over nanometer gaps has already been demonstrated [225, 150]. The limit $l \to \infty$ of a slab–slab geometry was recently studied in Ref. [309], with $\Phi(\omega)$ exhibiting a single peak at the SPP frequency $\approx 3 \times 10^{14}$ rad/s. The finite size of the nanorods leads to additional peaks at lower frequencies, corresponding to bulk plasmon resonances (red and blue solid lines) that provide additional channels of heat exchange, albeit at the expense of weaker SPP peaks, leading to a roughly 5-fold enhancement in RHT compared to slabs. More importantly and well known, such structured antennas provide a mechanism to tune plasmon resonances at near- and far-infrared wavelengths (much lower than many planar materials), which can result in more effective heat exchange. The contour plots in Fig. 3.6(b) show the spatial RHT distribution $\Phi_e(\omega, x)$ (in arbitrary units) at three separate frequencies $\omega = \{0.4, 0.8, 2.3\} \times 10^{14}$ rad/s, corresponding to the first, second, and SPP resonances, respectively. As expected, the largest-frequency resonance is primarily confined to the corners of the nanorods (approaching the slab SPP resonance as $l \to \infty$), with the lowest-order resonances contributing flux primarily from within the bulk. Such enhancements not only result in larger temperature gradients but also qualitatively change the resulting temperature distributions.

To begin with, we consider a situation in which the boundary $I$ of nanorod $a$ is kept at $T_1 = 800$ K while the entire nanorod $b$ is held at $T_b = 300$ K. Figure 3.7(a) shows
Figure 3.7: (a) Temperature profile along the $z$ coordinate of a nanorod (solid lines) when it is heated from one side to a temperature of 800 K, and is separated from an identical, constant- and uniform-temperature nanrod held at $T = 300$ K on the other side, by a gap size $d = 20$ nm. The nanorods have cross-sectional width $l = 20$ nm and thicknesses $t = 500$ nm, and are made up of AZO with results shown for multiple values of the doping concentration $\{2, 6, 11\}$ wt% (blue, red, and black lines). Also shown are the temperature profiles of slabs (dashed lines) of the same thickness (corresponding to the limit $l \to \infty$). (Inset:) Temperature distribution throughout the nanorod in the case of 11 wt%. (b) Temperature profiles of nanorods of width $l = 10$ nm under various separations $d = \{5, 10, 20, 30\}$ nm (black, blue, red, and green lines). (c) The flux value of radiative heat transfer for nanorods (red) and slabs (blue) in the presence (solid lines) or absence (dashed lines) of the temperature gradients induced by the interplay of conduction and radiation. (Inset:) The temperature $T_I$ at the interface I of nanorod $a$ as a function of heat pumping rate at I, under various separations $d = \{5, 10, 20, 30\}$ nm (black, blue, red, and green lines).

the temperature profile along the $z$ direction for the $l = d = 20$ nm nanorod geometry of Fig. 3.6(a), obtained by solving Eq. 3.1 under $Q = 0$. For the sake of generality, we show results under various doping concentrations $\{2, 6, 11\}$ wt% (green, red, and black solid lines), resulting primarily in different SPP frequencies [321], and assume an AZO thermal conductivity of $\kappa = 1$ W/m·K [284]. As illustrated for the particular case of 11 wt% concentration (inset), the temperature along the $x$–$y$ cross sections of the nanorods is nearly uniform (due to the faster diffusion rate along this dimension), allowing us to focus on variations along $z$. In all scenarios, the temperature gradient is significantly larger for nanorods (solid lines) than for slabs ($t \to \infty$, dashed lines), becoming an order of magnitude larger in the case of 6 wt% due to the larger SPP frequency, which lies closer to peak Planck wavelength near 800 K. Furthermore, while
slabs can only exhibit visible linear temperature profiles in situations involving the
largest possible RHT (since RHT here in planar media is dominated by the resonant
surface emission [27]), the bulk and de-localized nature of nanorod resonances lead
to nonlinear temperature gradients.

Figure 3.7(b) shows the temperature profile at various separations \( d = \{5, 10, 20, 30\} \) nm (black, blue, red, and green lines) and for nanorods of width
\( l = 10 \) nm and doping concentration 11wt%. The figure illustrates the sensitive
relationship between CR interplay and gap size. Notably, while RHT and therefore
temperature gradients are expected to increase as \( d \) decreases, the profile becomes
increasingly linear as the geometry approaches the slab–slab configuration. The
transition from bulk- to surface-dominated RHT is also evident from Fig. 3.7(c),
which shows the flux rate \( \varphi \times d^2 \) as a function of \( d \) for slabs (black lines) of thickness
\( t = 500 \) nm and nanorods (red lines) of equal thickness and width \( l = 10 \) nm, in the
presence (solid lines) or absence (dashed lines) of CR interplay (with the latter in-
volving uniform temperatures). While RHT between bodies of uniform temperatures
scales as \( 1/d^2 \) (dashed lines), the temperature gradients induced by CR interplay in
the case of nanorods begins to change the expected powerlaw behavior at \( d \approx 20 \) nm;
the same occurs for slabs but at a shorter \( d \lesssim 10 \) nm. As shown in Ref. [309], in
the limit \( d \to 0 \), RHT will asymptote to a constant (not shown) rather than diverge.
Finally, we also consider the typical and more experimentally relevant scenario of a
fixed thermal source of flux rate \( Q \) dumping heat evenly at boundary I of nanorod \( a \)
and thus raising the temperature \( T_1 \) at the interface. Such a pumping mechanism can
be implemented through direct laser heating [18], Joule heating [259], or contact with
a high-temperature body described by a finite transfer coefficient [135]. The inset
of Fig. 3.7(c) shows \( T_1 \) as a function of \( Q \) at multiple \( d = \{5, 10, 20, 30\} \) nm (black,
blue, red, and green lines), illustrating an decreasing slope \( \partial T_1/\partial Q \) with decreasing
\( d \), with RHT making it harder to maintain higher temperatures. Nevertheless, we
find that the previous scenario of a high-temperature boundary $T_1 = 800$ K requires a moderate heat flux $\sim 10$ nW at $d = 20$ nm, achievable via recent nanoscale thermal transportation techniques \[63\].

### 3.3.3 Concluding Remarks

As experiments continue to push toward larger RHT by going to smaller vacuum gaps or by nanostructuring, accurate descriptions of CR interplay will become increasingly important \[91, 63\]. In this work, we focused primarily on describing the extension and application of the FVC technique to such situations, providing only a proof-of-concept example where CR interplay is relevant while ignoring other practically important effects associated with the possibility of significant temperature gradients in multiple bodies or additional nonlinearities stemming from the temperature-dependent dielectric response of materials \[48\]. Future work along these directions could also focus on extending our work to periodic structures, which could potentially exhibit much larger RHT and hence CR effects. Finally, the asymptotic behavior of RHT at short separations is also impacted by surface roughness \[238\] and ballistic effects (treatable by thermokinetic approaches \[460, 345\]), making a full description of their combined effects of increasing importance.
Chapter 4

Inverse design of radiative heat transfer

The super-Planckian features of radiative heat transfer in the near-field are known to depend strongly on both material and geometric design properties. However, the relative importance and interplay of these two facets, and the degree to which they can be used to ultimately control energy flow, remains an open question. Recently derived bounds suggest that enhancements as large as $|\chi|^4 \lambda^2 / ((4\pi)^2 \text{Im}[\chi]^2 d^2)$ are possible between extended structures (compared to blackbody); but neither geometries reaching this bound, nor designs revealing the predicted material ($\chi$) scaling, have been previously reported.

In this chapter, we exploit inverse design to investigate any possible enhancement to radiative heat transfer via geometric structuring. The organization of this chapter is as the following. In Sec. 4.1, we briefly summarize challenges and recent efforts in tuning RHT via geometric design. In Sec. 4.2, we begin by applying inverse design methods to simple multilayer geometries involving only one-dimensional DOFs. Using a combination of analytical calculations and numerical gradient-based optimization, we show that in complicated multilayer geometries, the scattering and coupling rates
of slab resonances are altered over a broad range of evanescent wavevectors to allow weakly enhanced heat transfer between two slabs. We also explored the flux between a dipolar particle and an inhomogeneous slab, discovering multilayer structures that can allow orders of magnitude enhancement. A brief discussion of hyperbolic metamaterials and anisotropic media shows that they don’t offer any improvement over optimized inhomogeneous slabs.

Moving on to more complicated grating structures allowing two dimensional DOFs, where fast computation of RHT is essential to apply optimization algorithms, in Sec. 4.3 we introduce an efficient formulation enabled by the low-rank properties of elliptic operators for disjoint bodies. Surprisingly, we find that lossy metals such as tungsten, typically considered to be poor candidate materials for strongly enhancing heat transfer in the near infrared, can be structured to selectively realize flux rates that come within 50% of those exhibited by an ideal pair of resonant ideal metals for separations as small as 2% of a tunable design wavelength.

4.1 Overview

In the far field, where the traditional, ray-optical form of Kirchoff’s law (equating emissivity and absorptivity) is applicable [451], computational advances have made it possible to exploit a variety of optimization strategies (exploiting a variety of methods, e.g. random-walk, genetic and particle swarm algorithms, and the Taguchi method) to realize, for instance, selective and/or wide-angle absorbers whose emissivity can come close to the blackbody limits [70, 116, 103, 151, 393, 87, 86, 327, 450, 403, 454, 7], and whose objective is usually that of increasing the performance of a TPV device or solar cell. More recently, development of adjoint optimization techniques in combination with fast numerical EM solvers have allowed application of large-scale optimization [200] methods capable of efficiently tackling problems involving much
higher number of degrees of freedom. For instance, these inverse-design techniques have been exploited to enhance the far-field efficiency of solar-cell absorbers [151, 144], tailor the spectrum of incandescent sources [193], and to increase the functionality of photonic-crystal absorbers [52] and TPV systems [35].

Yet, a concrete understanding of what can be accomplished with near-field RHT remains elusive. Simple high-symmetry structures where analytic solutions are possible provide valuable insight, but appear to be far from ideal. In particular, the most well-studied platform for implementing selective RHT enhancement [444, 191], involving parallel metal plates [113, 233, 192, 243] supporting surface resonances (plasmon or phonon polaritons) [396, 288], has critical deficiencies. First, as dictated by the Planck distribution, there is a natural wavelength scale for observing significant thermal radiative effects near ambient temperatures that spans the near to mid infrared (1 to 10 μm) spectrum [26, 370, 316]. Typical (low-loss) optical materials do not support polariton resonances at these wavelengths, and often lack sufficient thermal stability to withstand longterm operation [320, 374, 118]. Second, the tightest known limits of RHT between extended structures, recently derived using energy conservation and reciprocity arguments [313], reveal that both practical material (|χ|^2/\text{Im } χ) and geometric (λ/d) factors seemingly enable enhancements relative to blackbody emission as large as

\[ F_{\text{limit}} \leq \left( \frac{1}{4\pi} \frac{\lambda}{d} \frac{|χ|^2}{\text{Im}[χ]} \right)^2, \]  

orders of magnitude larger than what is achievable with ideal planar media, suggesting that dramatic improvements are possible through the use of nanostructured surfaces [313]. (In this expression, d stands for separation, and χ material susceptibility, assumed to be the same in both bodies.) Moreover, the |χ|^4/\text{Im } [χ]^2 scaling in Eq. 4.1 indicates that materials exhibiting strong metallic response, far from the typical planar surface–plasmon polariton (SPP) condition Re [χ] = −2, have much greater potential for enhancement. To date, however, this behaviour has not been observed,
and tested geometries employing non-resonant materials [130, 102, 165, 289, 270] have yet to surpass the optimal rates associated with planar bodies.

The application of inverse design to selective RHT enhancement between extended structures is complicated in several ways. First, near-field RHT is controlled by evanescent electromagnetic fluctuations. The large density of these states makes it challenging to apply traditional resonant nanophotonic strategies for enhancing far-field emission over narrow spectral windows [358, 104]. Moreover, the characteristically large field amplitudes and sub-wavelength features of evanescent states make them sensitive to small variations in structural and material properties [271], and correspondingly accurate modeling of RHT requires fine numerical resolution [378, 380, 333, 310].) Second, unlike the far field, RHT can not be decomposed into approximately equivalent independent subproblems. Alterations in the structure of any one object affects the response of the entire system, meaning that the scattering properties of all bodies must be controlled simultaneously. Finally, Maxwell’s equations depend nonlinearly on the dielectric properties and shapes of all bodies, making the optimization non-convex [54] and any a priori guarantee of globally optimal solutions impossible.

Consequently, tractable general approaches for simply calculating near-field RHT have only recently been realized [378, 380, 333, 120], and nearly all previously studied geometries have been designed via trial-and-error approaches exploiting brute-force search over a handful of high-symmetry design parameters [274]. Beginning with bulk metamaterials [41, 169, 270, 279, 275, 399, 108], thin films [32, 134, 28, 433, 341], plasmonic materials [322, 341, 275, 120, 76], and more recently, metallic metasurfaces [274, 130, 478] and gratings [289, 273, 101, 276, 76], selective RHT enhancement has primarily been achieved by tuning the permittivity response, either real or effective, to create or mimic surface resonances. Notably, a recent silicon metasurface design [130] was predicted to have a larger integrated RHT than planar SiO$_2$, which
exhibits low-loss surface phonon polaritons, down to gap distances of 10 nm. Other similarly high symmetry approaches have sought to increase the photonic density of states by exploiting interference (hybridization \[362\]) among the localized plasmons of individual nanostructures \[42\]. Building from simple shapes, tunable RHT rates have been demonstrated in nanobeam (triangular, ellipsoidal, and rectangular unit cells) and nanoantenna arrays exploiting both Mie \[77\] and Fano \[24, 346\] resonances. Although conceptually promising, these approaches have been found to have diminishing returns at small separations (relative to the thermal wavelength). Less restrictive inverse-design techniques have been recently employed to improve the performances of heat-assisted magnetic recording (HAMR) head \[38\]. Relatedly, some authors have also addressed the modulation and optimization of a closely related quantity, the near-field electromagnetic local density of states (LDOS). For instance, Ref. \[257\] performed a parametric study of the LDOS close to a multilayer arrangement of silicon carbide and silicon thin slabs as a function of distance and number of layers while Ref. \[29\] employed genetic algorithms to optimize the LDOS in proximity of a multilayer binary structure composed of aluminium and lossless dielectric layers.

