The QCQP Approach: A Novel Design Strategy for Optical Computing Devices

I. INTRODUCTION

Artificial intelligence has a size problem. Existing infrastructure is struggling to keep up with the immense computational and energy demands imposed by ChatGPT and other cutting-edge AI models. Optics and photonics has the potential to provide a solution.

For AI computations, current digital computers require the highest performance processors available (GPUs), which combined with other computational requirements often warrants the use of a dedicated, resource-hungry data center. Microsoft has even sought to restart the Three Mile Island nuclear plant to offset the huge demands its AI facilities impose on the electrical grid [1]. By comparison, an optical computer small enough to fit on a desk could implement machine learning models the size of ChatGPT-4 while consuming orders-of-magnitude less energy [2]. However, existing approaches to designing such electromagnetic wave-based optical computing systems are inadequate. Algorithmic design methods suffer from impractically high computational complexity, requiring compromises that lead to suboptimal designs. As a result, optical computing devices currently occupy prohibitively large device footprints. A state-of-the-art optical device with an input of about 1% the size of a typical image requires a footprint about 50 *meters* long [3], a scale that is clearly infeasible.

In this project, I worked as the sole researcher to implement and evaluate a novel photonic optimization paradigm for the computational design of highly compact optical devices. The approach recasts the typical photonic design problem—an optimization problem constrained by Maxwell's equations—as a quadratically-constrained quadratic program (QCQP), thereby addressing the computational inefficiency standing in the way of traditional gradient-based approaches to designing compact devices.

II. METHODOLOGY

I devised and studied a binary-choice optimization problem that seeks to maximize the electric field intensity of an electromagnetic wave measured at a single point $\mathbf{x}^*$ by organizing a grid of electric dipoles (Eq. (1)).

$$\begin{aligned} \underset{\boldsymbol{\varrho}}{\text{maximize}} \quad & f(\boldsymbol{\varrho}) = |\mathbf{E}_{\text{net}}(\mathbf{x}^*; \mathbf{p}(\boldsymbol{\varrho}))|^2 \\ \text{subject to} \quad & \varrho_i \in \{\varrho_a,\ \varrho_b\} \text{ for } i = 1, 2, \ldots, \\ & (\mathbb{I} - \alpha_0 \operatorname{diag}(\boldsymbol{\varrho})\mathbb{G})\mathbf{p} = \alpha_0 \operatorname{diag}(\boldsymbol{\varrho})\mathbf{E}_{\text{inc}} \end{aligned} \tag{1}$$

The first constraint requires each dipole to be in one of two states ϱ_i, either ϱ_a or ϱ_b. This is the "binary-choice" variable. The waves emitted by the dipoles in response to the driving electromagnetic wave, $\mathbf{E}_{\text{inc}}$, are permitted to interact with each other and with the other dipoles, governed by the second constraint. These interactions determine the polarization p_i of each dipole (which is calculated by inverting the second constraint and is thus a function of $\boldsymbol{\varrho}$). The coefficient α_0 is simply the linear polarizability of the dipoles (i.e., what material they are), and $\mathbb{G}$ is a matrix of the dipoles' Green's functions.

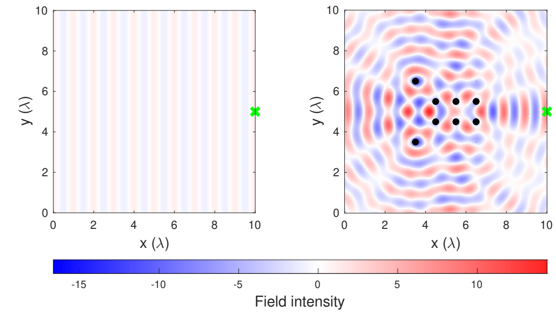

FIG. 1. Scattering system before (left) and after (right) optimization with the QCQP approach. On the left, the relatively weak driving field $\mathbf{E}_{\text{inc}}$ is parallel to the x-axis and has magnitude 1. After optimization, the dipoles (black dots) emit their own scattered waves, each driven by $\mathbf{E}_{\text{inc}}$ and the other dipoles' scattered waves. The optimized grid is laid out to create constructive interference in $\mathbf{E}_{\text{net}}$ focused at the target (green "X").

