Journal of the Optical Society of America B
OPTICAL PHYSICS

Understanding search behavior via search landscape analysis in design optimization of optical structures

SACHA VERWEIJ[1,2,4] AND SHANHUI FAN[3,5]

[1]Department of Applied Physics, Stanford University, Stanford, California 94305, USA
[2]Institute for Computational & Mathematical Engineering, Stanford University, Stanford, California 94305, USA
[3]Ginzton Laboratory, Stanford University, Stanford, California 94305, USA
[4]e-mail: sacha@stanford.edu
[5]e-mail: shanhui@stanford.edu

Received 2 May 2016; revised 16 August 2016; accepted 22 August 2016; posted 22 August 2016 (Doc. ID 264384); published 10 November 2016

Search algorithms play a crucial role in systematic design optimization of optical structures. We demonstrate how *search landscape analysis* sheds light on the behavior of search algorithms in this context and thereby leads to valuable insight. Specifically, we present a case study in which search landscape analysis demystifies the surprisingly good performance of a simple stochastic local search algorithm—restarted iterative best improvement—on a challenging design optimization problem—combinatorial design optimization of a multi-spatial-mode photonic crystal waveguide bend that preserves modal content. Moreover, the search landscape analysis leads to insight potentially valuable in design optimization of a broad class of optical structures.

OCIS codes: (220.0220) Optical design and fabrication; (350.4600) Optical engineering; (230.0230) Optical devices; (000.5490) Probability theory, stochastic processes, and statistics; (350.4238) Nanophotonics and photonic crystals.

http://dx.doi.org/10.1364/JOSAB.33.002457

1. INTRODUCTION

Systematic design optimization of optical structures is opening new and advancing existing frontiers in the control of light, from ultracompact wavelength- and mode-division multiplexing [1–13] to performance enhancement of photo- and thermophoto-voltaics [14–18]. To exploit systematic design optimization's full potential in optics, search algorithms capable of efficiently sifting large design spaces are critical [1–37].

The degree to which a search algorithm exploits problem structure determines its efficacy. Thus, in the methodical analysis that is often key to understanding, and thereby ultimately improving, search behavior, analysis of problem structure—or *search landscape analysis*—is a crucial component. Despite this, search landscape analysis is largely absent from the otherwise rapidly growing literature on search algorithms in design optimization of optical structures.

In this paper, we illustrate how search landscape analysis can provide insight into search behavior in design optimization of optical structures. To this end, we present a case study in which search landscape analysis sheds light on the surprisingly good performance of a simple search algorithm and thereby yields insight potentially valuable in design optimization of a broad class of optical structures.

Particularly, Section 2 introduces the model design optimization problem about which the case study revolves, and Section 3 defines the associated search landscape. Section 4 describes the search algorithms under study, and Section 5 characterizes these algorithms' performance on our model design optimization problem. Section 6 details the search landscape analyses, which enable understanding of the algorithms' observed performance, culminating in a successful performance retrodiction and determination of the key search landscape features impacting performance. Section 7 presents the physical intuition that sprang from the analyses detailed in Section 6. In closing, Section 8 addresses possible generalizations and extensions of the results in the preceding sections.

2. MODEL DESIGN OPTIMIZATION PROBLEM

Our case study revolves around design optimization of a multi-spatial-mode photonic crystal waveguide bend that preserves modal content; such multi-spatial-mode photonic crystal devices, for example those designed in [10–12], hold promise for mode-division-multiplexing in future photonic integrated circuits. Particularly, we ultimately seek a device like that schematized in Fig. 1(b):

A square lattice of silicon cylinders in air, which supports a broad transverse-magnetic (TM) mode bandgap, hosts the

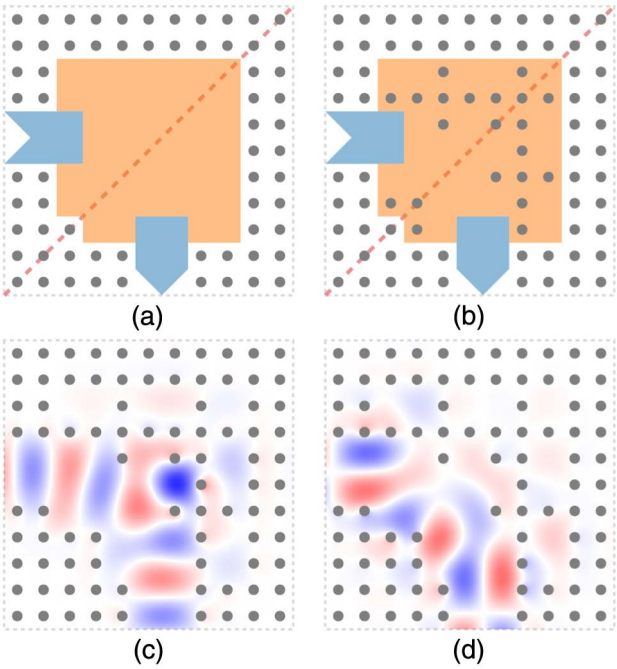

Fig. 1. (a) Design prototype for a multi-spatial-mode photonic crystal waveguide bend that preserves modal content. A square lattice of silicon cylinders (gray circles, $r/a = 0.2$, $n = 3.4$) in air (white backdrop), which supports a transverse-magnetic (TM) mode bandgap over frequencies $0.29 \cdot 2\pi c/a$ to $0.42 \cdot 2\pi c/a$, hosts the device. Cylinder-deletion defects in this photonic crystal constitute the device itself. Highlighted in blue, two adjacent horizontal lines of cylinder deletions form a waveguide incident from the left. Similarly highlighted in blue, two adjacent vertical lines of cylinder deletions form a waveguide exiting at the bottom. Each waveguide supports two TM spatial modes over frequencies $0.36 \cdot 2\pi c/a$ to $0.42 \cdot 2\pi c/a$. Highlighted in orange, a mirror-symmetric, multi-deletion junction terminates these waveguides at their intersection. Together, the waveguides and junction compose a symmetric multi-spatial-mode bend. The dashed red line indicates this prototype's diagonal mirror symmetry. We generate a candidate structure from this prototype, such as that shown in (b), by placing deletion cells at some of the structure's junction lattice sites consistent with the prototype's mirror symmetry. (c) and (d) respectively show the out-of-plane electric field patterns resulting from excitation, at frequency $0.39 \cdot 2\pi c/a$, of the first and second ingoing TM spatial modes on the horizontal waveguide of this particular structure.

device. Cylinder-deletion defects in this photonic crystal constitute the device itself. Two adjacent, horizontal lines of cylinder deletions form a waveguide incident from the left. Similarly, two adjacent, vertical lines of cylinder deletions form a waveguide exiting at the bottom. Each waveguide supports two TM spatial modes over the upper half of the crystal's bandgap. A mirror-symmetric, multi-defect junction terminates these waveguides at their intersection. Together, the waveguides and junction compose a symmetric multi-spatial-mode bend. Figures 1(c) and 1(d) respectively show the out-of-plane electric field patterns resulting from excitation of the first and second ingoing TM spatial modes on the horizontal waveguide; insofar as one can discern visually, these modes respectively strictly transmit to the first and second outgoing TM spatial modes on the vertical waveguide, exemplifying near-perfect modal-content-preserving transmission. Figure 1's caption provides additional detail, and Fig. 18 in Appendix A depicts the behavior of additional high-performing structures.

To design such a device, we begin from the prototype schematized in Fig. 1(a); this prototype is identical to the device described above but for a junction solely containing defective cells. We generate a *candidate structure* from this prototype by placing nondefective cells at some set of the structure's *independent lattice sites*—the twenty-five lattice sites in the junction's upper-left half, diagonal inclusive; enforcement of the prototype's diagonal mirror symmetry then dictates placement of nondefective cells in the junction's lower-right half. The desired device described above and shown in Fig. 1(b) exemplifies a candidate structure. We can generate 2^{25} distinct candidate structures via this procedure; the set of all such candidate structures constitutes our *search space* $\mathcal{S}$.