4.2 Inverse design in multilayer structures

4.2.1 Multilayer formulation

In what follows, we consider RHT in geometries involving slabs of arbitrarily varying dielectric profiles \(\varepsilon(z)\) along the direction \(z\) perpendicular to the slab–vacuum interfaces, depicted in Fig. 4.1(a). RHT in such a setup can be described via the fluctuational electrodynamics framework developed by Rytov, Polder, and van Hove (see Refs. \[27, 213\] and references therein). Specifically, we extend a recently developed formulation of this problem \[306, 307\] that expresses the flux in scenarios involving two and three uniform bodies as a function of their reflection and trans-
Figure 4.1: (a) Schematic of two inhomogeneous $\varepsilon(z)$ slabs (A and B) separated by a vacuum gap of size $d$ along the $z$ direction. The radiative heat transfer (RHT) rate between them depends on their local temperatures $T_{A,B}$ and reflection coefficients $R_{A,B}$. Associated with each slab is a coordinate system centered at the slab–vacuum interface and pointing away from the gap. (b) Enhancement factor comparing RHT between optimized inhomogeneous (solid lines) or optimized uniform (dashed lines) slabs against that of semi-infinite uniform plates of $\text{Re}[\varepsilon] = -1$, at a fixed vacuum wavelength $\lambda = 8 \mu m$, as a function of material loss $\text{Im}[\varepsilon]$, and for two representative separations $d = 10 \text{ nm}$ (red) and $500 \text{ nm}$ (blue). The black dotted line shows the theoretical bound described by Eq. 4.4. (c) Transmission coefficient $\mathcal{Z}$ corresponding to either inhomogeneous (red), uniform (blue), or semi-infinite (black) slabs, as a function of the dimensionless wavevector $k_Bd$ at fixed $d = 10 \text{ nm}$ and $\text{Im}[\varepsilon] = 10^{-2}$. The inset shows a typical dielectric profile $\text{Re}[\varepsilon(z)]$ needed to achieve $\mathcal{Z} \approx 1$ over a broad range of $k_B$. To enhance the readability, two different $x$-scales are used in the ranges $[0, 20] \text{ nm}$ and $[20, 100] \text{ nm}$, and the points resulting from numerical optimization are connected with segments.

mission matrices. This approach, together with a semi-analytic expression for the reflection matrix of a slab of arbitrary $\varepsilon(z)$, enables gradient-based optimizations of RHT. Since the system is time and translationally invariant in $x$–$y$, the reflection matrix $\mathcal{R}$ is diagonal in the frequency $\omega$, parallel wavevector $\mathbf{k}_B = (k_x, k_y)$, and polarization $p$, and can be cast as the solution of a differential equation, derived as follows. Consider for each slab a local coordinate system such that $z = 0$ lies at the interface between each slab and the vacuum gap (of size $d$) and points away from the interface. Given a slab occupying the region $[z,t]$ (where $0 < z < t$ and $t$ is the possibly infinite thickness of the slab), let $R(z)$ be the coefficient describing the reflection on the left side, i.e. at the interface $z$. Adding a film of infinitesimal thickness $\Delta z$ at $z$, the reflection coefficient of the combined system (at the $z - \Delta z$ interface) is given
by \( R(z - \Delta z) = \rho(z) + \tau^2(z)R(z)/(1 - R(z)\rho(z)) \), where \( \rho \) and \( \tau \) are the reflection and transmission coefficients of the film, respectively. Taking the limit \( \Delta z \to 0 \), one obtains the following nonlinear differential equation:

\[
\frac{dR(z)}{dz} = \frac{2i k_{zm}(z)}{1 - \tau^2(z)} \left[ r(z) \left( 1 + R^2(z) \right) - (1 + \tau^2(z)) R(z) \right],
\]

which, in combination with the boundary condition \( R(t) = 0 \), describing the absence of the slab (thus a vanishing reflection coefficient) for \( z = t \), completely specifies the reflectivity of the system. Here \( r \) is the ordinary Fresnel reflection coefficient, \( k_{zm}(z) = \sqrt{\varepsilon(z)\omega/c^2 - k_0^2} \) the perpendicular wavevector inside the slab. Note that in the limiting case of uniform \( \varepsilon \), Eq. 4.2 yields the well-known solution \( R(z) = r[1 - e^{2i k_{zm}(t-z)}]/[1 - r^2 e^{2i k_{zm}(t-z)}] \), going to \( r \) in the case of a semi-infinite slab \( (t \to +\infty) \).

Equation 4.2 can be directly solved to obtain the reflection coefficient of a slab of arbitrarily varying \( \varepsilon(z) \): in the case of the two slabs of Fig. 4.1(a), one would also need to specify the boundary conditions \( R(+\infty) = 0 \) for both slabs A and B. Once the function \( R(z) \) is known for each slab, \( R = R(0) \) represents the reflection coefficient needed to calculate RHT and analyze the possible enhancements arising from a given \( \varepsilon(z) \), investigated below via analytical and optimization techniques.

We seek dielectric profiles \( \varepsilon(z) \) that maximize the heat flux \( H[R] \) at a given frequency. In practice, given a choice of slab thickness, numerical evaluations require that the slab be discretized into segments, forming a multilayer geometry. We thus replace the function \( \varepsilon(z) \) with a piecewise-constant function that assumes values \( \varepsilon_i = \varepsilon(z_i) \), with all \( \{\varepsilon_i\}_{i=1,...,N} \) taken as variable degrees of freedom. Note that the size of individual layers typically needs to be very small in order to resolve the exponential decay of evanescent fields, with typical \( N \gg 100 \). Furthermore, while gradient information \( \frac{\partial H}{\partial \varepsilon_i} = \frac{\partial H}{\partial R} \frac{\partial R}{\partial \varepsilon_i} \) is typically needed for large \( N \) [61], \( \frac{\partial R}{\partial \varepsilon_i} \) can be straightforwardly obtained from Eq. 4.2. Because RHT can diverge with vanishing loss rate [313], (while
most plasmonic material have nonzero loss rate (unless doped with gain media [218]), we consider finite but uniform $\text{Im}[\varepsilon]$ throughout the slabs, focusing only on optimizing with respect to $\{\text{Re}[\varepsilon_i]\}$, which leaves modifications in the scattering rather than loss rate as the primary source of enhancement. Since this objective function is far from convex [54], we exploit local optimization algorithms [421, 330].

4.2.2 Optimizing heat transfer between plate–plate

We first consider the scenario of two inhomogeneous, parallel slabs, depicted in Fig. 4.1(a), with both slabs A and B assumed to be in local thermal equilibrium at temperatures $T_A$ and $T_B$, respectively. In this case, the well-known formalism for two slabs of uniform permittivities can be employed. The total heat transfer per unit surface is given by [32]:

$$H = \int \frac{d\omega}{2\pi} \left[ \Theta(T_A) - \Theta(T_B) \right] \Phi(\omega),$$

where the monochromatic spectral component $\Phi(\omega) = \sum_{s,p} \int \frac{d^2 k}{2\pi} k_\beta Z_{s(p)}(k_\beta, \omega)$, with $\Theta(T) = \frac{\hbar \omega}{[\exp(\hbar \omega/k_BT) - 1]}$ denoting the Planck function. Here, we focus on the $p$ polarization, which supports surface modes and hence dominates RHT at short separations $\omega d/c \lesssim 1$. $Z$ is known as the heat transmission coefficient, whose evanescent component is given by:

$$Z(k_\beta, \omega) = \frac{4 \text{Im}[R_A] \text{Im}[R_B] e^{-2\text{Im}[k_z]d}}{|1 - R_A R_B e^{-2\text{Im}[k_z]d}|^2} \quad (4.3)$$

The extension of Eq. 4.15 to the case of inhomogeneous slabs consists in replacing the reflection operators with those from Eq. 4.2.

It is well known that energy transfer is optimal when the scattering and absorption decay rates of the surface modes described by Eq. 4.15 are equal [269]. This corresponds to a maximum transmissivity $Z = 1$, realized at $R_A R_B^* = e^{2\text{Im}[k_z]d}$. For two uniform semi-infinite slabs, such a “rate-matching” condition can only occur at a single $k_\beta$, depending on the separation and loss rate [313], in which case $Z$ exhibits a typical Lorentzian lineshape as a function of $k_\beta$, whose peak lies close to a typical
cutoff wavevector $k_{\text{max}}$, above which $Z$ is exponentially suppressed. For small separations and loss rates and assuming operation close to the surface plasmon resonance of a uniform slab, i.e. $\text{Re}[-1/\chi_{\text{spp}}] = 1/2$, such a cutoff can be approximated by \[313\] $k_{\text{max}} \approx \frac{1}{2d} \ln\left[\frac{|\chi_{\text{spp}}|^4}{4(\text{Im} \chi_{\text{spp}})^2}\right]$, where $\chi = \varepsilon - 1$ is the susceptibility of the material. This leads to an upper bound on the RHT between uniform semi-infinite slabs, given by $\Phi_0 = \frac{1}{2\pi d^2} \ln\left[\frac{|\chi_{\text{spp}}|^4}{4(\text{Im} \chi_{\text{spp}})^2}\right]$ \[313\].

Relaxing the assumption of uniform $\varepsilon$ allows modes of different $k_\beta$ to experience different scattering and absorption rates, potentially allowing rate-matching to not only persist over all $k_\beta \lesssim k_{\text{max}}$ but even beyond $k_{\text{max}}$. The latter condition, however, appears to be prohibitive. As a matter of fact, already in the case of two uniform slabs of finite thickness, the coefficient $Z$ can in principle approach 1 at arbitrarily large $k_\beta$, but only at the expense of exponentially diverging $\text{Re}[\varepsilon] = -\text{Im}[\varepsilon] e^{k_\beta d}$, and vanishing thickness $t = \frac{2}{\text{Im}[\varepsilon] k_\beta} e^{-k_\beta d}$ and bandwidths $\Delta k_\beta = k_\beta e^{-k_\beta d}$, making such an interference effect highly impractical if at all feasible to sustain over a wide range of $k_\beta$. We find, however, that there exist structures that can achieve rate matching over all $k_\beta \lesssim k_{\text{max}}$ and therefore whose flux is described by a larger upper bound $\tilde{\Phi}_0$, obtained by integrating Eq. 4.15 with $Z = 1$ up to $k_{\text{max}}$, given by:

$$\frac{\tilde{\Phi}_0}{\Phi_0} = \frac{1}{8} \ln\left[\frac{|\chi_{\text{spp}}|^4}{4(\text{Im} \chi_{\text{spp}})^2}\right] + \frac{1}{2}, \quad (4.4)$$

The ratio $\frac{\tilde{\Phi}_0}{\Phi_0}$ depends only on material loss, increasing with decreasing loss, as shown in Fig. 4.1(b) (black dotted line). In practice, however, such an enhancement tends to be relatively small because of the logarithmic power law and the fact that inhomogeneity seems to barely increase $k_{\text{max}}$, which is instead primarily determined by the choice of loss rate and $d$.

Next, we exploit optimization to discover inhomogeneous structures that can achieve or approach the monochromatic bounds on $\Phi(\omega)$ above [Eq. 4.4]. Although
we consider the permittivities of the two slabs to be independent degrees of freedom, we find that the optimization always leads to a symmetric dielectric profile, guaranteeing the rate matching condition over a wide range of \( k_\beta \). The inset of Fig. 4.1(c) shows \( \varepsilon(z) \) for one such optimized slab, corresponding to the particular choice of \( \text{Im}[\varepsilon] = 10^{-2} \) and \( d = 10 \) nm at the vacuum wavelength \( \lambda = 8 \) \( \mu \)m (frequency \( \omega = 2\pi c/\lambda \approx 2.35 \cdot 10^{14} \) rad/s), with each dot representing the permittivity of a 1 nm-thick layer. The function \( \text{Re}[\varepsilon(z)] \) shows a strong variation near the slab–vacuum interface, approaching \(-1\) away from the interface. This somewhat unintuitive dielectric profile leads to nearly perfect \( Z = 1 \) for all \( k_\beta < k_{\text{max}} \approx 5/d \), shown in Fig. 4.1(c) (red solid line). In contrast, the transmissivity of either a uniform, semi-infinite slab of \( \varepsilon = -1 \) (black solid line) or a finite slab of optimal thickness \( t_{\text{opt}} \) and permittivity \( \varepsilon_{\text{opt}} \) (blue solid line) exhibit \( Z \sim 1 \) over a smaller range of \( k_\beta \). Moreover, we find that these enhancements are robust with respect to frequency and layer thicknesses on the order of 10 nm. Figure 4.1(b) shows the enhancement factor associated with two different structures, optimized to maximize RHT at either \( d = 10 \) nm (red lines) or \( d = 500 \) nm (blue lines), as a function of the loss rate. We find that at small \( d = 10 \) nm, the achievable enhancements agree well with the predictions of Eq. 4.4 while at larger \( d = 500 \) nm and smaller loss rates, larger enhancements are observed; such a discrepancy is expected since the non-retardation approximation employed in deriving Eq. 4.4 underestimates \( k_{\text{max}} \) at mid-range separations \( \omega d/c \gtrsim 1 \). Even then, the flux rates of inhomogeneous slabs (solid lines) tends to be larger than those of uniform slabs (dashed lines).

**Discussion on anisotropy**

While the configuration of isotropic (possibly inhomogeneous) parallel slabs explored above yields only a small enhancement factor, stemming primarily from the logarithmic power law and difficulty of increasing \( k_{\text{max}} \), one might ask whether it is possible
to further increase $k_{\text{max}}$ by exploiting more exotic media, e.g. electric and magnetic anisotropy. For the sake of simplicity, we restrict our discussion to uniaxial media, described by diagonal permittivity and permeability tensors given by:

$$
\varepsilon = \begin{pmatrix}
\varepsilon_\parallel & 0 & 0 \\
0 & \varepsilon_\parallel & 0 \\
0 & 0 & \varepsilon_\perp
\end{pmatrix}, \quad \mu = \begin{pmatrix}
\mu_\parallel & 0 & 0 \\
0 & \mu_\parallel & 0 \\
0 & 0 & \mu_\perp
\end{pmatrix}.
$$

For such media, RHT is still described by Eq. [4.15] but with a modified expression of the perpendicular component of the wavevector inside the medium, which reads $k_{zm} = \sqrt{\varepsilon_\parallel \mu_\parallel k_0^2 - \frac{\varepsilon_\parallel}{\varepsilon_\perp} k_0^2}$ and $\sqrt{\varepsilon_\parallel \mu_\parallel k_0^2 - \frac{\mu_\parallel}{\mu_\perp} k_0^2}$ for the $p$ and $s$ polarizations, respectively [187]. Moreover, the corresponding Fresnel reflection coefficients have to be modified and become $r^p = \frac{\varepsilon_\parallel k_0 - k_{zm}}{\varepsilon_\parallel k_0 + k_{zm}}$ and $r^s = \frac{\mu_\parallel k_0 - k_{zm}}{\mu_\parallel k_0 + k_{zm}}$ [187]. Since the reflection coefficients of the two polarizations are symmetric with respect to exchange of $\varepsilon$ and $\mu$, implying the existence of both electric or magnetic phonon-polaritons [132], one can focus on only one polarization, e.g. the $p$ polarization. In the extreme near-field regime, where the non-retarded approximation is valid, the corresponding reflection coefficient is well approximated by

$$
r^p \approx \frac{i\varepsilon_\parallel}{i\varepsilon_\parallel/\sqrt{-\frac{\varepsilon_\parallel}{\varepsilon_\perp}} - 1},
$$

and is therefore equivalent to the reflectivity of an isotropic medium of effective permittivity $\varepsilon_{\text{iso}} = i\varepsilon_\parallel/\sqrt{-\frac{\varepsilon_\parallel}{\varepsilon_\perp}}$. Thus, in analogy to the uniform isotropic medium at a fixed separation, the key to increase $k_{\text{max}}$ is to reduce $\text{Im}[\varepsilon]$ at the surface-resonance frequency, defined by the resonance condition $\text{Re}[\varepsilon_{\text{iso}}] = -1$, for which we have, assuming the same loss rate, $\text{Im}[\varepsilon_\parallel] = \text{Im}[\varepsilon_\perp] \equiv \text{Im}[\varepsilon] \ll 1$,

$$
\text{Im}[\varepsilon_{\text{iso}}] \approx \frac{1}{2} \left( \sqrt{\frac{\text{Re}[\varepsilon_\parallel]}{\text{Re}[\varepsilon_\perp]}} + \sqrt{\frac{\text{Re}[\varepsilon_\perp]}{\text{Re}[\varepsilon_\parallel]}} \right) \text{Im}[\varepsilon] \geq \text{Im}[\varepsilon].
$$
Thus, one concludes that the anisotropy does not allow one to decrease losses and hence increase the cutoff wavevector $k_{\text{max}}$.