Fig. 1 shows the system before and after solving Eq. (1) using the QCQP approach described in Section II C. Field intensity at the green "X" is maximized by adding dipoles (black circles). The optimized "degree of freedom" is whether each dipole is present ($\varrho_i = 1$) or absent ($\varrho_i = 0$). This problem has a similar degree of difficulty (nonconvexity) as many others in wave optics, so it can serve as a testbed for techniques to efficiently design true optical computing devices.

A. Traditional Gradient-Based Optimization

To start, I constructed a MATLAB simulation that uses the dyadic Green's function in 2D to calculate the electric field, $\mathbf{E}_{\text{net}}$, generated by the dipoles in response to a driving electromagnetic wave, $\mathbf{E}_{\text{inc}}$. Then using linear algebra techniques and the $\mathbb{CR}$-calculus [4], I derived the adjoint equations and accompanying $\mathbf{p}_{\text{adj}}$ to formulate a computationally efficient optimization approach requiring only one additional simulation to calculate the gradient (Eq. (2)) rather than the thousands of additional simulations required for numerical approaches. To solve the optimization problem, I used a standard gradient descent routine, MATLAB's `fmincon` function with the interior-point algorithm.

$$\frac{\partial f}{\partial \varrho_i} = 2\mathrm{Re}\left[\alpha_0 \left(\mathbf{E}_{\text{inc}} + \mathbb{G}\mathbf{p}\right)_i p_{\text{adj},\, i}\right] \tag{2}$$

As this technique, the adjoint state method, optimizes by gradient descent, it is often caught in low-quality *local* optima. Yet it is currently the dominant optimization paradigm in photonic design, so its performance provides an important point of comparison for the novel QCQP approach.

B. SDP Optimization for Fundamental Bounds

Using linear algebra techniques, I rederived the optimization problem of Eq. (1) as a QCQP, which is a well-studied class of computational problems that have not before been applied to photonic design [5].

$$\begin{aligned} \underset{P}{\text{maximize}} \quad & \mathrm{Tr}\left(\mathbb{A}_0 P\right) \\ \text{subject to} \quad & \mathrm{Tr}\left(\mathbb{B}_i P\right) = 0 \quad \text{for } i = 1, 2, \ldots, \\ & P \geq 0, \\ & \mathrm{rank}\ P = 1 \end{aligned} \tag{3}$$

The QCQP approach of Eq. (3) "lifts" the problem into a higher dimensional space, formulating it in terms of a matrix variable $P = \mathbf{p}\mathbf{p}^\dagger$, where $\dagger$ denotes the conjugate transpose operation. The operators $\mathbb{A}_0$ and $\mathbb{B}_i$ encode the objective function and constraints, respectively, from Eq. (1). This reformulation concentrates the nonconvexity (that is, the difficulty) of the problem into a single optimization constraint: that the matrix, P, has rank one. In a strategy called "semidefinite relaxation," I dropped this nonconvex constraint while maintaining all others (Eq. (4)), and then computationally solved the easier convex problem, called a semidefinite program (SDP), using the `cvx` package in MATLAB. The semidefinite relaxation is closely related to the original problem and provides a fundamental limit on its optimal value, but achieving this optimal value in a real device is often a physical impossibility.

$$\begin{aligned} \underset{P}{\text{maximize}} \quad & \mathrm{Tr}\left(\mathbb{A}_0 P\right) \\ \text{subject to} \quad & \mathrm{Tr}\left(\mathbb{B}_i P\right) = 0 \quad \text{for } i = 1, 2, \ldots, \\ & P \geq 0 \end{aligned} \tag{4}$$

C. Physically Feasible QCQP Optimization

Since the nonconvex constraint is ultimately important to the physical reality of the optimization problem, it must be accounted for when looking for a design that can be built in the physical world. To solve the more difficult (nonconvex) problem of Eq. (3), I implemented the "majorization-minimization" method [6] that iteratively solves a closely related SDP with the same constraints as the SDP in Eq. (4), but with the objective function slightly altered by inclusion of a convex rank penalty. At each step, the rank penalty is increased, so the solution to this rank-penalized SDP approaches a solution to the original, physically realistic QCQP including the nonconvex rank-one matrix constraint (Eq. (3)).