This discrete, *combinatorial* search space contrasts with continuous search spaces admitting, for example, continuous geometric or material-parametric deformations between candidate structures. Restriction to such a discrete search space is not artificial: Our interest in this problem derives from design for scalable, low-cost manufacturability, which strongly constrains admissible design spaces [38–43]. For example, both traditional semiconductor and dedicated silicon photonics foundries typically only allow designs consisting of select geometric elements with constraints on scale, relative position, and refractive index. In this context, combinatorial design spaces are practically relevant. This restriction enriches the problem but also makes it more challenging, necessitating either (1) strict combinatorial optimization techniques or (2) techniques involving continuous relaxation of the discrete problem and, for example, thresholding and/or penalty mechanisms to recover a solution to the discrete problem from a solution of the continuous relaxation. Here we consider the former class, though techniques in the latter class are also viable and often perform well, for example as in [1–9,19–25]. The insights developed in this paper inform both approaches.

Next we define our *design objective*, which is to find a candidate structure that exhibits satisfactorily high modal-content-preserving transmission or, equivalently, satisfactorily low reflection and modal crosstalk in transmission. We capture our design objective quantitatively in an *objective function* $f : \mathcal{S} \to \mathcal{R}$. Though many possible objective functions reasonably correspond to our design objective, we focus on minimization of the ℓ^2-type objective function

$$f(s) = 2^{-1/2} \sqrt{\sum_{i,j} R^2_{i\to j}(s) + \sum_{i,j\neq i} T^2_{i\to j}(s)}, \tag{1}$$

where, under monochromatic excitation of ingoing TM spatial mode i on one port of candidate structure $s \in \mathcal{S}$, $R_{i\to j}(s)$ denotes the excitation-normalized power reflected into outgoing TM spatial mode j on the same port, and $T_{i\to j}(s)$ that transmitted into outgoing TM spatial mode j on the other port. Accordingly, the first summation in this objective function captures reflection and the second modal crosstalk in transmission. Where i and j in the summations respectively run over all

ingoing and outgoing spatial modes, the factor of $2^{-1/2}$ normalizes the objective function to the interval [0, 1].

Finally, we form our *design optimization problem* from these elements: In principle, we seek a candidate structure s^* in search space $\mathcal{S}$ that minimizes objective function f, or $s^* = \arg\min_{s\in\mathcal{S}} f(s)$. In practice, however, we stop the search once we find an acceptable candidate structure s_a in search space $\mathcal{S}$ with objective (function) value below some acceptability threshold f_a, or $s_a \in \mathcal{S} : f(s_a) \leq f_a$.

We evaluate our objective function numerically via the Dirichlet-to-Neumann map method and extensions detailed in [12,13,44]. These techniques' efficiency enables single-frequency characterization of our complete search space (approximately 34 million candidate structures) within a few hundred core-hours on Xeon E5-2680 processors; this data set is available on request. Of course, obviating such exhaustive search is the objective of this and other search algorithms research. Complete characterization of our search space, however, enables us to analyze search behavior, our problem's structure, and their relationship in a statistically rigorous manner.

3. SEARCH LANDSCAPE

With our design optimization problem in place, introducing our search landscape and algorithms requires one addition—a sense of locality for our search space $\mathcal{S}$:

We measure the distance between candidate structures s and s' by the number of independent lattice sites at which they differ—a form of *Hamming distance* $\mathbf{H}(s, s')$. To illustrate, the candidate structures shown in Figs. 1(a) and 1(b) differ at twelve lattice sites in the upper-left half of the junction, diagonal inclusive, and so by this metric the (Hamming) distance between them is twelve.

With distance defined, we introduce a *neighborhood relation* $\mathcal{N} \subseteq \mathcal{S} \times \mathcal{S}$ that, intuitively, indicates which candidate structures are *neighbors* or *adjacent*: We say that s and s' are neighbors if and only if they differ at precisely one independent lattice site or, equivalently, their Hamming distance $\mathbf{H}(s, s')$ is one.

Our search space $\mathcal{S}$, neighborhood relation $\mathcal{N}$, and objective function f together constitute our *search landscape*. Figure 2 illustrates this concept via a toy search landscape: Each node represents a candidate structure. Lines connect neighboring candidate structures. Node position along the vertical axis reflects the candidate structure's objective value. Together the search space (nodes), neighborhood relation (lines), and objective function (vertical positions) form a sort of landscape that we can imagine traversing.

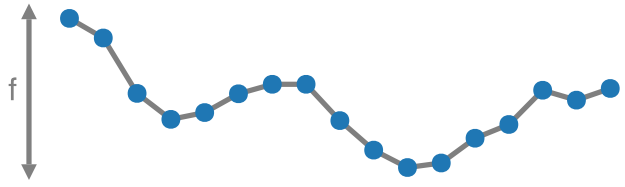

Fig. 2. A toy search landscape: Nodes (blue) represent our search space's candidate structures. Lines (gray) connect neighboring candidate structures, illustrating our neighborhood relation. Node position along the vertical axis reflects the candidate structure's objective value. Together the search space (nodes), neighborhood relation (lines), and objective function (vertical positions) form a search landscape.

4. SEARCH ALGORITHMS

We can view an algorithm for solving our design optimization problem as a strategy for traversing our search landscape, specifically in search of an acceptable candidate structure. The degree to which that strategy exploits search landscape structure determines the algorithm's efficacy. We describe three such strategies below, beginning with the least sophisticated.

Uninformed random picking (URP) is appropriately named: In each search step, we draw a candidate structure uniformly at random from our search space and determine its objective value. If the candidate structure's objective value is acceptable, we terminate the search. Otherwise, we continue with the next search step.

Evidently, uniformed random picking exploits no search landscape structure. Though as a result this algorithm performs comparatively poorly in search landscapes with exploitable structure, it performs as well as possible in search landscapes lacking exploitable structure. Hence uninformed random picking makes a logical performance baseline where we expect a dearth of exploitable structure.

Iterative best improvement (II) is slightly more clever: In each search step, we determine the objective values of all neighbors of an initially given *current* candidate structure. If any neighbor's objective value improves upon the current candidate structure's objective value, we declare the *best improving* neighbor the new current candidate and continue with the next search step. Otherwise, the current candidate structure—having no improving neighbors—is a *local minimum*. As iterative best improvement has no means to *escape* this local minimum, we then terminate the search. Additionally, if at any point this procedure uncovers a candidate structure with acceptable objective value, we also terminate the search. Figure 3(a) illustrates a single step in a run of iterative best improvement, in which we evaluate the initial current candidate structure's neighbors and move to the best improving neighbor. Figure 3(b) illustrates a sequence of such steps and the search's ultimate arrival—and subsequent termination—at a local minimum. Iterative best improvement is an instance of *local search*.

Intuitively, iterative best improvement exploits any search landscape local *basin* or *valley* structure to efficiently descend to a local minimum. (When we describe a method as performing "efficient descent" and thereby implicitly compare that method to others, we confine that comparison to other methods operating in discrete search spaces such as ours. We are not comparing the efficiency of such methods' descents to methods that operate in continuous search spaces.) In search landscapes possessing many local minima of varying quality—which is often the case in practice—descent to *a* local minimum typically fails to produce a satisfactory result. Going beyond *a* local minimum requires an *escape mechanism*.