### 4.2.3 Optimizing heat transfer between dipole–plate

![Figure 4.2: (a) Schematic of an inhomogeneous $\varepsilon(z)$ slab and a dipole separated by a vacuum gap of size $d$ along the $z$ direction. (b) Dielectric profile $\text{Re}[\varepsilon(z)]$ corresponding to inhomogeneous slabs optimized to increase RHT from a dipole a distance $d = 1 \, \mu m$ away from their $z = 0$ interface, at a fixed vacuum wavelength $\lambda = 8 \, \mu m$ (frequency $\omega \approx 0.785c/d$). The profiles are obtained under different constraints on the maximum possible permittivity $\varepsilon_{\text{max}} \equiv \max |\text{Re} \varepsilon| = \{5, 40\}$ (upper and lower figures) but correspond to the same uniform $\text{Im}[\varepsilon] = 10^{-3}$. The imaginary part of the reflection coefficient $\text{Im}[\mathcal{R}(k_{\beta})]$ and LDOS $\mathcal{L}(k_{\beta})$ (in SI units) at the location of the dipole and at $\omega$ are plotted in (b) and (c) as a function of $k_{\beta}$, along with those of optimized uniform slabs (blue solid line). The black dotted line shows the theoretical bound described by Eq. 4.9. The inset in (c) shows the $k_{\beta}$-integrated spectrum $\mathcal{L}(\omega')$ near $\omega$ as a function of the dimensionless frequency $(\omega' - \omega)/\omega \text{Im}[\varepsilon]$, indicating that contributions from smaller $k_{\beta}$ are increasingly sensitive to the wavelength.

In this section, we study RHT between an inhomogeneous slab and a dipole, depicted in Fig. 4.2(a). Beginning with a brief overview of the formulation, we establish an approximate bound for RHT in the case of semi-infinite, homogeneous bodies and exploit optimization to show that multilayer structures can come close to approaching these limits over a broad range of wavevectors, leading to orders-of-magnitude larger RHT. Our work extends recent studies of near-field RHT between dipoles and HMMs or thin films [312] to the mid-field regime.

Consider a small sphere of radius $R \ll d$, approximated as a dipolar particle of polarizability $\alpha$, $d$ being its distance from a planar substrate [see Fig. 4.2(a)].
what follows, we focus on the off-resonance regime \( \alpha D_{ll} \ll 1 \), in which the spectral transfer rate reads \( \Phi(\omega) = 4 \sum_{l=x,y,z} \text{Im}[\alpha(\omega)] \text{Im}[D_{ll}(\omega)] \), where \( D_{ll} \) denotes the Green’s function of the slab at the position of the dipole \[444\]. At short separations \( \omega d/c \lesssim 1 \), the relevant tensor components of the Green’s functions are:

\[
D_{xx} = D_{yy} = \frac{i}{2} \int_0^\infty dk_\beta k_z k_\beta \left( 1 - \text{Re} e^{2ik_zd} \right),
\]

\[
D_{zz} = i \int_0^\infty dk_\beta \frac{k_\beta^3}{k_z^2} \left( 1 + \text{Re} e^{2ik_zd} \right), \tag{4.8}
\]

with the reflection coefficient \( \mathcal{R} \) of the slab obtained again by solving Eq. 4.2. Note that by Poynting’s theorem, the RHT rate is proportional to the LDOS at the position of the dipole, \( \mathcal{L}(\omega) = \frac{1}{2\pi^2\omega} \sum_{l=x,y,z} \text{Im}[D_{ll}(\omega)] \), except in the regime \( k_\beta < \omega/c \) where the latter overestimates RHT since it also captures power radiating into the vacuum region \[332\].

Figure 4.3: (a) Peak LDOS of optimized inhomogeneous slabs (solid lines) and optimized uniform slabs (dashed lines), as a function of the dimensionless separation \( \omega d/c \), for multiple material loss rates \( \text{Im}[\varepsilon] = \{10^{-1}, 10^{-2}, 10^{-3}\} \) (black, blue, and red, respectively). The green dotted line marks the largest possible LDOS in the far field, given by Eq. 4.11. (b) Enhancement factor comparing the peak LDOS of optimized inhomogeneous and uniform slabs.

First, as in the plate–plate scenario, we investigate the possible enhancements in \( \Phi(\omega) \) or \( \mathcal{L}(\omega) \) that can arise from a spatially varying dielectric profile. Unlike the previous scenario, where \( \Phi[\mathcal{R}] \) was a highly non-monotonic function of \( \mathcal{R} \), here the
RHT integrand is linearly proportional to $\text{Im}[R]$. Assuming small losses $\text{Im}[\varepsilon] \ll 1$ and defining $k = ck_\beta/\omega$, a useful figure of merit is the maximum $\text{Im}[R(k)]$ and optimal permittivity of a uniform, semi-infinite slab at any given $k$:

$$\begin{align*}
\text{Im}[R(k)] &= \frac{1}{\text{Im}[\varepsilon]} \frac{2 + 4k^2(k^2 - 1) + 2\sqrt{1 + 4k^2(k^2 - 1)}}{(k^2 - 1)\sqrt{1 + 4k^2(k^2 - 1)}}, \\
\text{Re}[\varepsilon(k)] &= -\frac{1 + \sqrt{1 + 4k^2(k^2 - 1)}}{2(k^2 - 1)}.
\end{align*}$$

Both quantities are strongly divergent at $k = 1$, suggesting that the monochromatic LDOS $L(\omega)$ of an inhomogeneous slab can in principle be unbounded, with the main contribution to the divergence coming from wavevectors near the light cone $k_\beta = \omega/c$. We first observe that such a divergence would theoretically persist for any distance $d$, since the separation enters Eq. 4.8 only as a parameter through the exponential factors. However, Eq. 4.10 shows that at least for a uniform medium, maximizing $\text{Im}[R]$ in the limit $k \to 1$ requires a perfect metal ($\varepsilon \to -\infty$), which can be shown to screen the response at other $k$, resulting in a vanishing bandwidth $\Delta k = 2(k-1)^2 \text{Im} \varepsilon$ and $L(\omega) \to 0$. Consequently, the integrated response $L(\omega)$ of a uniform slab is finite and maximized by a finite thickness and permittivity, which can be found numerically.

More significant improvements, however, can be gained from a spatially varying $\varepsilon(z)$, which provides additional degrees of freedom with which to simultaneously tune the scattering rate at different $k$, allowing the response to approach the bounds described by Eq. 4.9 over wider bandwidths. Realizing such an enhancement presents, however, both conceptual and numerical challenges: waves approaching the light line have increasingly longer wavelength in the $z$ direction and are thus increasingly sensitive to spatial variations, requiring longer slabs and sharper variations in $\varepsilon(z)$. Any numerical optimization strategy will thus benefit only from finite enhancements coming from $k \gtrsim 1$ due to the finite number of layers needed to resolve $\varepsilon(z)$. One should also consider that, as shown above, in the simple case of a uniform slab the permittivity
that maximizes the LDOS at $k \gg 1$ equals -1, while Eq. 4.10 requires $\varepsilon = -\infty$ as $k \rightarrow 1$. In practice, the optimal profile results from a tradeoff between these two conditions, since very high values of $\varepsilon$ act to screen the response from other regions of the slab. Such distinct and challenging requirements make the optimization procedure highly nontrivial, increasing the computational cost of RHT calculations and slowing the convergence rate of the optimization algorithm, which can get easily trapped in multiple local optima. To illustrate these features, we perform separate optimizations with different constraints on the maximum allowed permittivity $\varepsilon_{\text{max}} = \max\{|\text{Re}\varepsilon|\}$, which limits potential enhancements coming from waves near the light line.

Figure 4.2 reports optimizations of the evanescent contribution to $\mathcal{L}(\omega)$ at a vacuum wavelength $\lambda = 8 \ \mu \text{m}$ and for $d = 1 \ \mu \text{m}$. Figure 4.2(b) shows representative profiles $\text{Re}[\varepsilon(z)]$ obtained under different $\varepsilon_{\text{max}} = \{5, 40\}$ and at fixed $\text{Im}[\varepsilon] = 10^{-3}$. Noticeably, the lower profile exhibits rapid, subwavelength variations over small (tens to hundreds of nanometers) regions: as discussed earlier, these are needed to maximize $\text{Im}[\mathcal{R}]$ near $k_{\beta} = \omega/c$ while avoiding screening effects at larger $k_{\beta}$, with higher $|\text{Re}[\varepsilon]|$ occurring away from the interface for the same reason. This explains why larger $\varepsilon_{\text{max}}$ lead to greater enhancements, illustrated in Figs. 4.2(c) and (d), which show $\text{Im}[\mathcal{R}]$ and $\mathcal{L}$ as a function of $k_{\beta}$. The results, which are also compared against those of uniform slabs of optimal thickness and permittivity (blue line), reveal that inhomogeneous structures can approach the bounds described by Eq. 4.9 (dashed black line) over much broader range of $k_{\beta}$. Although producing a significant increase with respect to uniform slabs, the optimization fails as $k_{\beta} \rightarrow \omega/c$ due to the practical limitations discussed above. Moreover, these enhancements will necessarily come at the expense of increased frequency selectivity, since waves near the light line are most sensitive to deviations in the long-range spatial pattern of the structure, here optimized to realize a specific interference pattern at $\omega$. This feature is apparent from the inset of Fig. 4.2(d), which shows the spectra $\mathcal{L}(\omega')$ of the optimized uniform
and inhomogeneous slabs from above (neglecting material dispersion): namely, the contribution of lower \( k \) states becomes increasingly restricted to frequencies \( \omega' \approx \omega \) as \( k \to \omega/c \). (Note that the factor of \( \text{Im} \varepsilon \) in the abscissa is there because just as in the case of a uniform medium, the bandwidth of the Lorentzian-like spectrum is proportional to the loss rate.)

We now explore the enhancement factor from optimized inhomogeneous slabs for a wide range of separations \( d \in [0.5, 9] \) \( \mu m \). To begin with, Fig. 4.3(a) shows the maximum LDOS of the optimized inhomogeneous (solid lines) and uniform (dashed lines) slabs, as a function of \( d \) and for multiple values of \( \text{Im}[\varepsilon] = \{10^{-3}, 10^{-2}, 10^{-1}\} \) (red, blue, and black lines respectively), with their ratio, the enhancement factor, depicted in Fig. 4.3(b). As shown, in both situations the LDOS increases rapidly with decreasing \( \omega d/c \) and decreasing material losses. In particular, the enhancement factor [in principle infinite for any \( d \), as suggested by Eq. 4.9] increases up to a maximum value (dictated by the smallest participating, enhanced wavevector) and then decreases, approaching 1 as \( d \to \infty \). Essentially, at large \( d \), the evanescent LDOS becomes increasingly dominated by wavevectors close to the light line (for which the optimization procedure fails). Consequently, beyond some separation the propagating contributions to \( \mathcal{L}(\omega) \) dominate. Such finite enhancements also imply that at small \( d \), where the LDOS becomes increasingly dominated by large \( kc/\omega \gg 1 \) waves, the optimal slab is one satisfying the typical resonant condition of \( \text{Re}[\varepsilon] = -1 \) and hence the enhancement factor approaches 1.

For comparison, the green dotted line in Fig. 4.3(a) denotes the largest achievable far-field LDOS in planar media,

\[
\max\{\mathcal{L}^{\text{prop}}\} = \left(\frac{4}{3} + \frac{\sqrt{2}}{3}\right) \frac{\omega^2}{\pi^2 c^3}, \quad (4.11)
\]
derived by summing the propagative contributions under rate-matched reflection and energy conservation, $|R_{p(s)}| \leq 1$, in which case $|\text{Re}[R_{p(s)}e^{2ik_zd}]| \leq 1$. More precisely, the limit Eq. [4.11] can be derived by observing that for the $p$ polarization,

$$
\mathcal{L}^{\text{prop}}_p = \frac{\omega^2}{\pi^2 c^3} + \frac{\omega^2}{\pi^2 c^3} \int_0^{k_0} dk_\beta \left( \frac{k_\beta^3}{k_z} - k_\beta k_z \right) \text{Re}[R_p e^{2ik_zd}]
$$

$$
\leq \frac{2\omega^2}{3\pi^2 c^3} (1 + \sqrt{2}),
$$

(4.12)

with the maximum achieved for structures with $\text{Re}[R_p e^{2ik_zd}] = 1 \ (-1)$ at $\frac{k_\beta^3}{k_z} - k_\beta k_z > 0 \ (< 0)$; a similar bound applies to the $s$ polarization, leading to Eq. [4.11]. Since it is derived under the assumption of an integrand maximized for any $k_\beta$, $\max\{\mathcal{L}^{\text{prop}}\}$ provides an upper bound that is challenging to realize. However, as shown in Fig. 4.3, it is still smaller than the evanescent part of the LDOS for optimized inhomogeneous slabs over a wide range of separations. In particular, we remark that at the separations where the enhancement factor peaks, the evanescent contribution to the LDOS of the optimized homogeneous slabs is more than an order of magnitude larger than $\max\{\mathcal{L}^{\text{prop}}\}$.