III. RESULTS

Recasting this computationally difficult photonic design problem as a quadratically-constrained quadratic program appears to lead to improved optimization performance when compared to the adjoint state method, the dominant paradigm in photonics.

Numerical experimentation shows that the QCQP approach reliably finds higher quality optima than the adjoint state method does, even when gradient descent is run 100 or more times from different random starting points. Fig. 2 compares the optimization performance of the adjoint state, SDP, and QCQP approaches in 100 trials on the same problem, each seeded with a random starting point. As this is a maximization problem, higher objective values correspond to better performance.

The "optimization landscape" for this wave-based problem is highly oscillatory—evident in the oscillations of the red line in Fig. 2—so the traditional adjoint state approach is unreliable. That is, the optimization performance varies widely between runs when seeded with a random starting point. Ideally in optimization, one can be confident that at least a near-global optimum has been reached whenever the routine terminates, but the adjoint state approach does not offer such assurances here. However, it is clear that the convex SDP (green line) provides a fundamental limit on performance as it is never exceeded, and that the nonconvex QCQP result is generally superior to that of the adjoint state method. The SDP and QCQP lines are horizontal since the performance of both techniques is independent of the starting point, unlike in the traditional adjoint state approach.

As problem complexity increases, the QCQP approach continues to outperform the adjoint state method. Fig. 3 shows the optimization objective value as a function of the number of dipoles in the grid being optimized (a proxy for problem difficulty), and the error bars on the red line give a two standard deviation confidence interval from 100 randomly seeded adjoint state optimizations for the same problem. As in Fig. 2, higher values indicate superior performance. In Fig. 3, the lines all meet at 1 dipole, which is to be expected because there is little to optimize in such a problem—each method natu-

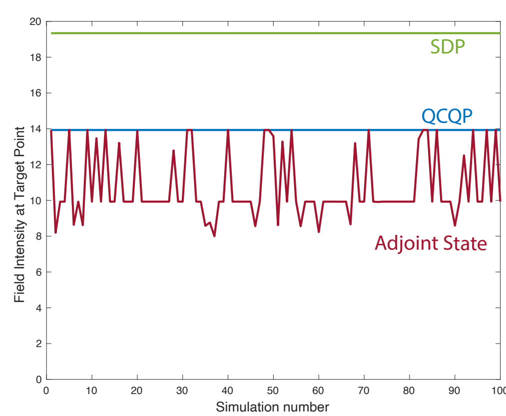

FIG. 2. The SDP of Eq. (4) provides a fundamental bound on the optimal objective value (y-axis). The QCQP (Eq. (3)) reliably converges to a high-quality and physically feasible optimum. The adjoint state method (Eq. (1)) is unreliable, oscillating between high- and low-quality optima when seeded with random starting points. Higher is better.

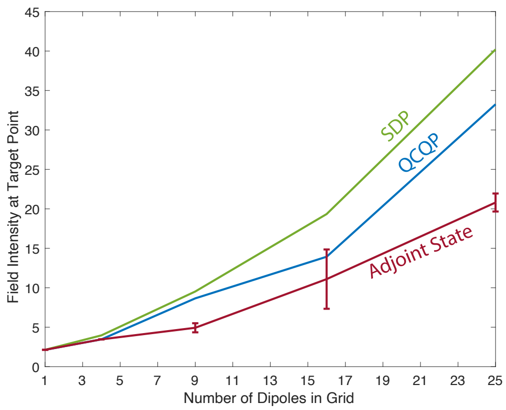

FIG. 3. As the problem increases in difficulty (i.e., more dipoles to position in the grid), the adjoint state (Eq. (1)) optima are, on average, worse than those of the QCQP (Eq. (3)) and SDP (Eq. (4)), and the error bars give a two standard deviation confidence interval from 100 randomly seeded optimizations on the same problem. Higher is better. The SDP and QCQP lines diverge more rapidly from the adjoint state line as the problem becomes more complex, reflecting the adjoint state method's inability to cope with the most difficult problems.