Restarted iterative best improvement (RII) augments iterative best improvement with a rudimentary escape mechanism: In each search *super*step, we draw a candidate structure uniformly at random from our search space and, from that candidate

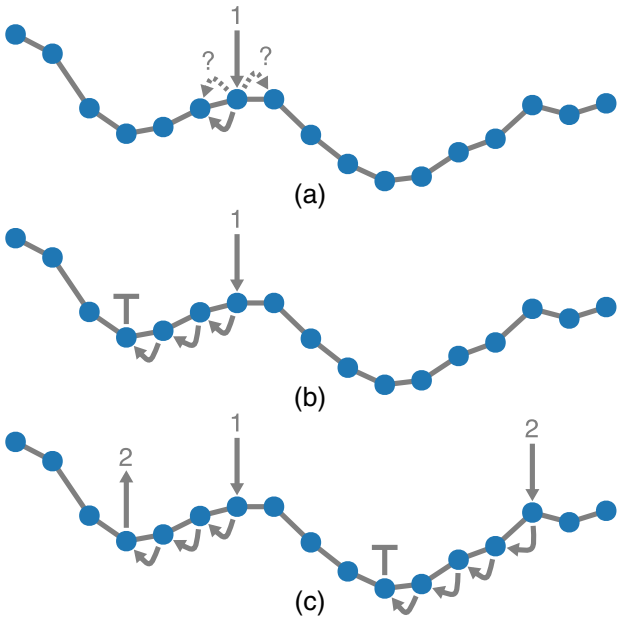

Fig. 3. (a) The first step in a run of iterative best improvement: The descending vertical arrow marks the (initial) current candidate structure. The curved dashed arrows represent evaluation of the (initial) current candidate structure's neighbors. The solid curved arrow expresses selection of the best-improving neighbor as the new current candidate structure, concluding the first step of iterative best improvement. (b) Iterative best improvement's following steps (solid curved arrows) and ultimate termination at a local minimum (tee or down tack). (c) Two successive supersteps in a corresponding run of restarted iterative best improvement.

structure, perform iterative best improvement. When iterative best improvement terminates at a local minimum, we *restart* the search by continuing with the next search superstep. If at any point this procedure uncovers a candidate structure with acceptable objective value, we terminate the search. Figure 3(c) illustrates two successive search supersteps in a run of restarted iterative best improvement. Restarted iterative best improvement is an instance of *repeated local search* where iterative best improvement is the *subsidiary local search procedure*.

Intuitively, restarted iterative best improvement exploits any search landscape local basin structure to perform an efficient sampling of the search landscape's local minima. Between repetitions of iterative best improvement, all information gained about our search landscape is lost; preferably we would retain and exploit what we learn between these repetitions, but more sophisticated algorithms that effect this are beyond the scope of this paper.

5. ALGORITHM PERFORMANCE

Both restarted iterative best improvement and uninformed random picking are *stochastic algorithms*. The behavior of an individual run of a stochastic algorithm is not particularly meaningful; rather, only the statistical behavior of a representative *ensemble* of runs necessarily has meaning. Figure 4, which compares restarted iterative best improvement's performance against uninformed random picking as baseline on the model

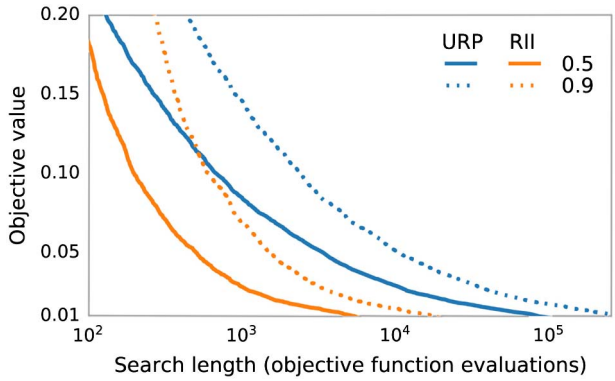

Fig. 4. Performance comparison of restarted iterative best improvement in orange and uninformed random picking in blue. Solid and dashed lines respectively reflect typical (0.5-quantile or median incumbent objective value) and near-worst-case (0.9-quantile incumbent objective value) performance. The objective function axis (vertical) terminates at the objective function's discretization-error noise floor ($f \approx 0.01$) and reflects operation at frequency $0.39 \cdot 2\pi c/a$.

design optimization problem outlined in Section 2, conveys such statistical behavior:

For each of 2000 independent runs of restarted iterative best improvement and uninformed random picking, we recorded the *incumbent objective value* at each search step, which is the objective value of the best candidate structure the search had found by that point. For each algorithm and search step, this procedure produced a sample of 2000 corresponding incumbent objective values, from which we calculated 0.5- and 0.9-quantile incumbent objective values. Figure 4's solid and dotted orange lines respectively show the calculated 0.5- and 0.9-quantile incumbent objective values at each search step for restarted iterative best improvement. The solid and dotted blue lines show the same for uninformed random picking.

Intuitively, our objective function captures the proportion of power lost to reflection and modal crosstalk in transmission. The solid and dotted lines thus respectively show the typical- and near-worst-case evolution of the best-encountered candidate structure's power loss as the search proceeds.

Figure 4 shows that, in the typical case, restarted iterative best improvement finds excellent structures in better than an order of magnitude less time than uninformed random picking. Moreover, even *near-worst-case* runs of restarted iterative best improvement find excellent structures in almost an order of magnitude less time than *typical* runs of uninformed random picking.

This result is remarkable: An algorithm can only outperform uninformed random picking to the degree that the search landscape possesses exploitable structure. Hence the above finding implies that our search landscape possesses considerable exploitable structure. Yet the conventional wisdom—that our search space consists of highly sensitive interferometric devices with substantial differences, so that even neighboring structures' behavior is often grossly unrelated—would suggest that our search landscape possesses little, if any, exploitable structure. This observation therefore motivates detailed search landscape analysis as described in the next section.

6. SEARCH LANDSCAPE ANALYSIS

As restarted iterative best improvement distills to a strategy for sampling a search landscape's local minima, two aspects of our search landscape's structure are paramount to understanding the above result: whether our search landscape's structure enables efficient sampling of *individual* local minima and whether our search landscape's structure facilitates efficient sampling of its *good* local minima, with *good* meaning low-objective-value. We additionally investigate the frequently discussed local minima density.

A. Efficient Sampling of Individual Local Minima

Consider a given local minimum. Associated with that local minimum is an *attraction basin*—the set of all candidate structures from which our local search procedure ultimately finds that local minimum. Figure 5 illustrates our toy search landscape's attraction basins. The attraction basin's structure determines whether descents to that local minimum proceed efficiently. The characteristic structure of *all* attraction basins in our search landscape determines how efficiently descents proceed *generally*. We investigate two aspects of characteristic basin structure: whether the basins' internal structure leads to *direct* or instead *tortuous* descents and how the basins' shapes impact typical descent length.

1. Descents Are Direct

Figure 6 contrasts direct descents with tortuous descents: In each pane a circle, representing the boundary of an attraction basin, encloses a star, representing that attraction basin's local minimum, and a square, representing a starting candidate structure for local descent. The solid line indicates the trajectory of local descent from starting candidate structure to local minimum, while the dashed line indicates the direct path between those candidate structures. Each line segment represents one search step. In Fig. 6(a), the descent trajectory closely approximates the direct path, requiring a near-minimal number of search steps—this descent is *direct*. In contrast, the descent trajectory in Fig. 6(b) departs wildly from the direct path and requires many more search steps—this descent is *tortuous*. Direct descents are efficient; tortuous descents are *in*efficient. An attraction basin's internal structure determines whether typical descents are direct or tortuous.

An attraction basin's *watershed paths*, or steepest-descent paths, reflect typical local descent trajectories in that basin; iterative best improvement walks precisely along watershed paths. The *watershed distance* between two candidate structures is the cumulative Hamming distance along a watershed path between them. If the watershed distance from a candidate structure to its *corresponding* local minimum—the local minimum to which attraction basin the candidate structure belongs—is comparable to their Hamming distance, the descent between them is relatively direct. If, to the contrary, the former significantly exceeds the latter, it is tortuous.

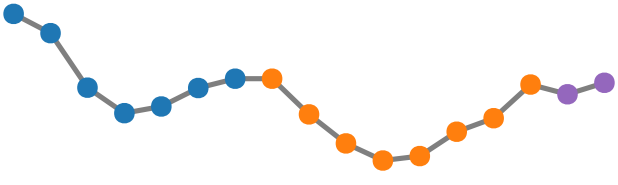

Fig. 5. For each of our toy search landscape's three local minima, color grouping indicates the associated attraction basin.