**Discussion on hyperbolic metamaterials**

Since LDOS enhancements from inhomogeneous slabs prove significant at mid-range separations, one may wonder whether similar enhancements can be achieved in previously studied planar geometries. One such geometry are HMMs, which consist of alternating metal and dielectric layers and are known to exhibit hyperbolic dispersion [169, 45]. While it has been demonstrated that RHT between a dipole and a HMM in the deep near field is no larger than that of an appropriately designed uniform thin film [312], we analyze below whether this remains true in the mid-field regime. Consider for instance, a HMM of period $a \ll \lambda, d$ described as an effective, uniform anisotropic medium, with permittivities $\varepsilon_\parallel$ (surface-parallel) and $\varepsilon_\perp$ (surface-
perpendicular), having real parts $\varepsilon_1$ and $\varepsilon_2$ respectively, and by assumption the same small imaginary part $\text{Im}[\varepsilon] \ll 1$. The hyperbolic regions of the spectrum are those in which $\varepsilon_1 \varepsilon_2 < 1$, i.e. when the real parts have opposite signs. More specifically, one typically defines two categories of HMMs: type I with $\varepsilon_1 > 0$ and $\varepsilon_2 < 0$, and type II with $\varepsilon_1 < 0$ and $\varepsilon_2 > 0$. While in the extreme near-field there is no distinction between the two types, given that only the product $\varepsilon_1 \varepsilon_2$ matters, the two exhibit very different behavior in the mid-field. Focusing on the dominant, $p$ polarization ($R \approx r^p$), the relevant quantity in type-I HMMs is

$$\text{Im}[R_1] \approx \frac{2\sqrt{k^2 - 1} \sqrt{\varepsilon_1 - \frac{\varepsilon_1}{\varepsilon_2} k^2}}{\varepsilon_1 (k^2 - 1) + 1 - \frac{1}{\varepsilon_2} k^2} \leq 1,$$

which is clearly far smaller than that of uniform isotropic slabs (scaling as $1/\text{Im}[\varepsilon]$), resulting in much smaller RHT in the limit of small $\text{Im}[\varepsilon] \ll 1$. The behavior of type-II HMMs, on the other hand, varies depending on two different $k$ regimes: $k > \sqrt{\varepsilon_2}$, in which case $\text{Im}[R_{II}] \approx \text{Im}[R_1] < 1$, and $1 \leq k < \sqrt{\varepsilon_2}$ (requiring $\varepsilon_2 > 1$), in which case the peak

$$\text{Im}[R_{II}] \approx \frac{4\varepsilon_1 \varepsilon_2 (\varepsilon_2 - 1)}{\varepsilon_1 - 1 + \varepsilon_2 - \varepsilon_2^2 \text{Im}[\varepsilon]}$$

occurs at $k_m = \sqrt{\frac{(1-\varepsilon_1)\varepsilon_2}{1-\varepsilon_1 \varepsilon_2}}$. Note that similar to the case of isotropic slabs, $\text{Im}[R_{II}] \to \infty$ as $\varepsilon_1 \to -\infty$ and $\varepsilon_2 \to \infty$. However, as before, such a large dielectric constant results in a significant screening effect and thus narrow bandwidth, whose full width at half maximum $\Delta k \approx \frac{|\varepsilon_1 - 1 + 3\varepsilon_2 (\varepsilon_2 - 1)|}{\sqrt{(1-\varepsilon_1 \varepsilon_2)^2 \varepsilon_2^2 (1-\varepsilon_1)}} \text{Im}[\varepsilon]$. In the limit of a diverging permittivity, $\Delta k \to 0$ faster than $\text{Im}[R_{II}]$ diverges, leading to vanishing RHT. Hence, one finds once again that HMMs are in principle not better than uniform, isotropic thin films at mid-field separations and are therefore never “discovered” by the optimization method.
4.2.4 Concluding Remarks

We remark that in the scenario of inhomogenous slabs, the LDOS decreases smoothly with increasing separation and material loss, while in the case of optimal uniform slabs, material losses become irrelevant at distances $d \gtrsim 3c/\omega$. In order to explain this feature, we observe that for for $\text{Im}[\varepsilon] = \{10^{-3}, 10^{-2}\}$ (red and blue lines respectively), the first-order derivative of the peak LDOS for uniform slabs is a discontinuous function of separation at a given $\tilde{d}$, depending on $\text{Im}[\varepsilon]$. This results from an abrupt transition between two mechanisms of enhancement. In particular, depending on the separation, the uniform slab either maximizes the LDOS at some intermediate $k \gg 1$ through a resonant $\text{Re}[\varepsilon] \approx -1$ (small $d$) or near $k \approx 1$ with $\text{Re}[\varepsilon] \to -\infty$ (large $d$). In the latter case, the imaginary part of the permittivity becomes irrelevant.

4.3 Inverse design in grating structures

Building on our earlier examination of RHT between multilayer bodies in the previous section, we now employ inverse design \[318\] to investigate RHT between generalized two-dimensional gratings, and restrict the analysis to realistic materials and fabricable structures.

In the case of periodic gratings, RHT is

$$\Phi(\omega) = \frac{1}{2\pi} \int \frac{d^n k}{(2\pi)^n} T(\omega, k),$$

(4.15)

where the integration is carried out over Bloch-vectors $k$ in the first Brillouin zone (BZ), and the scattering properties of the structures are captured by the transfer function $T(\omega, k)$ described by Eq. \[4.19\]. To maximize the transfer function, both the density and coupling efficiency of the participating states \[317, 269\] must be made as large as possible at all $k$. A similar problem arises in the design of far-field
emitters, where resonant structures are often intuitively designed to match the absorption and coupling rates of a wide range of externally excited states, leading to a complete suppression of scattered fields. But the task here is more complicated, as the electromagnetic field must be regulated over a much larger (evanescently coupled) range. For intuition based structures with a few tunable parameters, there does not seem to be enough design freedom to reach this level of control, with the range of rate-matched states occurring in current designs falling short of those achievable in planar high index dielectrics ($n > 3$). Without a viable means of addressing these questions, we turn to inverse techniques.

As an initial step towards the broader development of this area, and for computational convenience, we focus exclusively on two-dimensional gratings (fields and dielectrics are translation-invariant along the third dimension). This choice has major consequences for the underlying physical processes. Particularly, in moving from three to two dimensions, the geometric and material scaling of the density of states decreases and as a result, the maximum RHT rate between two ideal planar metals, $\text{Re}[\chi(\omega)] = -2$, becomes finite in the limit of vanishing loss

\[
\mathcal{F}_{\text{pl}}^{2D}(\omega) = \frac{\lambda}{2\pi d}, \quad (4.16)
\]

$\mathcal{F}(\omega) = \Phi(\omega)/\Phi_0(\omega)$, and $\Phi_0(\omega) = \omega/\pi^2 c$ ($\omega^2/4\pi^2 c^2$) the spectral emission rate per unit area of a two (three) dimensional planar blackbody. In contrast, $\mathcal{F}_{\text{pl}}^{3D}(\omega) = \lambda^2 \ln (2/\text{Im}[\chi(\omega)]) / (2\pi^2 d^2)$ exhibits stronger geometric and material enhancement factors. Consequently, achieving a strong material response $\chi(\omega)$ at the desired frequency window along with broadband rate-matching through nanostructuring is expected to have more significant impact in three dimensions. As Eq. (4.16) is useful standard for comparing the efficacy of any given design it will be used as a normalization throughout the remainder of the manuscript.
4.3.1 Low-rank formulation

Here we describe a computational method that allows fast computations of RHT between arbitrarily shaped gratings of period $\Lambda$, separated by a vacuum gap $d$. Within fluctuational electrodynamics, the calculation of RHT consist of determining the absorbed power within a body $B$, $\Phi(\omega) = \frac{1}{2}\omega\varepsilon_o \int_{V_B} d\mathbf{r}' \, \text{Im} [\chi(\mathbf{r}', \omega)] \langle |\mathbf{E}(\mathbf{r}', \omega)|^2 \rangle$, resulting from thermally excited current sources originating within a different body $A$. Given a discretized computational grid and assuming local media, these sources obey the fluctuation–dissipation relation \[ \langle j_{\gamma,i}^* j_{\beta,j} \rangle = \frac{4\omega\varepsilon_o}{\pi} \delta_{ij} \delta_{\gamma \beta} \text{Im}[\chi_i(\omega)] \Theta(\omega,T_i). \] (4.17)

Here, $\Theta(\omega,T) = \frac{\hbar \omega}{(e^{\hbar \omega/k_B T} - 1)}$ is the Planck function, $\langle \ldots \rangle$ a thermal ensemble average, $\{i, j\}$ the index of a given location or pixel within the computational grid, and $\{\gamma, \beta\} = \{x, y, z\}$ the vector polarizations. Equation (4.17) can be used in conjunction with knowledge of the electric Green’s function $G$ of the system \[ \Phi(\omega) = \frac{\omega^4 \Theta(\omega,T)}{2\pi c^4} \sum_{\beta,\gamma} \int_{V_A} d\mathbf{r} \int_{V_B} d\mathbf{r}' \, \text{Im} [\chi(\mathbf{r}, \omega)] \text{Im} [\chi(\mathbf{r}', \omega)] |G_{\beta\gamma}(\mathbf{r}', \mathbf{r}, \omega)|^2. \] (4.18)

Writing this in matrix form, with superscripts denoting projections onto the respective body, and $G$ denoting the matrix form of the electric Green’s function, it follows that RHT can be written as a Frobenius norm,

\[ \Phi(\omega) = \frac{\omega^4 \Theta(\omega,T)}{2\pi c^4} \| \sqrt{\text{Im}[\chi^A]} G^{AB} \sqrt{\text{Im}[\chi^B]} \|_F^2. \] (4.19)

The main challenge in evaluating Eq. (4.19) lies in the need to repeatedly evaluate and multiply $G^{AB}$, the inverse of a sparse matrix. Direct application of either sparse-direct \[255\] or iterative solvers \[354\] would demand extraordinary computational
resources, especially in three dimensions. In particular, without additional simplifications, within a particular numerical discretization, the number of computations at each iteration of an optimization required is at least the rank of the matrix $\sqrt{\text{Im} [\chi^B]}$, or three times the number pixels in $B$ (polarizations). However, because $G_{AB}$ does not describe fields created by current sources within the same body (but only disjoint bodies), it admits a low-rank approximation [75]. Hence, Eq. 4.19 can be well approximated by a singular value decomposition of the matrix $Z_{AB} = \sqrt{\text{Im} [\chi^A]} G_{AB} \sqrt{\text{Im} [\chi^B]}$,

$$
\Phi (\omega) = \frac{\omega^4 \Theta (\omega, T)}{2\pi c^4} \sum_i |\sigma_i|^2,
$$

requiring only a small set of singular values $\{\sigma_i\}$. Applying the fast randomized SVD algorithm [174], detailed in Supplemental Material, we find that typically no more than 8 singular values are needed to reach an error estimate better than $1 \cdot 10^{-3}$, reducing the number of required matrix solves to $\lesssim 16$. (As derived in Supplemental Material, this trace formulation also enables fast gradient computations via the adjoint method [318].)

The inverse problem is then to maximize $\sum_i |\sigma_i|^2$ with respect to variations in $\chi$. Such an optimization can be carried out in the framework of topology optimization using the adjoint method [318], allowing a huge range of design parameters (each pixel within the optimization domain). We find, however, that local, gradient-based optimization leads to slow convergence to fabricable structures and comparatively suboptimal designs. To avoid these difficulties, we instead considered a range of shape optimizations [318], detailed in Sec. 1.1.2. While limiting the space of discoverable structures, this choice allows for application of statistical Bayesian algorithms [62, 299] in combination with fast, gradient-based optimization. In what follows, this procedure, implemented with a simple 2d FDFD Maxwell solver [465], is applied to selectively enhance RHT at the thermal wavelength $\lambda_{opt} = 3 \mu m$ corresponding to
Figure 4.4: Near-field RHT enhancement for inverse-designed tungsten gratings $\mathcal{F}$, along with that of planar silicon carbide for comparison, relative to the enhancement achieved with ideal (lossless) planar metals, $\mathcal{F}_{pl}^{2D}$ given in Eq. 4.16 with respect to frequency $\omega$ (left). Profiles of the structures are displayed as insets. Color plots depicting the $k$-dependent transfer functions $\mathcal{T}(\omega, k)$ over the chosen frequency range are shown on the right. The depicted optimization proceeds from an unstructured planar system (0) to structured gratings (N) by successively introducing additional ellipsoidal degrees of freedom “N” to the design space. Performance is qualified by the magnitude of RHT at a single design frequency $\omega_{opt} = \frac{2\pi c}{3\mu m}$ ($\lambda_{opt} = 3 \mu m$), where tungsten behaves as a highly lossy metal far from the planar surface plasmon resonance. The gap separation and the period of gratings are $d(\Lambda) = 0.02 (0.04) \lambda_{opt}$, respectively.

peak emission at $T = 1000$ K. The surface–surface separation between the two bodies is fixed to be $d = 0.02 \lambda_{opt}$ (deeply in the sub wavelength regime).

4.3.2 Lossy metal grating optimization

We begin by considering tungsten gratings, which at $\lambda_{opt}$ exhibits a highly metallic response $\chi \approx -170 + 37i$ far from the planar SPP condition. Fig. 4.4 depicts the spectral enhancement factor $\mathcal{F}/\mathcal{F}_{pl}^{2D}$ for both unstructured plates “(0)” and optimized gratings “(N)” obtained by successively increasing the number of (ellipsoidal) shapes allowed in each unit cell, $N$. The spectra of the two optimized gratings, illustrated as insets\(^1\), both peak at $\lambda_{opt}$ (black dashed line), with magnitudes $\mathcal{F}/\mathcal{F}_{pl}^{2D} = \{0.40, 0.53\}$

\(^1\)Full geometric characterizations of these structures and those presented later in the text are available upon request.
increasing with the number of ellipsoids, \(\{1, 3\}\)\(^2\). On the one hand, enhancements of this magnitude for lossy metals, \(|\chi|/\text{Im} (\chi) \gtrsim 1\), are considered challenging \(^3\) at small separations, \(d \ll \lambda_{\text{opt}}\). On the other hand, large RHT is known to be possible in ultra-thin films through the interference of coupled SPPs \(^{170}\). However, to reach the magnitudes obtained here, unrealistically small thicknesses \(\lesssim 1\) nm are needed. In contrast, no feature in the gratings of Fig. 4.4 is smaller than 10 nm. Notably, although the optimization is carried out at a single frequency, the discovered enhancement peaks are always broadband due to the high level of material absorption. Consequently, the frequency-integrated RHT at \(T = 1000\) K exhibited by grating (3) is found to be roughly 10% larger than that of two planar silicon carbide (black line), a low-loss polaritonic material. To explain this enhancement, Fig. 4.4(right) examines the transfer function \(T(\omega, k)\) versus frequency and wavenumber \(k\). In moving from (1) to (3), the color plot demonstrates (frequency axis) that additional modes are successively created and pushed towards \(\omega_{\text{opt}}\), enhancing the density of states. Owing to the large size of the BZ (small periodicity \(\Lambda = 0.04 \lambda_{\text{opt}}\)), we find that the range of rate matching achieved here is considerably larger than that observed in previously examined grating structures \(^{101}\).

Another key finding is depicted in Fig. 4.5, which plots RHT enhancement for representative optimizations \(^3\) across an array of material and geometry combinations, as a function of the material scaling factor \(|\chi|^4/\text{Im} [\chi]^2\) of Eq. 4.1. Three different classes of design are explored: collections of ellipsoids (circles), single polyline interfaces (squares), and Fourier curves (triangles). Uniformly, every one of these structures is observed to enhance RHT by at least an order of magnitude compared to the corresponding planar systems and within factors of unity of the ideal planar bound, Eq. 4.16. Regardless of the particular parameterization, over the range of

---

\(^2\)The optimizations involving four elliptical bodies usually require \(\approx 10^2\) iterations to reach convergence; roughly 200 hours of total computation time for each structure.

\(^3\)The structures with the best performance features over all trials with the same susceptibility.
Figure 4.5: The figure highlights the material trend in enhancement $F/F_{pl}^2$ found by optimizing grating geometries over a wide range of material susceptibilities $\chi$. Quantities displayed in parenthesis correspond to the relative enhancement of a given grating design compared to a planar geometry of the same material. Various shape parameterizations such as polylines, Fourier curves, and ellipsoids marked as squares, triangles, and circles are considered. The susceptibility values are chosen to vary along either the real or imaginary axis: gratings {1, 2, 4} correspond to $\chi = \{-76, -101, -151\} + 50i$, 5 to $\chi = -151 + 40i$, 3 to $\chi = \{-121, 169.5\} + 37.3i$, and {6 → 9} to $\chi = -169.5 + \{50, 37.3, 30, 25\}i$. Results for tungsten are colored red. The upper inset illustrates the unit cell of correspondingly numbered grating. (There is no visible difference for gratings 6-9 and so only one of these is shown.) The lower inset depicts the RHT of tungsten grating 7 when the susceptibility is varied without altering the structure. The dashed line represents the susceptibility scaling predicted by recently derived limits on RHT [313].