rally converges to an optimal solution that includes the standalone dipole. The QCQP approach (blue line) outperforms the adjoint state method (red line) more dramatically in more difficult problems, which indicates its promise as a paradigm for designing highly sophisticated wave scattering devices like optical computers. Indeed, it is the poor performance of gradient descent in these highly nonconvex problems that has limited optimal design of large-scale optical computing devices.

These results suggest that recasting a difficult photonic design problem as a QCQP improves optimization performance and could enable the computational design of compact yet fully sophisticated optical computers that meet AI's exploding computational demands.

IV. CONCLUSION AND FUTURE DIRECTIONS

The great advantage of optical computing devices is that they leverage the properties of wave scattering to perform matrix-vector products entirely in the physical domain, making them highly energy- and time-efficient. But this wave behavior makes typical methods of computational design—those based on gradient descent—poorly suited because the optimization landscape becomes highly oscillatory as the design complexity increases (that is to say, there are lots of "peaks and valleys" that will cause the gradient descent optimization routine to terminate at a point that is not globally optimal; see Fig. 2). Yet human-intuition based designs are prohibitively large, and more reliable gradient-based alternatives to gradient descent are too computationally intensive for designing the most complex optical devices. The QCQP approach that I use here, on the other hand, "smooths out" the optimization landscape by lifting the problem to a higher dimensional space, resulting in more reliable convergence and higher-quality designs.

These findings contribute to a larger vision of a novel design methodology that computationally designs reasonably sized optical computing devices by (1) decomposing the device into simple, uncoupled components and then (2) designing each component computationally. Although the QCQP approach is currently incapable of designing the most complicated devices in a single shot, it has strong potential as a method of designing such a series of separate, uncoupled sub-devices. This technique could efficiently produce highly compact, yet fully sophisticated, optical computing devices with substantial advantages over digital devices.

Future inquiries will be necessary to realize the ultimate project aim of creating a fully-optical AI device. My next step will be to equip the QCQP approach with an optimization objective that designs a bona fide optical computing device, as well as a more physically realistic representation of electromagnetic scatterers (including the magnetic dipole moment and precise measures of material energy dissipation). Then to design more complex devices, I will devise an algorithmic strategy for breaking down a large device into simple, uncoupled sub-devices, each of which can then be designed independently with the QCQP approach and assembled into a compact yet highly sophisticated optical computer, as noted above. With the computational design methodology complete, I will collaborate with an experimental research group to manufacture these computationally designed sub-devices and assemble them into an actual compact optical computer. After that, we will be ready to integrate optical devices into the existing AI ecosystem as an environmentally friendly and less resource-intensive alternative to large data centers.

[1] M. Greshko, Nuclear power for ai: what it will take to reopen three mile island safely, Nature **634**, 272 (2024).

[2] E. Cartlidge, Scattered light holds promise for efficient ai (2024).

[3] H. Larocque and D. Englund, Universal linear optics by programmable multimode interference, Opt. Express **29**, 38257 (2021).

[4] K. Kreutz-Delgado, The complex gradient operator and the cr-calculus (2009), arXiv:0906.4835 [math.OC].

[5] S. Gertler, Z. Kuang, C. Christie, H. Li, and O. D. Miller, Many photonic design problems are sparse qcqps, Science Advances **11**, eadl3237 (2025), https://www.science.org/doi/pdf/10.1126/sciadv.adl3237.

[6] T. Liu, B. Sun, and D. H. K. Tsang, Rank-one solutions for sdp relaxation of qcqps in power systems, IEEE Transactions on Smart Grid **10**, 5 (2019).