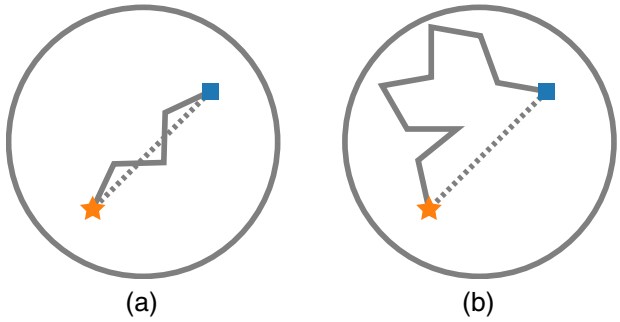

Fig. 6. (a) and (b) contrast direct descents with tortuous descents: In each pane a circle, representing the boundary of an attraction basin, encloses a star, representing that attraction basin's local minimum, and a square, representing a starting candidate structure for local descent. The solid line indicates the trajectory of local descent from starting candidate structure to local minimum, while the dashed line indicates the direct path between those candidate structures. Each line segment represents one search step. In (a), the descent trajectory closely approximates the direct path, requiring a near-minimal number of search steps—this descent is *direct*. In contrast, the descent trajectory in (b) departs wildly from the direct path and requires many more search steps—this descent is *tortuous*.

To illustrate, suppose that the descent trajectories (solid lines) in Figs. 6(a) and 6(b) reflect watershed paths, and each line segment has unity Hamming distance. Then the starting candidate structure in Fig. 6(a) would have watershed distance four and Hamming distance three to its corresponding local minimum, indicating a relatively direct descent; the equivalents for Fig. 6(b) would be nine and three, indicating a tortuous descent.

Table 1 shows that watershed and Hamming distances are usually comparable in our search landscape. We interpret the data as follows: For restarted iterative best improvement, 94.1% of descents are perfectly direct ($d_{ws} = d_H$), 5.3% of descents involve a only a single indirect step and subsequent correction ($d_{ws} = d_H + 2$), and a vanishing fraction (0.6%) of descents involve two or more indirect steps and corrections ($d_{ws} \geq d_H + 4$). Moreover, most descents involving one or more indirect steps are relatively long—when indirect steps occur, they typically occur in small number relative to the descent length. Furthermore, even for the longest descents—taking up to sixteen of twenty-five possible steps in the space—about 80% are direct.

In sum, the characteristic internal structure of our search landscape's attraction basins is strikingly conducive to direct descents.

2. Basins Are Extended

Figure 7 illustrates the impact of basin shape on typical descent length: In each pane an ellipse, representing the boundary of an attraction basin, encloses a star, representing that attraction basin's local minimum. The solid lines show typical descent

Table 1. Comparison of Watershed to Hamming Distances in Our Search Landscape[a]

d_H	$d_{ws} = d_H$		$d_{ws} = d_H + 2$		$d_{ws} \geq d_H + 4$		+
0	0.4%	(100%)	0%	(0%)	0%	(0%)	0.4%
1	6.2%	(100%)	0%	(0%)	0%	(0%)	6.2%
2	15.5%	(98.9%)	0.2%	(1.1%)	<.1%	(<.1%)	15.7%
3	21.0%	(96.9%)	0.6%	(2.9%)	<.1%	(0.2%)	21.6%
4	19.8%	(94.4%)	1.1%	(5.1%)	0.1%	(0.5%)	21.0%
5	14.5%	(91.6%)	1.2%	(7.5%)	0.1%	(0.9%)	15.8%
6	8.8%	(88.8%)	1.0%	(10.0%)	0.1%	(1.3%)	9.9%
7	4.6%	(86.1%)	0.6%	(12.2%)	<.1%	(1.7%)	5.3%
8	2.1%	(83.4%)	0.4%	(14.3%)	<.1%	(2.3%)	2.5%
9	0.8%	(81.0%)	0.2%	(16.2%)	<.1%	(2.7%)	1.0%
10	0.3%	(78.9%)	<.1%	(18.1%)	<.1%	(3.0%)	0.4%
11	<.1%	(77.6%)	<.1%	(19.1%)	<.1%	(3.4%)	0.1%
12	<.1%	(76.8%)	<.1%	(19.5%)	<.1%	(3.7%)	<.1%
13	<.1%	(75.7%)	<.1%	(20.7%)	<.1%	(3.6%)	<.1%
14	<.1%	(80.5%)	<.1%	(17.9%)	<.1%	(1.6%)	<.1%
15	<.1%	(83.7%)	<.1%	(16.3%)	0%	(0%)	<.1%
16	<.1%	(89.5%)	<.1%	(10.5%)	0%	(0%)	<.1%
+	**94.1%**		**5.3%**		**0.6%**		**100%**

[a]Each entry provides the proportion of our search landscape's candidate structures having Hamming distance d_H and watershed distance d_{ws} to their corresponding local minima. For a given entry, the corresponding row of the leftmost column indicates d_H, and the corresponding column of the topmost row indicates d_{ws} expressed relative to d_H. Each *parenthesized* entry provides the same proportion, but instead normalized by the number of candidate structures with Hamming distance d_H to corresponding local minima rather than the total number of candidate structures; the rightmost column provides the proportion of candidate structures with Hamming distance d_H. Finally, the bottommost row provides the proportion of candidate structures with watershed distance d_{ws}, irrespective of d_H.

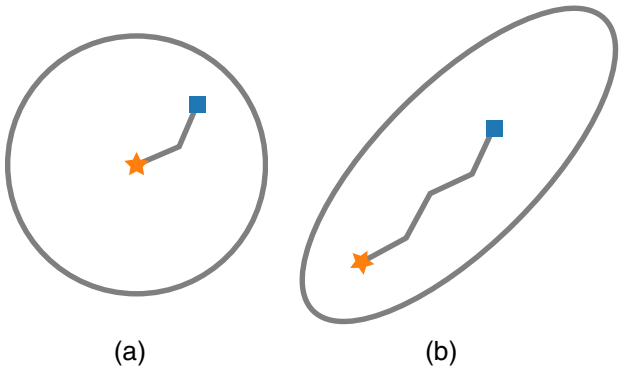

Fig. 7. Illustration of the impact of basin shape on typical descent length: In each pane an ellipse, representing the boundary of an attraction basin, encloses a star, representing that attraction basin's local minimum. The solid lines show typical descent trajectories in each basin, assuming descents are comparably direct between the two basins. Though these basins contain the same number of candidate structures (area), they have different *shapes*: The basin in (a), being circular and with its local minimum centrally located, is *compact*—its candidate structures (points in the basin area) are close-packed around its local minimum. In contrast, the basin in (b), being highly eccentric and with its local minimum located by a focus, is *extended*—its candidate structures extend away from its local minimum.

trajectories in each basin, assuming descents are comparably direct between the two basins. Though these basins contain the same number of candidate structures (area), they have different *shapes*: The basin in Fig. 7(a), being circular and with its local minimum centrally located, is *compact*—its candidate structures (points in the basin area) are close-packed around its local minimum. In contrast, the basin in Fig. 7(b), being highly eccentric and with its local minimum located by a focus, is *extended*—its candidate structures extend away from its local minimum. Due solely to this shape difference, the typical descent trajectory is longer in Fig. 7(b)—the extended basin—than in Fig. 7(a)—the compact basin.

To investigate the characteristic shape of our search landscape's basins, we examine the *occupancy* of basin *k-shells*: Consider a basin. For each integer k between zero and twenty-five, our search landscape possesses $\binom{25}{k}$ candidate structures at Hamming distance k from that basin's local minimum. These candidate structures together constitute the basin's *k-shell*. The conceptual equivalent of k-shells in Fig. 7 would be a series of concentric annuli centered about the basin's local minimum. For each of a basin's k-shells, the number of candidate structures in the k-shell that belong to the basin is that k-shell's *occupancy*. For example, the basin shown in Fig. 7(a) would have complete occupancy of all k-shells lying strictly within the basin's boundary, partial occupancy of perhaps one k-shell containing the basin's boundary, and zero occupancy of all k-shells lying strictly beyond the basin's boundary. In contrast, the basin shown in Fig. 7(b) would have complete occupancy of only a few k-shells lying strictly within the basin's boundary, progressively decreasing partial occupancy of many k-shells lying partially within the basin's boundary, and then zero occupancy of k-shells lying strictly beyond the basin's boundary. The sharp transition between filled and empty k-shells in the case of Fig. 7(a) reflects that basin's compactness, whereas the partial filling of many k-shells in the case of Fig. 7(b) indicates that basin's extended character.