Examined Re[$\chi$] and Im[$\chi$], a clear linear trend in $F$ (red dotted line) as a function of $|\chi|^4/\text{Im}[\chi]^2$ is observed. This linear scaling becomes increasingly difficult to observe at larger values of $\chi$, where larger resolutions are needed to accurately capture resonant behavior and the optimization requires increasingly larger number of iterations to find structures along the fit line. As should be expected, based on the fact that this behavior has not been previously reported, material scaling consistent with Eq. 4.1 is seen only for gratings optimized specifically for each particular value of $\chi(\omega_{opt})$. For a fixed geometry, Fig. 4.5 (insets), varying either Re[$\chi$] or Im[$\chi$] shifts the resonance away from $\omega_{opt}$ and diminishes RHT. The appearance of this linear trend indicates
that aspects of the arguments leading to the energy bounds in [313] are coming into play. However, the fact that the slope of the fit line is too flat to pass through the origin indicates that a substantial increase in available design space has failed to significantly bridge the magnitude gap, or that approaching these bounds (at least in 2d) may prove challenging with practical structures. Conversely, it should be emphasized that our results do not preclude the existence of structures with much larger enhancements. First, although the design space we have investigated is substantially larger than previous work, it is still relatively limited. Second, the complexity of the strutures and degree of enhancement are limited by the spatial resolution of the chosen discretization, \(0.0005 \lambda_{opt}\) (1/40th of the gap size).

4.3.3 Concluding Remarks

To summarize, in investigating potential radiative heat transfer enhancements through inverse design, we have found evidence supporting the material scaling recently predicted by shape-independent bounds [313], a feature that to our knowledge had yet to be confirmed. While the observed heat transfer rates are still far from matching the magnitudes predicted by general bounds, we have found that RHT rates between fabricable tungsten gratings (a highly lossy metal), for subwavelength gap separations as small as 2% of the design wavelength, can selectively approach 50% of the rate achieved by ideal planar materials (lossless metals satisfying the SPP condition) in the infrared. The results represent nearly two orders of magnitude greater RHT rates in structured compared to planar materials. It remains to be seen to what degree similar strategies might enhance RHT in three dimensions, where the photonic density of states is larger and its associated dependence on material losses significantly stronger. The FVC formulation developed in Sec. 2.1 can be applied to efficiently capture RHT involving heterogeneous objects.
Chapter 5

Nonlinear frequency conversion and multimode couplers

Nonlinear frequency conversion can be enhanced in a resonator \[335\], which offers both longer interaction timescales and higher field confinements. Beginning with large-etalon cavities \[127\] initially considered in the mid-60s, and then moving from millimetre- \[143\] to micrometre-scale whispering gallery mode resonators \[282, 347, 40\], and more recently proposed wavelength-scale photonic crystal cavities \[376, 55\], these ideas have continued to be pushed to realize higher efficiencies (lower pump powers), more compact architectures and faster operating timescales (wider bandwidths). While these conventional designs fall short of simultaneously meeting the many design challenges associated with resonant frequency conversion, chief among them being the need to support multiple modes with highly concentrated fields, exactly matched resonant frequencies, and strong mode overlaps \[377\], recently, we developed and proposed powerful, large-scale optimization techniques to allow computer-aided photonic designs that can address all of these challenges.

In this chapter, extending our previous proof-of-concept work on optimizing resonators for efficient second harmonic generation \[265, 267\], we demonstrate inverse
designs that address two crucial problems in moving onto the stage of on-chip integration for such nonlinear devices. First, in Sec. 5.1 we demonstrate that inverse design can be applied to discover fully three-dimensional structures exhibiting huge nonlinearities; examples include compact cavities and extended metasurfaces. Second, nonlinear frequency conversion requires efficient coupling between on-chip sources and cavity modes over a set of far-apart frequencies. However, despite the importance of this basic functionality, also present in many other problems in silicon photonics, there are few systematic design tools to simultaneously control coupling between multiple tightly confined modes with far-separated wavelengths, and a single waveguide. In Sec. 5.2 we propose a large-scale optimization formulation to produce wavelength-scale waveguide–cavity couplers operating over tunable and broad frequency bands. We numerically demonstrate couplers that can achieve critical or nearly critical coupling between multi-ring cavities and a single waveguide, at up to six widely separated wavelengths spanning the 560–1500 nm range of interest for on-chip nonlinear optical devices.

5.1 Optimizing nonlinear processes in three-dimensional structures

Our recently demonstrated optimization framework allows automatic discovery of novel cavities supporting tightly localized modes at several desired wavelengths and exhibiting large nonlinear mode overlaps. As a proof of concept, we proposed doubly-resonant structures, including multi-layered, aperiodic micro-post cavities and multi-track ring resonators, capable of realizing second-harmonic generation (SHG) efficiencies exceeding $10^4 \, \text{W}^{-1}$ [265, 267]. However, in those designs, we either explored only low dimensional DOFs for high-symmetry geometries such as multi-track ring resonators [267] and optical fibers [404], or manually extruded structures optimized in
two dimensions to three dimensions \[265\]. In situations of more restricted design space where the above approaches can fail, one needs to apply inverse design techniques directly to three-dimensional structures. However, a typical large-scale optimization routine involves solving Maxwell’s equations several thousands of times, which is computationally intensive in three dimensions. Recently, advances in high-performance hardware such as graphics processing units significantly speed the iterative solvers of sparse matrix, enabling us to explore the following inverse designed metasurfaces and microcavities in three dimensions.

Below we briefly introduce our optimization formulation in the context of SHG processes; the same formulation can be easily generalized to include more complicated nonlinear processes \[404\]. Instead of solving a set of nonlinear equations, our formulation frames the problem of creating two highly coupled resonances as a set of coupled linear scattering problems, involving both the fundamental \(\omega_1\) and second-harmonic \(\omega_s = 2\omega_1\) frequencies; within first order perturbation theory, such an approach can accurately describe the physics of SHG. We follow the notation introduced in Sec. 1.1, where the dielectric constant at every pixel \(\mathbf{r}_a\) is \(\varepsilon_\alpha = \varepsilon_b + x_\alpha (\varepsilon_m - \varepsilon_b)\), in which \(\varepsilon_{b(m)}\) is the permittivity of the background (nonlinear) medium, and the variable \(x_\alpha \in [0, 1]\) is the DOF in the topology optimization, or described by some shape parameters in the shape optimization. Consider the SHG process as a frequency mixing scheme in which two photons at \(\omega_1\) interact to produce an output photon at the \(\omega_s\). Given the nonlinear tensor component \(\chi^{(2)}_{ijk}\), with \(i, j, k \in \{x, y, z\}\), mediating an interaction between the field components \(E_i(\omega_s)\) and \(E_j(k)(\omega_1)\), we begin with two point dipole currents at \(\omega_1\), such that \(\mathbf{J}_n(\omega_1) = \hat{\mathbf{e}}_{n\nu} \delta(\mathbf{r} - \mathbf{r}')\), where the two currents can have different profiles, denoted by the subscript \(n\), and \(\hat{\mathbf{e}}_{n\nu}\) is a polarization vector chosen so as to excite the desired electric-field polarization components \((\nu)\) of the corresponding mode at an appropriate position \(\mathbf{r}'\). Given the choice of incident currents \(\mathbf{J}_n(\omega_1)\), we solve Maxwell’s equations to obtain the corresponding electric-field response \(\mathbf{E}_n(\omega_1)\), from which one
can construct a nonlinear polarization current \( J(\omega_s) = x(\mathbf{r}) \prod_{n=1}^{2} E_{nv}(\omega_1) \hat{e}_i \), where \( E_{nv}(\omega_1) = E_n(\omega_1) \cdot \hat{e}_{nv} \) and \( J(\omega_s) \) can be generally polarized (\( \hat{e}_i \)) in a (chosen) direction that differs from the constituent polarizations \( \hat{e}_{nv} \). Finally, maximizing the radiated power, \(-\text{Re}\left[ \int J(\omega_s)^* \cdot E(\omega_s) \, d\mathbf{r} \right] \), due to \( J(\omega_s) \), one is immediately led to the following nonlinear optimization problem:

\[
\max \mathcal{F} = -\text{Re}\left[ \int J(\omega_s)^* \cdot E(\omega_s) \, d\mathbf{r} \right],
\]

Writing down the objective function \( \mathcal{F} \) in terms of the nonlinear polarization currents, it follows that solution of Eq. 5.1, obtained by employing any mathematical programming technique that makes use of gradient information, e.g. the adjoint variable method [202], maximizes the nonlinear coefficient (mode overlap) associated with the SHG process. The above framework can be easily extended to consider propagating modes once we take into account the appropriate Bloch boundary conditions that may arise from any desired wave vectors imposed at the requisite frequencies [422]. In the case of optical fibers or metasurfaces (or, more generally, any waveguiding system), such an extension naturally guarantees perfect phase- and frequency-matching of the relevant modes in the optimized structure.

### 5.1.1 Metasurface

Metasurfaces offer an advantageous platform for realizing complicated beam generation and wavefront shaping over extended surfaces [473] and have recently been exploited in conjunction with nonlinear materials as a means of generating and con-
trolling light at multiple wavelengths [252, 311, 217, 391]. A typical nonlinear metasurfaces can suffer from poor frequency-conversion efficiencies due to a combination of weak confinement, material absorption, and sub-optimal mode overlaps. In particular, typical designs exploit plasmonic [66, 457, 250, 458] or all-dielectric [272, 467] elements comprising simple shapes distributed over a unit cell, including split ring resonators [66, 458, 457], cross-bars [250], and cylindrical posts [272], with the main focus being that of satisfying the requisite frequency- and phase-matching condition [234]. Here, we show that inverse design can not only facilitate the enforcement of frequency- and phase-matching requirements but also allow further enhancements stemming from the intentional engineering of nonlinear modal overlaps, often neglected in typical designs.

![Figure 5.1: (a) Schematic illustration of SHG in a square-lattice metasurface of finite thickness \( t \) and period \( \Lambda \times \Lambda \). Shown to the right are dielectric profiles and mode profiles \( |E|^2 \) corresponding to two inverse-designed metasurfaces, both over single unit cells and \( (z = 0) \) cross sections. The structures are optimized to ensure frequency- and phase-matching for light incident at an angle \( \theta = 3^\circ \) (i) or normal incidence (ii). Dark (white) represents gallium phosphide (vacuum) regions. (b) Convergence of the objective function with respect to iteration number, leading to structure (ii).](image)

To achieve large SHG efficiencies, a metasurface must support two extended resonances at frequencies \( \omega_1 \) and \( \omega_s = 2\omega_1 \) and wavevectors satisfying the phase-matching
condition $k_s = 2k_1$. As illustrated schematically in Fig. 5.1(a), a typical setup consists of an incident wave of power per unit cell $P_1$ from free space at some frequency and angle (described by wavenumber $k_1$) and a corresponding output harmonic wave of power per unit cell, $P_s$. In the small-signal regime, the output power $P_s \propto P_1^2$ scales quadratically with $P_1$, resulting in a conversion efficiency per unit cell, derived from the coupled mode analysis \[124\],

$$\eta = \frac{P_s}{P_1^2} = \frac{32Q_1^4Q_s^2}{\omega_1 Q_{1,r} Q_{s,r} |\beta_2|^2}$$  \hspace{1cm} (5.2)

where $Q$ and $Q_r$ denote total and radiative dimensionless lifetimes and $\beta_2 = \beta_1^*/2$ the nonlinear overlap factor \[377\], a generalization of the familiar phase matching condition to wavelength-scale structures,

$$\beta_1 = \frac{1}{4} \frac{\int dV \varepsilon_0 \sum_{ijk} \chi^{(2)}_{ijk} (E_{ik}^* E_{1j}^* E_{sk} + E_{ik}^* E_{s} E_{1j}^* E_{lk})}{\left(\int dV \varepsilon_0 \varepsilon_1 |E_1|^2 \right) \left(\int dV \varepsilon_0 \varepsilon_s |E_s|^2 \right)}$$ \hspace{1cm} (5.3)

Note that here the conversion efficiency is defined as the efficiency per unit cell for such an extended surface, hence the volume integration is performed inside a unit cell.

We now apply our optimization framework to discover new all-dielectric three-dimensional metasurfaces, with the permittivity of the medium $\varepsilon_{GaP}$ taken to be that of gallium phosphide (GaP) near telecom wavelengths \[51\ \[400\], and the $\chi^{(2)}$ tensor to be diagonal (realized through rotations) as a proof of principle. Note, however, that the same framework can be easily extended to design plasmonic surfaces. The metasurfaces, illustrated schematically in Fig. 5.1, are square photonic-crystal slabs of in-plane period $\Lambda \times \Lambda$ and finite thickness $t$. To ensure fabricability, here we consider $z$-invariant structures, in which the optimization parameters are taken to lie in the plane perpendicular to $z$-axis, resulting in a structures that can be fabricated...
Table 5.1: Representative second-harmonic generation figures of merit for both hand-and inverse-designed metasurfaces, including $\chi^{(2)}$, fundamental wavelength $\lambda_1$, and conversion efficiency $\eta$ per unit cell. Most of the literature only states the net conversion efficiency for metasurfaces with finite $N$ unit cells, which we convert to the efficiency $\eta$ per unit cell.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$\chi^{(2)}$ (nm/V)</th>
<th>$\lambda_1$ (µm)</th>
<th>$\eta/(\chi^{(2)})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>gold split resonators [66]</td>
<td>250</td>
<td>10</td>
<td>$2.1 \times 10^{11}$</td>
</tr>
<tr>
<td>gold split resonators [457]</td>
<td>1.3</td>
<td>3.4</td>
<td>$3.8 \times 10^{11}$*</td>
</tr>
<tr>
<td>gold cross bars [250]</td>
<td>54</td>
<td>8</td>
<td>$1.4 \times 10^{13}$*</td>
</tr>
<tr>
<td>all-dielectric cylinders [272]</td>
<td>0.2</td>
<td>1.02</td>
<td>$1.6 \times 10^{17}$*</td>
</tr>
<tr>
<td>optimized design [Fig. 5.1]</td>
<td>0.1</td>
<td>1.2</td>
<td>$9.6 \times 10^{24}$</td>
</tr>
</tbody>
</table>

by etching. As a proof of principle, we consider metasurfaces suspended in air, while the same framework can be easily applied to include any substrate [265].

Figure 5.1 shows cross-sections of the unit cell of two GaP metasurfaces of thicknesses $t = 612$ nm and $\Lambda = 480$ nm, designed for operation at a fundamental frequency $\omega_1 = 1.57 \times 10^{15}$ rad/s ($\lambda = 1.2$ µm) so as to satisfy both frequency- and phase-matching conditions. Also shown are the corresponding fundamental and harmonic mode profiles. The structure on the left is optimized for operation at an incident angle $\theta \approx 3.6^\circ$ relative to the out-of-plane axis and is found to exhibit large radiative lifetimes $Q_{1(\text{rad})}^\text{rad} \approx 6(2) \times 10^4$ and overlap factor $|\beta_2|^2 = 1.0 \times 10^{-4}(\chi^{(2)}/\varepsilon_0 \lambda^3)$. The structure on the right is instead optimized for operation at normal incidence, resulting in a slightly smaller $|\beta_2|^2 = 2.5 \times 10^{-5}(\chi^{(2)}/\varepsilon_0 \lambda^3)$. Due to the symmetry of the structure, the modes exhibit infinite lifetimes (and hence are technically dark modes), though in practice, fabrication imperfections necessarily lead to finite lifetimes. Furthermore, Fig. 5.1(b) illustrates the convergence of TO optimization process to achieve structure (i), converged within $\sim 10^3$ iterations. Table 5.1 compares a few of the relevant figures of merits for representative metasurface designs, which include both plasmonic and dielectric structures. Although comparing $\beta_2$ appears to be impossible due to a surprising lack of relevant modal parameters in these studies [66, 457, 250, 272], such as the absence of radiative and dissipative quality factors, we find that the optimized
designs exhibit orders of magnitude larger conversion efficiencies. While it is difficult to distinguish the relative impact of the mode lifetimes and overlap factors, arguably, the optimized structures overcome several limitations associated with previous designs. On the one hand, plasmonic structures exhibit tightly confined modes and therefore lead to large nonlinear overlaps, but absorptive losses and weak material nonlinearity imply that they suffer from small lifetimes. On the other hand, several of the proposed all-dielectric metasurfaces have had negligible material losses and hence larger lifetimes, but have not been designed to ensure large nonlinear overlaps.