Figure 8 demonstrates that our search landscape's basins are far from compact: Each orange histogram bin provides the proportion of our search landscape's candidate structures that are

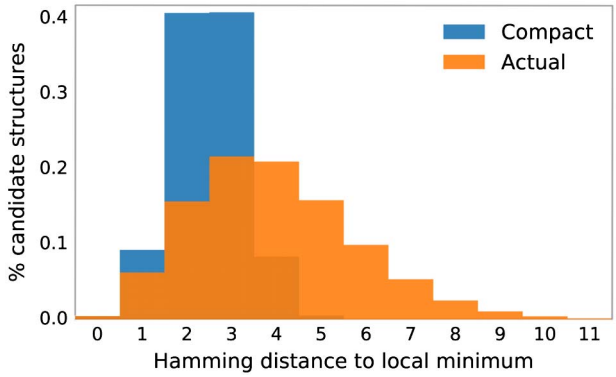

Fig. 8. Each orange histogram bin provides the proportion of our search landscape's candidate structures that are separated from corresponding local minima by the Hamming distance indicated on the horizontal axis. For comparison, the blue histogram shows the same for the hypothetical situation in which all of our search landscape's basins were strictly compact.

separated from corresponding local minima by the Hamming distance indicated on the horizontal axis. For comparison, the blue histogram shows the same for the hypothetical situation in which all of our search landscape's basins were strictly compact, generated as follows.

For each basin in our search landscape, we determined the number of candidate structures in that basin—its *cardinality*. We then considered a hypothetical basin with the same cardinality, but compact. From the k-shell occupancies of this hypothetical basin, we constructed a corresponding list of that basins' candidate-structure-to-local-minimum Hamming distances. By accumulating all such lists across all basins, we generated a data set equivalent to that from which we generated the orange histogram in Fig. 8 but for the hypothetical version of our search landscape wherein each basin is strictly compact. This data set underlies the blue histogram.

Comparing the orange (actual) and blue (hypothetical) distributions, we see that our search landscape's candidate structures distribute significantly further from corresponding local minima than would be the case were our search landscape's basins strictly compact.

Figure 9 shows the characteristic extension of our search landscape's basins more concretely: For each basin in our search landscape, we calculated the basin cardinality and occupancy of each k-shell. From that data, for each k-shell we calculated a running median of the k-shell's occupancy with increasing basin cardinality. Specifically, for each unique basin cardinality we calculated the median k-shell occupancy over all basins with that cardinality, the closest smaller cardinality, and the closest larger cardinality. The solid lines in Fig. 9 show the results of this procedure for each k-shell. For comparison, the dashed lines in Fig. 9 show k-shell occupancy with increasing basin cardinality for the hypothetical case of strictly close-packed basins.

Intuitively, each solid line in Fig. 9 illustrates how typical k-shell occupancy evolves as basin cardinality increases, by k-shell order k. The data shows that higher k-shells typically populate long before lower k-shells fill. Moreover, higher k-shells often populate well beyond the *maximum* occupancy of lower k-shells long before those lower k-shells fill. In other words, our search landscape's basins typically appear more like starfish with bulbous arms than spheres.

In sum, our search landscape's basins are extended rather than compact. This structure contributes to longer, less efficient descents. But as Fig. 8 shows, this structure should increase typical descent lengths by less than a factor of 2 over the hypothetical compact-basin case.

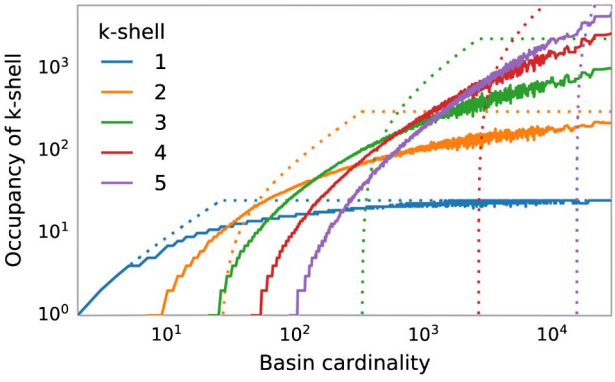

Fig. 9. Each solid line shows the typical occupancy of k-shell k with increasing basin cardinality. For comparison, each corresponding dashed line shows the hypothetical occupancy of that k-shell for a compact basin.

B. Efficient Sampling of Good Local Minima

Recall that restarted iterative best improvement distills to a strategy for sampling a search landscape's local minima via a sequence of independent local searches. If the probability of sampling an acceptable local minimum on any given local search is high, a small number of local search repetitions should suffice to find an acceptable local minimum and the search will be efficient. If instead that probability is low, a large number of local search repetitions may be required and the search will be inefficient.

The probability of sampling a *particular* local minimum in *a* given local search is the probability that local search begins within that local minimum's attraction basin. Given that we select the initial candidate structure for local search uniformly at random from our search space, the probability of a local search beginning within a particular local minimum's attraction basin is proportional to that attraction basin's cardinality.

Figure 10 illustrates this point: The two labeled circles represent attraction basins. Attraction basin B contains twice as many candidate structures (proportional to area) as attraction basin A. Hence if we select a starting candidate structure for local search (a point) uniformly at random from among all candidate structures (the figure area), we are twice as likely to begin local search within attraction basin B as within attraction basin A. Ultimately, sampling attraction basin B's local minimum is thus twice as likely as sampling attraction basin A's.

Figure 11 shows that restarted iterative best improvement's sampling of local minima skews to good local minima, specifically due to trends in basin cardinality with local minimum objective value: The blue histogram shows our search landscape's

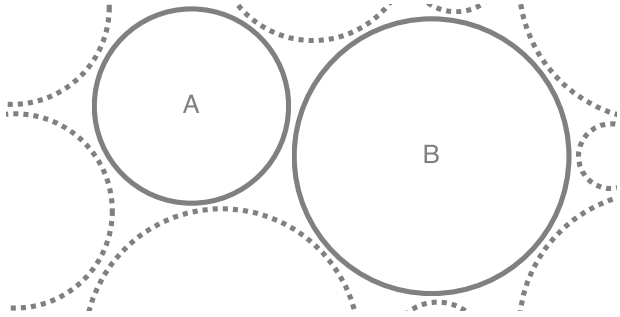

Fig. 10. Illustration of the relationship between attraction basin cardinality and sampling probability: Circles represent attraction basins. Attraction basin B contains twice as many candidate structures (proportional to area) as attraction basin A. Hence if we select a starting candidate structure for local search (a point) uniformly at random from among all candidate structures (the figure area), we are twice as likely to begin local search within attraction basin B as within attraction basin A. Ultimately, sampling attraction basin B's local minimum is thus twice as likely as sampling attraction basin A's.

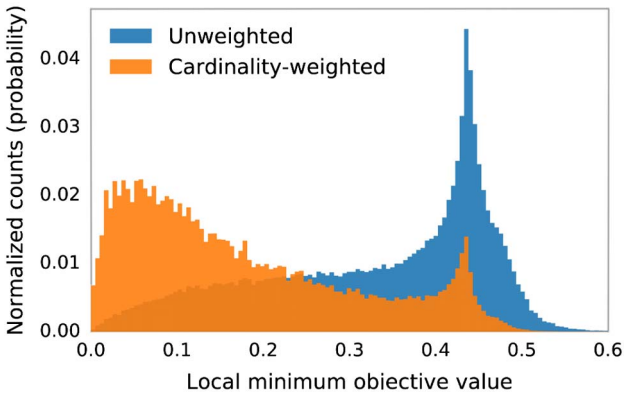

Fig. 11. The blue histogram shows our search landscape's local minima binned uniformly by objective value, and with bin height normalized such that the bin heights sum to one. The orange histogram shows the same as the blue histogram, but with each local minimum weighted by its attraction basin's cardinality. Intuitively, this figure shows that the distribution from which supersteps of restarted iterative best improvement sample local minima (orange) is strongly skewed toward good local minima relative to the bare distribution of local minima by objective value (blue).

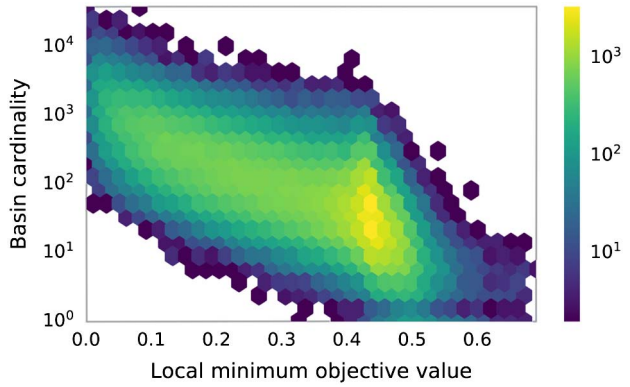

Fig. 12. Distribution of local minima by objective value on the horizontal axis and corresponding basin cardinality on the vertical. White cells represent an absence of local minima, dark cells a small number of local minima, and bright cells a large number of local minima, with logarithmic intensity scaling. Intuitively, this figure shows that local minimum attraction basin cardinality grows roughly *exponentially* with decreasing objective value.

local minima binned uniformly by objective value, and with bin height normalized such that all bin heights sum to one. Intuitively, this histogram reflects the distribution of local minima by objective value. Were local minima drawn uniformly at random from the local minima population, this histogram would approximate the underlying probability distribution. Evidently, good local minima—specifically those with objective value on the order of 0.01—represent only a small fraction of our search landscape's local minima. Hence drawing a good local minimum uniformly at random from that population would occur only rarely.

The orange histogram, however, shows the same as the blue histogram, but with each local minimum weighted by its attraction basin's cardinality. Intuitively, this histogram approximates the probability distribution from which restarted iterative best improvement draws local minima. Evidently, drawing a good local minimum in a given superstep of restarted iterative best improvement is dramatically more likely than the local minima population statistics (the blue histogram) suggest.

Figure 12 reveals the structure underlying this skew, showing the distribution of local minima by objective value on the horizontal axis and corresponding basin cardinality on the vertical. White cells represent an absence of local minima, dark cells a small number of local minima, and bright cells a large number of local minima, with logarithmic intensity scaling. Intuitively, Fig. 12 shows that local minimum attraction basin cardinality grows roughly *exponentially* with decreasing objective value.

In sum, our search landscape's structure is strongly conducive to efficient sampling of good local minima.

C. Local Minima Density

A search landscape's *local minima density*—the fraction of candidate structures that are local minima—gives some indication of the search landscape's structure, and typically suggests how difficult the associated problem is to solve via stochastic local search algorithms such as restarted iterative best improvement [45].

Consider a search landscape with the same underlying graph structure as our search landscape (the same search space and neighborhood relation) but potentially arbitrarily different objective values. (Figure 13 illustrates this concept via toy search landscapes.) If such a search landscape lacks structure or, more precisely, is *uncorrelated*, approximately one in every twenty-six structures is a local minimum [46]. In contrast, scanning our search landscape reveals that its local minima density is almost an order of magnitude lower—approximately 1 in every 235 structures.

Such comparatively low local minima density is neither necessary nor sufficient for restarted iterative best improvement to outperform uninformed random picking. But excluding pathological cases, it strongly suggests the existence of

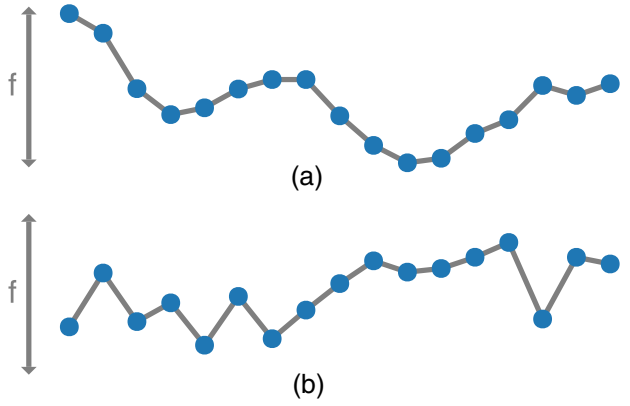

Fig. 13. (a) Our toy search landscape. (b) Search landscape with the same underlying graph structure as our toy search landscape—the same search space and neighborhood relation—but potentially arbitrarily different objective values. Additionally, this search landscape is largely *uncorrelated*: We assigned each candidate structure's objective value independently and randomly.

significant local basin structure (which restarted iterative best improvement can exploit).

D. Performance Retrodiction and Feature Impact

We have found that our search landscape possesses substantial structure relevant to restarted iterative best improvement's performance. Two questions follow: First, does the explored structure account for restarted iterative best improvement's observed performance? Second, which explored features are most important? To investigate these questions, we develop a simple model of restarted iterative best improvement's performance, ascertain that the explored features account for the observed performance via a performance retrodiction, and examine how the explored features impact the performance retrodiction.

1. Performance Model

We define the *search cost* C as the number of objective function evaluations restarted iterative best improvement requires to find a candidate structure with objective value below some threshold. Estimating this cost is straightforward: We model C as the sum of the costs of the local descents that constitute a run of restarted iterative best improvement,

$$C = \sum_{k=1}^{N_d} C_{d,k}, \tag{2}$$

where N_d is the number of descents performed to find a below-threshold, or acceptable, local minimum, and $C_{d,k}$ is the cost of the kth descent, both of which we treat as random variables. We can write the expectation of C in terms of the conditional expectation of C given N_d,

$$\mathbf{E}[C] = \mathbf{E}[\mathbf{E}[C|N_d]], \tag{3}$$

which, using Eq. (2), readily evaluates to

$$\mathbf{E}[C|N_d = n] = \mathbf{E}\left[\sum_{k=1}^{n} C_{d,k}\right] = \sum_{k=1}^{n} \mathbf{E}[C_{d,k}]. \tag{4}$$

As the $C_{d,k}$ are independent and identically distributed with common expectation $\mathbf{E}[C_d]$, the preceding conditional expectation reduces to

$$\mathbf{E}[C|N_d = n] = \sum_{k=1}^{n} \mathbf{E}[C_d] = n\mathbf{E}[C_d], \tag{5}$$

and so Eq. (3) yields

$$\mathbf{E}[C] = \mathbf{E}[N_d\mathbf{E}[C_d]] = \mathbf{E}[N_d]\mathbf{E}[C_d]. \tag{6}$$

We first consider $\mathbf{E}[N_d]$. We can view restarted iterative best improvement's sequence of descents as a Bernoulli process in which the outcomes are (1) descending to an acceptable local minimum with probability p_a and (2) to an unacceptable local minimum with complementary probability $1 - p_a$. The geometric distribution models the number of Bernoulli trials necessary to have one success, and so N_d is geometrically distributed with parameter p_a. Hence

$$\mathbf{E}[N_d] = \frac{1}{p_a}. \tag{7}$$

We now turn to $\mathbf{E}[C_d]$, which we write in terms of the conditional expectation of C_d given H_d,

$$\mathbf{E}[C_d] = \mathbf{E}[\mathbf{E}[C_d|H_d]] = \sum_{k=0}^{25} \mathbf{Pr}(H_d = k)\mathbf{E}[C_d|H_d = k], \tag{8}$$

where H_d is the Hamming distance between the local descent's starting candidate structure and the terminating local minimum. Combining Eqs. (6)–(8) yields

$$\mathbf{E}[C] = \frac{1}{p_a}\sum_{k=0}^{25} \mathbf{Pr}(H_d = k)\mathbf{E}[C_d|H_d = k]. \tag{9}$$