5.1.2 Microcavity

To achieve efficient second-harmonic generation in a cavity, the latter must also support two resonances at both frequencies $\omega_1$ and $\omega_s = 2\omega_1$. Consider a typical on-chip setup where the input/output power is guided through a waveguide, and coupled to the cavity mode evanescently. The conversion efficiency, defined in terms of the input power $P_1$ and output power $P_s$ in the small-signal regime, can be conveniently derived from coupled mode equations [265],

$$\eta = \frac{P_s}{P_1^2} = \frac{128}{\omega_1} |\beta_2|^2 Q_s^2 Q_s \left( 1 - \frac{Q_1}{Q_{1,r}} \right)^2 \left( 1 - \frac{Q_s}{Q_{s,r}} \right)$$  \hspace{1cm} (5.4)

where $Q$ and $Q_r$ denote total and radiative quality factors and $\beta_2$ the nonlinear overlap factor defined in Eq. [5.3]. The slight difference between Eq. [5.4] for a cavity and Eq. [5.2] for a metasurface arises from the different coupling mechanisms at hand. To enhance $\eta$, apart from frequency matching and lifetime engineering, one also needs to design the modal profiles at the two frequencies to have large nonlinear overlap $\beta$, and to tune the coupling strength to reach critical coupling condition.

\footnote{Note the volume integration will be instead over the whole space.}
We now apply our optimization framework to a GaP on oxide integrated photonic platform, currently being pursued by our experimental collaborators at the U. of Washington. The structure aims to build on our recent experimental demonstration of high conversion efficiency in robust, compact, and wide bandwidth integrated cavities \[282\]. More explicitly, as illustrated schematically in Fig. 5.2(a), we consider a 250 nm thick layer of (100) GaP on a thermal SiO\(_2\) substrate. The \(\chi^{(2)}\) tensor of the GaP in the zincblende crystal phase is off-diagonal, allowing nonlinear interactions only between mutually perpendicular field components. Consequently, the fundamental mode must be TE-like (E-field is mostly in-plane) and the second-harmonic mode TM-like (E-field is mostly out-of-plane). Under such a specific modal requirement and given the thin thickness of GaP \((t/\lambda_1 \approx 0.16)\), it is already challenging to confine...
light to a wavelength-scale cavity at $\omega_1$ via either traditional design principles or the previous two-dimensional inverse design strategies \cite{265}, not to mention the further need to enhance the overlap factors. Our optimization formulation in three dimensions, on the other hand, can easily discover compact cavities that tightly confine fields to the appropriate modes that also exhibit a large nonlinear overlap.

Fig. 5.2(a) depicts an optimized microcavity that supports a TE-like mode at the fundamental wavelength $\lambda_1 = 1550$ nm, and a TM-like mode at the second-harmonic wavelength $\lambda_s = 775$ nm. The cavity is $z$-invariant for the ease of fabrication, and has wavelength-scale length and width $L_x = L_y = 3\lambda_1$ in the other two dimensions. Its dielectric profile in the $z = 0$ cross section is shown in Fig. 5.2(b), with a minimum feature size $\gtrsim 60$ nm. Eigenmode analysis shows that the cavity is able to support modes of moderate lifetimes $Q_{1(2)} = 3000(1000)$ at $\omega_{1(2)}$, enabling fast operating timescales. The mode profiles along $x - y$ cross section (left) and $y - z$ cross section (right) are shown in Fig. 5.2(c), illustrating tight confinement of light along $z$-direction, and the effort of the optimization algorithm to arrange the field distributions to enhance the overlap factor, approaching $|\beta_1| = 0.01\frac{x^{(2)}}{\sqrt{\epsilon_0}\lambda_1^3}$. The resulting $\beta_1$ factor is at least two orders of magnitude larger than that of typical ring resonators \cite{282}, making it possible for the resonator to have high efficiency $\eta \approx 1.6 \text{ W}^{-1}$, assuming critical coupling, even with moderate $Q_s$.

5.1.3 Concluding Remarks

We have demonstrated a novel optimization approach for the design of nonlinear metasurfaces and microcavities in three dimensions. The optimized structures exhibit order-of-magnitude larger overlap factors than traditional designs. Inverse design not only overcomes efficiency limitations associated with traditional microcavities and metasurfaces but also greatly reduces challenges and difficulties inherent to the design process.
5.2 Compact multimode cavity couplers

In this section, we show that large-scale optimization algorithms can be applied to design compact on-chip devices that efficiently couple light consisting of multiple, widely separated wavelengths from a single waveguide into a wavelength-scale multi-resonant cavity. Motivated by practical problems in nonlinear optics, we pursue three illustrative examples: compact multi-resonant cavities with resonant features mimicking those used for second-harmonic, sum-frequency, and frequency comb generation. In each situations, we demonstrate either total or near total critical coupling.

5.2.1 Overview

To operate on-chip, each mode in a device making use of these cavities must be coupled to a source or detector in a controlled way; and until presently, we have not addressed how this can be done. Using a typical evanescent scheme, tuning the gap separation to control evanescent overlap between the waveguide and cavity [79], realization of high efficiency devices using these cavities, and similar future designs, may be difficult. Beyond the issues of layout intricacy, bending loss [439, 142], and waveguide crosstalk [114, 198] that would be introduced by requiring multiple waveguides to intersect in a wavelength-scale area, modes in the best performing cavities designs may be tightly confined to the core [267], precluding the possibility of achieving critical or over coupling by simply decreasing the separation.

This problem of efficiently coupling light between sources and predefined volumes appears in many branches of nanophotonics. For instance, it is the defining goal of wide-area absorbers—surfaces that can perfectly absorb a wide range of incident propagating waves. Broadly, the main approach in this setting is to create structures supporting many resonances in order to tune the radiative and absorptive decay rates in each scattering channel [152]. This behaviour can be introduced in a
wide variety of ways, including adiabatic tapers [461, 263], metasurfaces [13, 278], epsilon-near-zero thin films [316, 369], chirped gratings [407, 203], multi-resonant photonic crystals [373], and more recently, unintuitive structures obtained via inverse design [145, 419, 140]. A similar objective also appears in the context of free-space to on-chip couplers, with the primary aim being to reduce losses, i.e. reflections, of light incident on a on-chip device from either a fiber or free space. Rate matching is more difficult to implement in these situations, as any signal decay (e.g. material absorption in the coupling region) reduces performance; and common approaches based on adiabatic tapers lead to couplers that are several wavelengths long and are only typically designed to operate over narrow, selective bands [6, 69, 141, 427]. Based on motivations similar to those of this present study, there is a current push to exploit inverse design [329, 293], metasurface concepts [481, 258, 253], and chaotic deformations [205] in this area. Likewise, a need to control coupling arises between on-chip devices, including filters, rectifiers, multiplexers, and frequency converters. In these situations, the usual goal is to efficiently couple two or more separately designed devices in the smallest possible footprint. Again, much in the spirit of the results presented here, within the past few years inverse design approaches have started to be applied in this setting, leading to experimental demonstrations of compact wavelength-division multiplexers operating over several far-apart wavelengths (\(\Delta \lambda / \lambda \gtrsim 5\%\)) [352, 136].

### 5.2.2 Formulation

Our conception of the coupling problem is depicted in Fig. 5.3. Starting from an isolated cavity supporting \(N\) resonances with frequencies \(\omega_i\) and radiative lifetimes \(Q_{i,r}\), \(i = \{1, 2, \cdots, N\}\), we aim to design a wavelength-scale device that tunes the external coupling rate of each mode to a single nearby waveguide to any desired value. That is, we seek to independently control the dimensionless coupling quality factor \(Q_{i,c}\) of every individual mode of a given set. Generically, the presence of a coupler
Figure 5.3: Schematic of a general coupler (black region), which couples light between a wavelength-scale, multimode cavity and a multimode port (waveguide). The design freedom of the scatter enables controllable coupling between the two devices at several wavelengths (red, yellow, and green arrows).

or a nearby waveguide can significantly alter the radiative decay of an isolated cavity modes, either enhancing or degrading temporal confinement. To ensure that the resonant features of the cavity are not destroyed by the coupler, we simultaneously constrain $Q_{i,r}^0 \leq \alpha_i Q_{i,r}$, with $Q_{i,r}$ denoting the radiative quality factor of the cavity in the presence of the waveguide and coupler, and $\alpha_i$ an adjustable scale factor. Based on this description, a structure for any desired collection of coupling characteristics is discoverable using a gradient adjoint-variable topology optimization approach [260] that seeks to solve the minimax problem,

$$
\min_{\{\varepsilon\}} \mathcal{F}(E, \varepsilon) \tag{5.5}
$$

s.t. $\{\mathcal{G}_i(E, \varepsilon) \leq 0\}, \ i = 1, 2, \ldots, N$

$$
\varepsilon_{\text{sub}} \leq \varepsilon \leq \varepsilon_{\text{st}},
$$

where

$$
\mathcal{F}(E, \varepsilon) = \max_{i=1}^{N} [Q_{i,c}(E, \varepsilon) - \xi_i Q_{i,r}(E, \varepsilon)]^2 \tag{5.6}
$$

$$
\mathcal{G}_i(E, \varepsilon) = Q_{i,r}^0 - \alpha_i Q_{i,r}(E, \varepsilon),
$$

104
with $\xi_i$ denoting the target ratio of $Q_{i,c}/Q_{i,r}$. In this method, the dielectric permittivity at every spatial point inside the coupling region, $\{\varepsilon\}$, is as a continuous degree of freedom, bounded by the substrate $\varepsilon_{\text{sub}}$ and structure $\varepsilon_{\text{st}}$ materials. (To produce binary, smooth, fabricable systems additional regularization and filter projection steps are applied in conjunction with this base algorithm [449].) In order to circumvent numerical issues associated with optimizations of electromagnetic eigenvalues [260], each $Q$ is computed by solving a set of scattering problems. This makes both the objective $F$ and constraints $G_i$ explicit functions of the electric field $E$, computed as the solution of the steady-state equation

$$\left[ \nabla \times \frac{1}{\mu} \nabla \times -\omega_i^2 \varepsilon(\omega_i, r) \right] E(\omega_i, r) = i\omega J(\omega_i, r).$$

To setup this problem, electric current sources, the duals $J(\omega_i) \propto \text{Re} \left[ \varepsilon(\omega_i) E^*(\omega_i) \right]$ of the modes in the energy norm [82], are first calculated (without the waveguide and coupler) at each individual frequency. The waveguide and coupler are then added, and the field quantities of interest determined: the electromagnetic energy density inside the cavity volume, $U_i = \frac{1}{2} \int_V dV \varepsilon(\omega_i, r) |E(\omega_i, r)|^2$, and the Poynting flux into the waveguide and radiated into vacuum, $P_i = \frac{1}{2} \int_\Sigma ds \cdot \text{Re}[E(\omega_i, r)^* \times H(\omega_i, r)]$, with $\Sigma$ denoting the corresponding flux surfaces. The radiative and coupling lifetimes $Q_{i,c(r)} = \omega_i U_i / P_{i,c(r)}$ are then used to evaluate $F$ and $G_i$.

In many applications of interest, one of two coupling characteristics are often desired: over coupling [168, 349], minimizing unwanted losses and increasing energy efficiency, or critical coupling, maximizing field amplitudes in the cavity [265]. For cavities designed to enhance nonlinear frequency conversion, such as the illustrative examples considered below, maximum power conversion occurs under critical coupling, $\xi_i = 1$, at each frequency. In such cases, the general scheme presented above can be simplified. When the cavity is pumped from a single channel with power $P_{\text{in}}$, the energy in the cavity is related to the quality factors by [418]

$$\frac{\omega_i U_i}{P_{\text{in}}} = \frac{4 Q_{i,r}}{2 + Q_{i,r}/Q_{i,c} + Q_{i,c}/Q_{i,r}},$$

(5.7)
reaching a relative maximum of $Q_{i,r}$ as the system moves toward critical coupling ($\xi_i \to 1$). Technically, Eq. 5.7 is only applicable to unidirectional couplers, i.e. when each cavity mode couples only to one port (direction) of the waveguide. (A simple example of a unidirectional coupler is the usual waveguide–ring resonator system, where the direction of coupling is constrained by momentum conservation [156].) However, as the introduction of any additional coupling channel always reduces the energy stored in the cavity [118], in practice, there is no loss of generality in considering this expression. By maximizing the energy in the cavity, the algorithm naturally proceeds towards unidirectional couplers, which in turn makes Eq. 5.7 an increasingly good approximation. Since the behavior of Eq. 5.7 is then ultimately equivalent to the more complicated Eq. 5.6 we are able to consider the simpler optimization problem,

$$F'(\bar{\varepsilon}) = \max_{\bar{\varepsilon}} \left\{ \min_{i=1}^{N} \left[ \frac{U_i(E, \bar{\varepsilon})}{U_i^0} \right] \right\}.$$  \hfill (5.8)

Where the $U_i^0 = Q_{i,r}^0 P_i^{in}/\omega_i$ is an energy normalization factor given by the bare radiative lifetime.

As a proof of concept, we consider two illustrative cavities designed to enhance two $\chi^{(2)}$ nonlinear processes: up-conversion of $\omega_1$ and $\omega_2$ to the summed frequency $\omega_s = \omega_1 + \omega_2$ (SFG), and second-harmonic generation (SHG) corresponding to degenerate SFG with $\omega_1 = \omega_2$. For these processes, the relative coupling rates largely dictate the achievable intensities in the cavity, and hence power requirements (in the undepleted regime [265]). Mathematically, this is captured by the figure of merit

$$\text{FOM} = |\beta|^2 \prod_{i=1,2,s} \frac{Q_{i,r}}{2 + Q_{i,r}/Q_{i,c} + Q_{i,c}/Q_{i,r}},$$  \hfill (5.9)

with $\beta$ denoting the overlap coefficient of the cavity fields, which to first order is not affected by the external waveguide or coupler. Like Eq. 5.7, Eq. 5.9 is maximized when all three modes achieve critical coupling, giving $\text{FOM}_{\text{max}} = |\beta|^2 Q_{1,r} Q_{2,r} Q_{s,r}/64$.  

106
5.2.3 Coupler in two dimensions

Figure 5.4: Optimized coupler for SHG showing critical coupling between the double-ring resonators and the waveguide. All structures (black) are made of GaP, while the substrate (white) is assumed to be vacuum. The width of the waveguide is 150 nm, the diameter of the outer ring 2.6 µm, and the area of the designed coupling region 3.75 µm × 1.5 µm. (c) shows the energy spectrum inside the resonator near the fundamental and second-harmonic wavelengths $\lambda_{\{1,s\}} = \{1500, 750\}$ nm, with matched azimuthal wavenumbers $m_1 = 8$ and $m_s = 2m_1$, (black, red) normalized by $U^0$. The middle figures show the TM-polarized electric fields at the respective wavelengths. The complete suppression of outgoing/transmitted power through the waveguide provides a visual confirmation of critical coupling.