2. Performance Retrodiction

As above, the vast majority of local descents in our search landscape are direct. Accordingly, we approximate $\mathbf{E}[C_d|H_d = k]$, the expected cost of local descent beginning at Hamming distance $H_d = k$ from the terminating local minimum, as the cost of a length-k descent. A length-k descent visits $k + 1$ candidate structures, and for each visited candidate structure we evaluate our objective function twenty-five times—once for each of that visited candidate structure's neighbors. Hence the cost of a length-k descent is $25(k + 1)$, and slightly optimistically

$$\mathbf{E}[C_d|H_d = k] \approx 25(k + 1). \tag{10}$$

As we draw each starting structure for local descent uniformly at random from our search space, the probability of a descent beginning at Hamming distance $H_d = k$ from the terminating local minimum ($\mathbf{Pr}(H_d = k)$) is simply the probability that a given candidate structure is Hamming distance $H_d = k$ from the corresponding local minimum. Table 1's rightmost column contains the probability distribution of candidate structures by Hamming distance from their corresponding local minima, and Fig. 8 visualizes it. Including this information in our model incorporates our observation of extended basins. We integrate this information as shown in Table 2, truncating and rounding slightly pessimistically.

Finally we estimate p_a as

$$p_a \approx \frac{\sum_{m\in\mathcal{M}_A} \mathbf{Card}(m)}{2^{25}}, \tag{11}$$

where $\mathcal{M}_A$ denotes the set of local minima with objective value below acceptability threshold f_a, and $\mathbf{Card}(m)$ denotes the cardinality of local minimum m's attraction basin. The numerator is the number of candidate structures residing in attraction basins of local minima with acceptable objective values. The denominator is the total number of candidate structures in our search space. The ratio p_a is thus the probability

Table 2. Probabilities $\mathbf{Pr}(H_d = k)$ in Performance Retrodiction

$\mathbf{Pr}(H_d = 0) \approx 0.000$	$\mathbf{Pr}(H_d = 9) \approx 0.010$
$\mathbf{Pr}(H_d = 1) \approx 0.062$	$\mathbf{Pr}(H_d = 10) \approx 0.004$
$\mathbf{Pr}(H_d = 2) \approx 0.157$	$\mathbf{Pr}(H_d = 11) \approx 0.001$
$\mathbf{Pr}(H_d = 3) \approx 0.216$	$\mathbf{Pr}(H_d = 12) \approx 0.001$
$\mathbf{Pr}(H_d = 4) \approx 0.210$	$\mathbf{Pr}(H_d = 13) \approx 0.001$
$\mathbf{Pr}(H_d = 5) \approx 0.158$	$\mathbf{Pr}(H_d = 14) \approx 0.001$
$\mathbf{Pr}(H_d = 6) \approx 0.099$	$\mathbf{Pr}(H_d = 15) \approx 0.001$
$\mathbf{Pr}(H_d = 7) \approx 0.053$	$\mathbf{Pr}(H_d = 16) \approx 0.001$
$\mathbf{Pr}(H_d = 8) \approx 0.025$	$\mathbf{Pr}(H_d \geq 17) \approx 0.000$

of selecting, uniformly at random from our search space, a candidate structure residing in an attraction basin with an acceptable local minimum. The definition of p_a is equivalent to integrating the basin-cardinality-weighted local minimum distribution shown in orange in Fig. 11 from $f = 0$ through $f = f_a$, and it incorporates our observation of exponential growth of attraction basin cardinality with decreasing local minimum objective value.

With the preceding estimates of p_a, $\mathbf{Pr}(H_d = k)$, and $\mathbf{E}[C_d|H_d = k]$, and with objective value acceptability threshold $f_a = 0.01$ (approximately the discretization-error noise floor of the objective function), evaluating our performance model [Eq. (9)] yields $\mathbf{E}[C] \approx 7275$. This result agrees well with the empirical value of $\mathbf{E}[C] \approx 7659$ obtained from the 2000 runs of the restarted iterative best improvement described in Section 5 (considering the corresponding value for uninformed random picking is 123309). Evidently, the explored features account for restarted iterative best improvement's observed performance.

3. The Basin Cardinality Trend Is Key

If instead of estimating p_a as above we estimate p_a assuming uniform sampling probability among local minima,

$$p_a \approx \frac{N_a}{N_n}, \tag{12}$$

where N_a is the number of acceptable local minima and N_n the total number, $\mathbf{E}[C]$ becomes approximately 106249—about a factor of 14 worse than the observed value and comparable with uninformed random picking's performance. Our estimates of $\mathbf{E}[C_d|H_d = k]$ are already maximally optimistic. Making our estimates of $\mathbf{Pr}(H_d = k)$ unrealistically optimistic (by assuming compact basins and calculating $\mathbf{Pr}(H_d = k)$ accordingly) would improve $\mathbf{E}[C]$ by about a factor of 2 at best. Hence we conclude that the exponential growth of attraction basin cardinality with decreasing local minimum objective value is the most important feature in accounting for restarted iterative best improvement's observed performance.

7. UNDERLYING PHYSICAL INTUITION

The above analyses' results are particular to our model problem. We can, however, argue for some of the observed search landscape structure on physical grounds applicable to a broader array of systems.

A. Critical Paths and the Impact of Structural Modifications

Consider Figs. 14(a) and 14(b). Figure 14(a) shows the field induced by excitation of the first ingoing TM spatial mode on the horizontal waveguide of the candidate structure first introduced in Fig. 1(b). Figure 14(b) shows the equivalent for the second spatial mode. Notice that, for each mode, there exists some *critical path* along which the vast majority of the power flows, and off which the field is relatively weak.

Modifying this candidate structure along its critical paths may significantly impact field behavior. Figures 14(e) and 14(f) demonstrate this point, showing the equivalents of Figs. 14(a) and 14(b) for a candidate structure separated by Hamming distance one from the original candidate structure, where the modified lattice sites fall on the original candidate structure's critical paths. The modifications substantially impact field behavior.

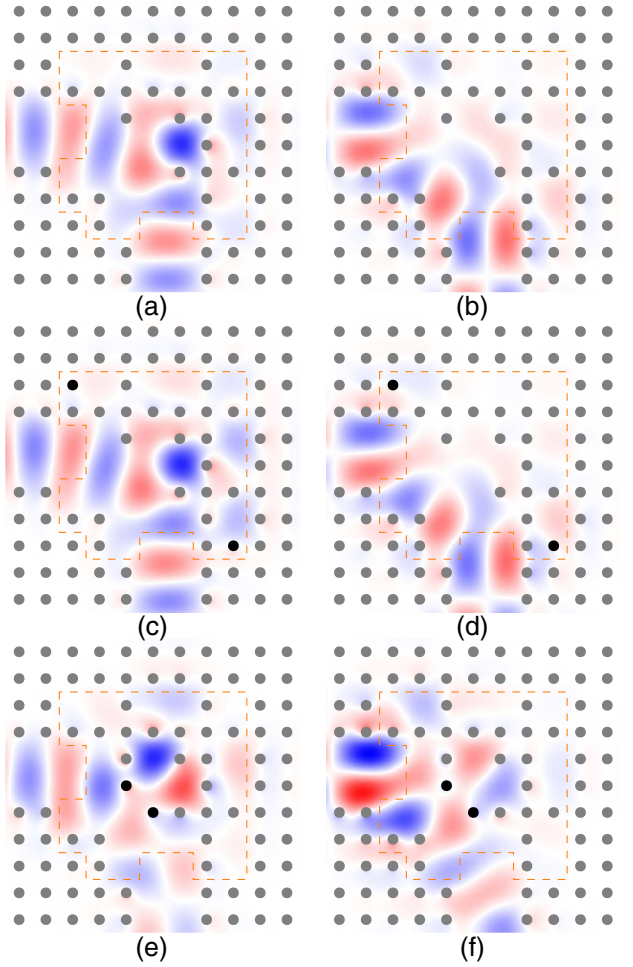

Fig. 14. (a) Field induced by excitation of the first ingoing TM spatial mode on the horizontal waveguide of the candidate structure first introduced in Fig. 1(b). The dashed orange rectangle outlines the optimization region. (b) Equivalent for the second spatial mode. (c), (d) Equivalents of (a) and (b) for a candidate structure separated by Hamming distance one from that in (a) and (b), where the modified lattice sites (black) fall off the critical paths of the candidate structure in (a) and (b); the modifications have little impact on behavior. (e), (f) Equivalents of (a) and (b) for another candidate structure separated by Hamming distance one from that in (a) and (b), but where the modified lattice sites (black) fall on the critical paths of the candidate structure in (a) and (b); the modifications substantially impact behavior.