As a platform for testing our algorithm, we begin by considering a two-dimensional system consisting of hand-designed multi-track ring-resonators supporting TM-polarized resonances of moderate radiative lifetimes $Q^0_{i,r} \lesssim 10^5$, and a rectangular admissible coupler region covering the separation between the cavity and the waveguide. The size of this design region is determined on a case-by-case basis as a compromise between compactness and functionality. Starting from a base of 3.75 µm × 1.5 µm the size of the coupling region is increased whenever the algorithm is unable to find suitable coupling structures. To guide the algorithm towards more easily fabricable structures, the coupler is always seeded with connected, smooth dielectric profiles, i.e. a random ribbon. (For numerical accuracy, the grid resolution is chosen to be smaller than $\lambda/45$ for the smallest wavelength considered, $\approx 13$ pixels per wavelength inside the highest index media.)

Our findings begin with the SHG and SFG systems depicted Fig. 5.4 and Fig. 5.5. For practical considerations, these simulations suppose a gallium phosphide material
Figure 5.5: Optimized coupler for SFG showing near critical coupling between the triple-ring resonators and the waveguide. The middle figures show the TM-polarized electric fields at the wavelengths \(\lambda_{1,2,s} = \{1500, 907, 565\}\) nm, \(m_{1,2,s} = \{9, 20, 28\}\). The width of the waveguide is 134 nm, the diameter of the outer ring 2.8 \(\mu\)m, and the area of the coupling region is 5.4 \(\mu\)m \(\times\) 2 \(\mu\)m.

(including material dispersion) for all dielectric regions: the cavity, waveguide, and coupler. The initial SHG system is a two-track multi-ring supporting TM-polarized resonances at \(\lambda_{1,s} = \{1500, 750\}\) nm, quality factors of \(Q_{1,s,r}^0 = \{1.4, 4.6\} \times 10^3\), with power coupled into the device through a narrow waveguide at a gap separation of 1.5\(\lambda_1\). (The azimuthal numbers of these modes satisfy the phase-matching condition \(m_s = 2m_1 = 16\) for the (111) plane of a GaP crystal. Given a different nonlinear tensor and requisite polarizations, the phase-matching condition for \(m\) can be slightly different \[39\].) As indicated in the field profiles of Fig. 5.4, in the presence of the coupler each mode shows vanishing transmission and reflection (\(\lesssim 2\%\)), and large field amplitude inside the cavity. Quantitatively, Fig. 5.4(c) examines the energy spectrum inside the resonator channeled from the waveguide around \(\lambda_{1,2}\), normalized by \(U_1^0\). After optimization, the cavity mode lifetime is more than doubled, with \(Q_{1,s,r}/Q_{1,s,r}^0 = \{2.9, 2.2\}\). As expected, eigenmode analysis reveals the system to be totally asymmetric, with the cavity coupling exclusively to the lower waveguide (downwards propagation). The coupler is also observed to be both binary and smooth, having no feature smaller than 120 nm. Nearly identical results are seen for the triply resonant system (non-degenerate SFG) illustrated in Fig. 5.5. Moving to a three-track cavity designed to support modes at \(\lambda_{1,2,s} = \{1500, 907, 565\}\) nm, with
$Q^0_{1,2,s},r = \{640, 5.3 \times 10^4, 3.2 \times 10^4\}$, the algorithm is again able to realize critical coupling at all three wavelengths, resulting in transmission $\lesssim 1\%$. Cavity radiative lifetimes are also similarly enhanced, with $Q_{1,2,s},r/Q^0_{1,2,s},r = \{2.2, 1.7, 3.1\}$. In either inverse design, the coupling mechanism is found to be more intricate than just the overlap of evanescent fields used for single wavelengths. This is most pointedly seen in (2) and (3) of Fig. 5.5, where over 99.5\% of the energy density is in the cavity, yet critical coupling occurs at the two wavelength of separation due to the fields in the coupler. Moreover, for some cavity modes, even at a single wavelength, it would not be possible to achieve critical coupling using the evanescent tails of a waveguide mode. For example, due to its tight confinement to the inner ring, even if the waveguide is made to touch the cavity, it is not possible to couple to the mode displayed in Fig. 5.4(1) with better than 70\% efficiency. (Reducing the waveguide cross-section offers no improvement due to the creation of phase mismatch.) Note that the radiative quality factors of the cavities we have designed are smaller than those typically considered for nonlinear processes. This choice was made primarily to test the algorithm in cases involving dissimilar waveguide and cavity mode profiles. Nevertheless, we note that for equivalent nonlinear performance, larger overlaps $\beta$ and smaller radiative lifetimes are often preferable to higher quality factors.

As a final benchmark, Fig. 5.6 demonstrates a system attaining near critical coupling at 6 frequencies (over an octave), a frequency comb with large tooth spacing. (A more practical frequency comb coupler, e.g. exhibiting critical coupling at over 100 frequencies, will be considered in future work.) In this case, we begin with a wavelength-scale ring resonator having unevenly distributed modes $m = \{5 \rightarrow 10\}$ at intervals $\Delta f = \{0.101, 0.099, 0.096, 0.098, 0.096\}c/1.5\mu m$, with $f_m = 2c/(4.5\mu m) + \sum_{i=0}^{m-5} \Delta f_i$. To simplify future comparisons, material dispersion is ignored and the cavity, waveguide and coupler are all assumed to have a constant permittivity of $\varepsilon = 9.3514$. As opposed to our first two examples, where the modal resonance fre-
Figure 5.6: Optimized coupler for comb generation showing near critical coupling over 6 frequencies. The width of the waveguide is 300 nm, the diameter of the ring 1.8 µm, and the area of the designed coupling region 4.5 µm × 4.5 µm. The figures show the TM-polarized electric field profiles at the respective azimuthal number \( m = \{5 \rightarrow 10\} \), corresponding to frequencies \( f = \{0.667 \rightarrow 1.157\} \) c/1.5 µm, with equal spacing \( \Delta f = 0.098 \) c/1.5 µm.

Frequencies are fixed constraints, adding a degree of modularity to the total system design, here, we assume that the resonance frequencies must be tuned by the coupler. This conceptual shift requires some small modifications to stabilize the optimization algorithm. Namely, we now include an initial phase where the usual energy density objectives are replaced by field overlap integrals with the eigenmodes of the cavity 

\[
\int_V dV \text{Re} [\varepsilon(\omega_i, \mathbf{r}) \mathbf{E}_m^*(\omega_i, \mathbf{r}) \cdot \mathbf{E}(\omega_i, \mathbf{r})],
\]

where the \( m \) subscript denotes the \( m^{th} \) mode of the bare cavity, and, as before, the integration is restricted to lie within the outermost material boundary of the cavity. The frequencies \( \{\omega_i\} \) where these computations are carried out are initialized to match those of the bare cavity, and then slowly transitioned to the desired resonances, i.e. an evenly distributed set. The converged output of this procedure is effectively a new cavity having characteristics well-matched to the original coupling optimization algorithm. The number of iterations needed for this optimization was roughly the same as those of the previous examples, leading to approximately linearly scaling of the total computation time with the number of frequencies. The coupler displayed in Fig. 5.6 achieves
the desired wavelength tuning and critical-coupling functionality. The resonance frequency intervals are equally distributed as $\Delta f = \frac{0.098c}{1.5\mu m}$, and good coupling (transmission and reflection below 15%) is visibly present in each of the field profile plots. Explicitly, the summed transmitted and reflected powers of are found to be $\{2\%, 5\%, 13\%, 2\%, 1\%, 4\%\}$. However, smaller minimal feature sizes, $\approx 15$ nm, and a larger total footprint $4.5 \mu m \times 4.5 \mu m$ were required to achieve these effects.

When extending our work to a practical comb over $10^2$ frequencies, one will certainly need a much larger coupler. However, its size might still be comparable to that of the large resonator present to support such a comb in the first place.

5.2.4 Coupler in three dimensions

Figure 5.7: (a) On top of a silica substrate (gray) is a microcavity optimized for both large nonlinear overlap factors and critical coupling to the same waveguide at both frequencies. The cavity has a finite thickness $t = 250$ nm in the $z$-direction, and $L_x = 4.65 \mu m$, $L_y = 6.2 \mu m$ along the other two dimension. The waveguide has the same thickness $t = 250$ nm, and a small width $w = 465$ nm. Shown in (b) is its dielectric profile over the $x-y$ cross section, where black (white) represents gallium phosphide (vacuum) regions. The mode profiles along the $x-y$ cross section at both fundamental ($\omega_1$) and second harmonic ($\omega_s$) frequencies are shown in (c), where we are plotting the $E_x$ ($E_z$) component at $\omega_1(\omega_s)$.

In Sec. 5.1.2 we show that inverse design can be applied to discover a three-dimensional compact cavity for high-performance SHG. However, the mode profiles of
inverse designed structures are far more complicated than that of ring resonators, and there is not yet an established approach to couple energy efficiently to those modes. Our optimization formulation in Sec. 5.2.2 is general and can be readily applied to address such coupling problems. Here for the sake of more compact devices, we pursue a slightly different formulation that, instead of separately treating the cavity design and the two coupling problems, can simultaneously optimize the cavity and the coupler regions, via directly maximizing the conversion efficiency. As described in Eq. 5.4, the output power $P_s$ at $\omega_s$, guided away along the waveguide, maximizes when both the cavity and the coupler approach the optimal design criteria. In the small signal region, $P_s$ can be approximately computed by coupling the two linear Maxwell’s equations at the two frequencies, as introduced in Eq. 5.1. More explicitly, the optimization algorithm is as follows,

\[
\max \mathcal{F} = \frac{1}{2} \text{Re} \int_{wvg} dS : [E(\omega_s) \times H^*(\omega_s)]
\]

\[
\mathcal{M}(\varepsilon, \omega_1)E_n = i\omega_1 J_n, J_n: \text{excite waveguide mode at } \omega_1,
\]

\[
\mathcal{M}(\varepsilon, \omega_2)E(\omega_2) = i\omega_s J(\omega_s), J(\omega_s) = x \prod_{n=1}^{2} E_n \hat{e}_t,
\]

\[
\mathcal{M}(\varepsilon, \omega) = \nabla \times \frac{1}{\mu} \nabla \times -\varepsilon(r)\omega^2,
\]

where $x \in [0, 1]$ is the design variable describing vacuum (GaP) at value 0(1), the surface integral in the objective $\mathcal{F}$ is over a cross section of the waveguide away from the device to compute $P_s$, and the current source $J_n$ is chosen to excite the waveguide mode at $\omega_1$ (in the absence of the cavity) to simulate $P_1$. While in principle the optimization can start from any initial structures, e.g. random ribbons, to approach a high local optimal, it is preferred to seed the algorithm with an optimized microcavity to ensure the existence of two appropriate modes at $\omega_1$ and $\omega_s$. 

112
We now apply the formulation to the same platform described in Sec. 5.1.2, where a 250 nm thick GaP wafer can be nanostructured into various optical components such as waveguides and nonlinear cavities, sitting on top of a SiO$_2$ substrate. As schematically shown in Fig. 5.7(a), we consider a single port system where a single waveguide guides both the input and output powers. The waveguide is assumed to pass through the center of the cavity, allowing for larger field overlaps. Fig. 5.7(b) depicts the dielectric profile of an optimized structure along the $z = 0$ cross section, which has the same width $L_x = 3\lambda_1$ as the previously optimized cavity, but slightly longer $L_y = 4\lambda_1$ that plays the role of coupling. The minimal feature size is $\gtrsim 60$ nm, corresponding to an aspect ratio of 4 that is within many of current experimental capacities. The efficient coupling between the nonlinear device and the waveguide is visibly apparent by inspecting the modal profiles of the entire structure, as shown in Fig. 5.7(c). The mode is highly confined inside the resonator, and dissipates the energy into the waveguide and far-field in close quantities (critical coupling corresponds to equal dissipations into the waveguide and far-field). A numerical analysis demonstrates that the coupling efficiency is $\gtrsim 90\%$ at both frequencies.

5.2.5 Concluding Remarks

In summary, we have shown that, in both two and three dimensions, inverse design provides a practical means of efficiently coupling light at multiple widely separated wavelengths from a single channel (a waveguide) into a compact, multimode cavity. Drawing from our recent work on the design of compact microcavities for high-efficiency nonlinear frequency conversion, we have successfully treated suggestive examples for second-harmonic (SHG), sum-frequency (SFG), and frequency comb generation (albeit for large tooth spacing). Critical coupling was achieved, or nearly achieved, at all relevant wavelengths without incorporating sharp components in the first two cases. Our results continue the promising trend seen in application of inverse
design to free-space and on-chip couplers, rectifiers, and multiplexers, indicating the potential of these techniques to enable significant improvements in integrated nonlinear photonics.
Chapter 6

Conclusion and Outlook

Recent advances in inverse-design approaches have significantly broadened the scope of nanophotonics to include large bandwidth or multi-frequency applications [318, 254]. In this thesis, we’ve demonstrated the versatility of inverse design in several challenging topics. First, we’ve introduced several powerful numeric techniques to efficiently compute near-field radiative heat transfer (NFRHT) between bodies of arbitrary shapes and dielectric profiles, enabling us to systematically study the optimization of frequency-selective NFRHT between two multilayer objects or grating structures. In the low-loss multilayer scenario, we showed that inhomogeneous dielectric profiles can allow perfect absorption over a wide range of wavevectors, resulting in a logarithmic enhancement factor depending on material losses. When restricted to highly lossy media, we find that, NFRHT between optimized grating structures, which admit more design variables, can be boosted to fall within factors of order unity of that between ideal planar bodies.

Next, built upon earlier work of simple two dimensional optimizations of nonlinear cavities [265, 267], we’ve addressed some practical on-chip integration concerns by introducing a set of on-chip components that are optimized in three dimensions for efficient nonlinear frequency conversions, including wavelength-scale microcavi-
ties and couplers. For the sake of robustness and fast operational timescales, the inverse designed microcavity trades off temporal confinement for nonlinear field overlaps, enabling nearly state-of-art nonlinear conversion efficiencies upon modes of short radiative lifetimes $Q \lesssim 10^3$. To operate on integrated platforms, we showed that optimization algorithms can discover compact couplers that efficiently couple light from a single waveguide into those wavelength-scale multi-resonant cavities at all participating frequencies. Beyond those components studied in this thesis, extending inverse design to active devices such as modulators and lasers, which are the performance limiting components of many optical systems, would also be extremely useful. Another direction is to explore even broader band applications, such as efficient coupling to frequency combs at hundreds of frequencies.

One more exciting direction is that the optimization framework not only promises practical impacts in locating structures of high performance, but also provides guidelines to explore fundamental limitations on device performance, e.g. upper bounds on NFRHT. The underlying reason is that inverse-design methods can, at least in principle, explore the full space of fabricable devices, and approach nearly globally optimal solutions by making use of appropriate regularizations. This capability opens the possibility of estimating an achievable maximum theoretical performance of an optical device. For instance, our optimization of NFRHT was initially motivated by a theoretical bound [313] that suggests NFRHT can be potentially improved by many orders of magnitude via nanostructuring in large-area bodies. However, the optimization results in Chapter 4 fail to fill in such a huge gap, which in turn strongly inspired people to inspect the tightness of the bound. Indeed, motivated by our results, recently we derived a much tighter bound [438], which surprisingly suggests that NFRHT from our grating structures are already within factors of order unity of the tighter bound.
To further advance inverse-design methods in practical applications, there are at least following few key factors to consider in the future work. First and foremost is to improve the robustness of design methods to handle process variations in photolithography, which would enable high throughput fabrication. This factor is particularly relevant in situations where the fabrication errors are so large that even the perturbation theories fail to provide any guidelines. While for traditional devices, the direction to overcome large errors can come from intuitions or simple theories built upon highly symmetric geometries, for those seemingly “irregular” structures from optimizations, constructing a library of selected inverse designed structures of similar patterns and performance can be helpful to provide similar intuitions for the tunability. Next, improvements to the underlying simulations and optimization algorithms could enable design of larger devices, greatly improving the breadth and scope of problems that can be tackled by inverse design. Along these lines, several recent works have begun exploring applications of machine learning in nanophotonics [482], paving the way, potentially, for improvements in fast iterative Maxwell solvers. In the enduring quest for optimal photonic designs, the widespread integration of inverse-design tools seems not only sensible but unavoidable.
Appendix A

Material properties of GST alloy

A.1 Material parameters

The Ge$_2$Sb$_2$Te$_5$ alloy is a phase-change chalcogenide glass that exhibits a large thermo-optic effect [385] and three possible (amorphous, cubic, and hexagonal) phases corresponding to transition temperatures of 438 K (separating the amorphous and cubic phases) and 623 K (separating the cubic and hexagonal phases) [256, 464]. Because there are yet no experimental characterizations or semi-analytical models of the dielectric dispersion $\varepsilon_{\text{GST}}(\omega, T)$ of the GST from 300 K to its melting point 870 K [428], we instead model the dispersion via a simple linear-interpolated fit of available experimental data at five different temperatures (spanning amorphous, cubic and hexagonal phases) [256, 401]. Figure A.1 shows both the real (red solid line) and imaginary (blue dashed line) parts of $\varepsilon_{\text{GST}}$ at a single wavelength $\lambda = 5.8\mu$m over this temperature range (with circles denoting experimental data). Together with the temperature profiles of the structures $T(\vec{x})$ and dispersion relations of Ti [300], Si$_3$N$_4$ [224], and AZO [220], this provides all of the information needed to perform the calculations of thermal radiation from the bodies of Fig. 2.2 in the main text. On the other hand, Fig. 2.4 of the main text explores single-frequency radiation from bodies with piece-
wise constant temperature profiles (constant $T = 870$ K in the GST and 300 K in the remaining regions), which allows us to employ typical permittivity values for these materials at mid-infrared wavelengths; specifically, we choose $\varepsilon_{\text{GST}} = 30 + 10i$ [401], $\varepsilon_{\text{Ti}} = -100 + 80i$ [300], $\varepsilon_{\text{Si}_3\text{N}_4} = 5 + 0.1i$ [224], and $\varepsilon_{\text{AZO}} = -25 + 15i$, corresponding to a doping density $\approx 1\text{wt}\%$ [220].

![Figure A.1: Real (solid red) and imaginary (dashed blue) permittivity $\varepsilon_{\text{GST}}(T, \lambda)$ of a bulk GST glass at $\lambda = 5.8\mu\text{m}$, obtained via simple linear interpolation of experimental data at multiple temperatures (circles).](image)

A.2 Temperature gradients

Interfaces play a crucial role in nanoscale thermal transport. For instance, they enable thermal boundary resistance (TBR) to radically alter the surrounding temperature distribution [368, 304, 413, 295], leading to small-scale thermal discontinuities across the interface. TBR consists of both contact and intrinsic “Kapitza” resistance, with the former arising from poor mechanical connection between materials (due to surface roughness) and the latter from acoustic mismatch between materials (and hence persisting even under perfect-contact situations) [413]. Typical values of intrinsic resistance at room temperature are on the order of $10^{-9} \sim 10^{-7}\text{m}^2\text{W/K}$ [413], whereas those arising from contact resistance vary depending on the surface and thermophys-
ical properties of the intervening medium. In our setup (described schematically in
Fig. 2.1 of the main text), there four interfaces at which TBR can arise. These are de-
noted and described by the resistances $R_{sh|c}$, $R_{h|su}$, $R_{sh|su}$, and $R_{h|c}$, of the shell–GST,
heater–substrate, shell–substrate, and heater–GST interfaces, respectively. Note that
the thermal resistance associated with graphene can be made extremely small \[394\]
and hence in our calculations, we assume negligible $R_{h|c} = 0$. In order to obtain
large temperature gradients, it is important to operate with materials that can dissi-
pilate heat away from the shells rapidly \[196\]; hence, we assume small shell–substrate
interface resistances $R_{sh|su} = 10^{-8} \text{m}^2\text{W/K}$. For simplicity, we consider conditions
under which the interface resistances $R_{sh|c} = R_{h|su} = R_{th}$ are equal and obtain var-
ious degrees of temperature localization by varying $R_{th}$, with $R_{th} = \infty$ leading to
perfect temperature-localization and $R_{th} = 0$ leading to uniform temperature distri-
butions. In particular, we consider five different operating conditions, corresponding
to realistic values of (i) $R_{th} = 0.5 \times 10^{-7} \text{m}^2\text{W/K}$, (ii) $R_{th} = 10^{-7} \text{m}^2\text{W/K}$, and (iii)
$R_{th} = 2 \times 10^{-7} \text{m}^2\text{W/K}$ and extreme (unrealistic) values of (iv) $R_{th} = \infty$, and (v)
$R_{th} = 0$.

![Figure A.2](image)

Figure A.2: Temperature distribution $T(\vec{x})$ of the Ti (left) and Si$_3$N$_4$ (right) hemi-
spheroid composites described in Fig. 1. Both structures rest on a SiO$_2$ substrate
(thickness 0.3µm and radius = 1.5× shell radius) whose bottom surface is in contact
with a heat reservoir at 300 K. All other surfaces are exposed to vacuum and there-
fore satisfy adiabatic boundary conditions; material interfaces on the other hand are
subject to thermal boundary resistance in accordance with operating condition (iii)
and (ii) described in the text, for the left and right body, respectively.

In order to solve the heat–conduction equation to obtain the steady-state tem-
perature distribution $T(\vec{x})$, one must also specify the boundary conditions associated
Table A.1: The thermal properties used in our COMSOL simulation

<table>
<thead>
<tr>
<th>Material</th>
<th>k(W/m/K)</th>
<th>C(J/kg/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ti</td>
<td>21.9</td>
<td>523</td>
</tr>
<tr>
<td>Si3N4</td>
<td>40</td>
<td>1100</td>
</tr>
<tr>
<td>GST</td>
<td>κ(T) = 208.3</td>
<td></td>
</tr>
<tr>
<td>SiO2</td>
<td>1.38</td>
<td>703</td>
</tr>
</tbody>
</table>

with vacuum–material interfaces, which we assume to be adiabatic ($\nabla T \cdot \vec{n} = 0$), corresponding to negligible conduction, convection, and radiative-heat dissipation through air. The substrate is chosen to be a 0.3 µm thick SiO$_2$ film in contact with a heat reservoir at 300 K through the bottom interface, leading to large heat dissipation away from the shell and hence large temperature localization in the GST with decreasing substrate thickness. We choose the substrate lateral (cylindrical) dimensions to be large enough to remove large thermal diffusion away from the GST. Figure A.2 shows the temperature distribution of both Ti (left) and Si$_3$N$_4$ (right) structures under the operating condition (iii) and (ii), respectively, assuming the material conductivities and heat capacities given in Table I. As shown, the temperature in the substrate and shells is almost uniform and close to 300 K, thanks to the presence of boundary resistance between the heater and substrate (which bars heat from flowing into the substrate) along with the high thermal conductivities of Ti and Si$_3$N$_4$, which act to dissipate heat away from the GST.

A.3 Geometries

Figure A.3 provides the size and dimensions of the various geometries explored in Fig. 2.4 of the text.
Figure A.3: Parameter descriptions for the bodies associated with Fig. 2.4 of the main text.
Appendix B

Randomized singular value decomposition for near-field radiative heat transfer

B.1 Approximate Singular Value Decomposition

In this section, we sketch how the low-rank nature of the RHT matrix $Z^{AB} = \sqrt{\text{Im}[\chi^A]} G^{AB} \sqrt{\text{Im}[\chi^B]}$ entering (6) in the main text allows application of fast randomized singular value decompositions (SVD), greatly speeding up calculations of RHT. We begin by splitting the problem into two domains, referred to as bodies $A$ and $B$, with associated superscripts denoting projection. It will be assumed that sources occur in body $A$ and that the fields of interest lie only in body $B$. (From Lorentz reciprocity, the Green function must be symmetric under the exchange of observation and source positions, and so there is no loss of generality in either choice.)

The algorithm, described in detail in [174], proceeds iteratively as follows.

Draw a random Gaussian distributed current vector $\tilde{j}_{i+1}$ with values in body $A$. Let $\tilde{\Omega}_{i+1}^A$ denote the set of all previously drawn randomly Gaussian distributed current...
vectors (vertically concatenated to form a matrix). Solve Maxwell equation’s to obtain the associated field, \( \tilde{E}_{i+1} = \mathcal{G} \tilde{j}_{i+1} \). Let \( \tilde{O}_i^B \) denote the set of all previously computed, orthonormalized, electromagnetic field vectors, and \( \tilde{Y}_i^B \) the set of all previous electromagnetic fields as originally calculated. Compute \( \tilde{o}_i^{uB} = (I - \tilde{O}_i^B \tilde{O}_i^{B\dagger}) \tilde{E}_{i+1} \), and normalize the result \( \tilde{o}_i^{uB} = \tilde{o}_i^{uB} / \langle \tilde{o}_i^{uB} | \tilde{o}_i^{uB} \rangle \), recording the value of \( \epsilon_{i+1} = \langle \tilde{o}_i^{uB} | \tilde{o}_i^{uB} \rangle \).

Concatenate \( \tilde{o}_i^{uB} \) onto \( \tilde{O}_i^B \) and \( \tilde{E}_{i+1} \) onto \( \tilde{Y}_i^B \), producing \( \tilde{O}_i^{B+1} \) and \( \tilde{Y}_i^{B+1} \). Similarly, draw random currents \( \tilde{j}_i^{A} \) in body \( A \) and repeat the previous procedure, producing \( \tilde{O}_i^{A+1} \) and \( \tilde{Y}_i^{A+1} \). The iterations are stopped when both \( \epsilon_{i+1} \) and \( \epsilon_{i+1}^\dagger \) (for both \( A \) and \( B \) calculations) are smaller than a prescribed singular value tolerance.

Given the above matrices, a low-rank approximation of the SVD of \( Z_{AB} \) is obtained by expanding it onto the basis functions \( \tilde{O}^A \) and \( \tilde{O}^B \). From random matrix theory [174], given that \( \epsilon_{i+1} \) is small for a given number of successive iterations, this basis approximately spans the domain and range of the matrix. It follows from Maxwell’s equations that \( \tilde{Y}^A = \mathcal{G}^{AB} \tilde{\Omega}^B \) and hence,

\[
\tilde{O}^A \tilde{Z}^{AB} \tilde{O}^B \tilde{\Omega}^B \sqrt{\text{Im}[\chi^B]}^{-1} \tilde{\Omega}^B \approx \tilde{O}^A \sqrt{\text{Im}[\chi^A]} \tilde{Y}^A.
\]

Multiplication by the inverse of \( \tilde{O}^B \sqrt{\text{Im}[\chi^B]}^{-1} \tilde{\Omega}^B \) then produces a low-dimension \( k \times k \) matrix on the right hand side, amenable to standard SVD at minimal cost,

\[
\tilde{O}^A \tilde{Z}^{AB} \tilde{O}^B \approx \tilde{O}^A \sqrt{\text{Im}[\chi^A]} \tilde{Y}^A \left( \tilde{O}^B \sqrt{\text{Im}[\chi^B]}^{-1} \tilde{\Omega}^B \right)^{-1}.
\]

The singular value approximation of \( Z^{AB} \) is then derived from the small \( k \times k \) matrices \( U \), \( \Sigma \), and \( V \),

\[
Z^{AB} \approx \tilde{O}^A U \Sigma V^\dagger \tilde{O}^B \dagger
= \left( \tilde{O}^A \right) \Sigma \left( \tilde{O}^B \right)^\dagger
\]
Analysis of the convergence and performance properties of similar algorithms have been previously produced in [174].

B.2 Fast Gradient Adjoint

To exploit gradient-based optimization, knowledge of \( \partial \Phi / (\partial p_\alpha) \) is required for each optimization parameter \( p_\alpha \). Letting \( \partial_\alpha \) denote a partial derivative with respect to \( p_\alpha \), and retaining the notation of the main text and Appendix A,

\[
\partial_\alpha \Phi = \frac{\omega^4 \Theta}{2\pi c^4} \partial_\alpha \text{Tr} \left[ \text{Im} \left[ \chi^B \right] \mathcal{G}^{AB\dagger} \text{Im} \left[ \chi^A \right] \mathcal{G}^{AB} \right]. \tag{B.3}
\]

Using the symmetry of \( \chi^A \) and \( \chi^B \), along with the usual cyclic properties of the trace,

\[
\text{Tr} \left[ \text{Im} \left[ \chi^B \right] \partial_\alpha \mathcal{G}^{AB\dagger} \text{Im} \left[ \chi^A \right] \mathcal{G}^{AB} \right] = \\
\text{Tr} \left[ \mathcal{G}^{AB\dagger} \text{Im} \left[ \chi^A \right] \partial_\alpha \mathcal{G}^{AB*} \text{Im} \left[ \chi^B \right] \right] = \\
\text{Tr} \left[ \text{Im} \left[ \chi^B \right] \mathcal{G}^{AB\dagger} \text{Im} \left[ \chi^A \right] \partial_\alpha \mathcal{G}^{AB*} \right],
\]

one finds that the second and fourth terms in the partial derivative expansion are complex conjugates. (Here, \( \tau \) denotes transposition without complex conjugation.) Given that \( \mathcal{G} = \left[ (\nabla \times \nabla \times) - \omega^2 \epsilon \right]^{-1} \) is the inverse Maxwell operator,

\[
\partial_\alpha \mathcal{G}^{AB} = \omega^2 \mathcal{G}^A \partial_\alpha \chi \mathcal{G}^{B*},
\]
and so one finds:

\[
\partial_\alpha \Phi = \frac{\omega^4 \Theta}{2\pi c^4} \text{Tr} \left[ G^{AB\dagger} \text{Im} [\chi^A] G^{AB} \text{Im} [\partial_\alpha \chi^B] +
\right.
\]

\[
G^{AB} \text{Im} [\chi^B] G^{AB\dagger} \text{Im} [\partial_\alpha \chi^A] +
\]

\[
2\omega^2 \text{Re} \left[ G^B \text{Im} [\chi^B] G^{AB\dagger} \text{Im} [\chi^A] G^A \partial_\alpha \chi \right] \right].
\] (B.4)

The \(G^{AB}\) matrices involved in Eq. B.4 are low rank, and can computed alongside \(Z^{AB}\) at almost no extra cost. Specifically, if \(\sqrt{\text{Im} [\chi^A]}\) and \(\sqrt{\text{Im} [\chi^B]}\) are excluded from the error estimates \(\epsilon_{i+1}\) and \(\epsilon_{i+1}^{\dagger}\), one obtains an approximation for \(G^{BA}\) rather than \(Z^{AB}\). Both computations can be carried out at the same time, using the same current sources and field solutions. If these matrices have rank \(k\), the parameter independent matrices of Eq. B.4 are solved in order \(k\) computations. Hence, determination of the gradient is essentially no more costly than the determination of \(\Phi\).
Bibliography


132


147


149


153


157