In contrast, modifying the original candidate structure *off* its critical paths typically should impact field behavior relatively little. Figures 14(c) and 14(d) illustrate this point, showing the equivalents of Figs. 14(a) and 14(b) for a candidate structure separated by Hamming distance one from the original candidate structure, where the modified lattice sites fall off the original candidate structure's critical paths. The modifications impact field behavior little.

When generalized, this observation seems almost self-evident: In the typical case, structural perturbations where fields are strong may significantly impact field behavior, whereas structural perturbations where fields are weak should impact field behavior relatively little. Of course, the utility of this observation in design optimization depends critically on the extent of the fields over the optimization region.

B. General Consequences for Search Landscape Structure

From this observation, we can argue for general, statistical consequences for search landscape structure: Consider a pair of candidate structures separated by Hamming distance one. In the case where these candidate structures differ only off their critical paths, the intuition above implies that the expected difference between these candidate structures' objective values should be small relative to objective function variation over the entire search space. Hence, to the degree that this case occurs, the typical objective value difference within a pair of candidate structures separated by Hamming distance one should be small relative to variation over the entire search space. In more precise terms (discussed further below), the objective values of candidate structures separated by Hamming distance one should be *correlated*.

This argument generalizes immediately to pairs of candidate structures separated by Hamming distances greater than one, but with two caveats: First, though individual off-critical-path structural modifications should impact field behavior little, the expected difference in field behavior within a pair of candidate structures should grow as the number of off-critical-path modifications differentiating the pair grows. Second, as the number of modifications differentiating a pair of candidate structures grows, the likelihood that all modifications lie off the candidate structures' critical paths diminishes. Together, these caveats imply that the typical field behavior, and thus objective value, difference within a pair of candidate structures separated by Hamming distance h should grow with increasing h, from small to large relative to typical objective value variation over the entire search space. In more precise terms (discussed further below), the objective values of candidate structures separated by Hamming distance h should be correlated for small separation h, and that correlation should diminish with increasing separation h, eventually becoming *anticorrelation* where separation h approaches the maximal separation.

Intuitively, the above says that our search landscape should exhibit some measure of local *smoothness*. Figure 13(a) illustrates the *smooth* character of a search landscape with local correlations by way of contrast with the uncorrelated landscape in Fig. 13(b). Fundamentally, this correlative search landscape structure underpins the specific search landscape features investigated above and is the foundation on which we can build efficient search algorithms. We can make arguments analogous to those above for a broad array of optical systems, particularly those where we can argue for the existence of critical paths and reason about their extent relative to the optimization region.

1. Verification

We can verify the above intuition by calculating the *landscape correlation function* [47] for our search landscape,

$$\rho(h) = \frac{\mathbf{E}_h[(f(s) - \mathbf{E}[f])(f(s') - \mathbf{E}[f])]}{\mathbf{Var}[f]}, \tag{13}$$

where $\mathbf{E}[f]$ and $\mathbf{Var}[f]$ are the expectation and variance of objective function f over all candidate structures in search space $\mathcal{S}$, and the expectation $\mathbf{E}_h$ is over all pairs of candidate structures (s, s') separated by Hamming distance h, or $\{(s, s') \in \mathcal{S} \times \mathcal{S} \,|\, \mathbf{H}(s, s') = h\}$. For fixed Hamming distance, this most closely resembles Fischer's intraclass correlation coefficient [48].

Intuitively, the landscape correlation function provides a measure of how similar the objective values of pairs of candidate structures separated by Hamming distance h tend to be, relative to objective function variation over the entire search space. This measure provides insight into the search landscape's *smoothness* or *ruggedness* and thereby gives some indication of the degree to which search landscape structure exists. For given h, a landscape correlation value approaching one indicates pairs of candidate structures with that separation tend to have similar objective values, a value approaching negative one indicates that such pairs' objective values tend to opposing ends of the objective value spectrum, and a value around zero indicates the absence of a detectable (linear) relationship between such pairs' objective values.

Figure 15 shows the landscape correlation function for our search landscape. On the left we see values approaching one, indicating substantial objective value similarity among candidate structures separated by short Hamming distance. Furthermore, the correlation decreases with increasing Hamming distance, eventually becoming weakly negative (anticorrelation) as Hamming distance h approaches its maximum of twenty-five. These results grossly confirm the intuition presented in the preceding sections.

C. Specific Structural Consequences

Arguing intuitively for specific search landscape features observed above is also possible, though we primarily regard the

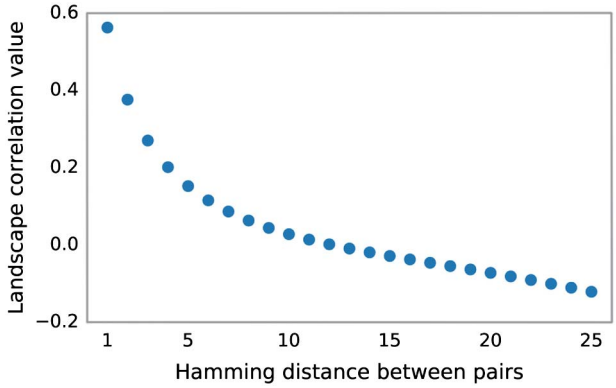

Fig. 15. Our search landscape's correlation function. For given Hamming distance h, a landscape correlation function value approaching one indicates pairs of candidate structures with that separation tend to have similar objective values, a value approaching negative one indicates that such pairs' objective values tend to opposing ends of the objective value spectrum, and a value around zero indicates the absence of a detectable (linear) relationship between such pairs' objective values.

following as interesting speculation. We begin by arguing for a form of local basin structure, continue with arguments for rapid basin cardinality growth with basin depth, and close by touching on basin extension.

1. Local Basin Structure

Consider a set of candidate structures that differ from each other only at lattice sites lying off their critical paths. Per the preceding sections' arguments, these candidate structures should exhibit largely similar field behavior and thus objective values, forming a basin or collection of closely related basins—hereafter a *pseudobasin*. Figure 16(a) simplistically illustrates a pair of adjacent pseudobasins separated by an on-critical-path modification: The blue nodes on the left represent one pseudobasin, the purple nodes on the right represent the other pseudobasin, and the orange node separating the blue and purple node collections represents the on-critical path modification.

The set of all candidate structures constituting off-critical-path modifications to a given candidate structure, or some subset thereof, likely constitutes a pseudobasin in which the given candidate structure lies. Hence our search space may consist of a large collection of pseudobasins, providing a plausible intuitive origin for our search landscape's observed local basin structure.

2. Basin Cardinality Growth with Depth

The preceding sections' arguments also imply that, within a pseudobasin, candidate structures' objective values should be correlated, with neighboring candidate structures' objective values particularly strongly correlated. For simplicity, we ignore objective value correlations beyond those between neighboring candidate structures and, admittedly simplistically, model objective value correlations between neighbors as hard constraints. Figure 16(b) illustrates this conceptualization.

Consider progressively perturbing a candidate structure in the right pseudobasin, specifically the candidate identified by the arrow in Fig. 17(a). Due to the objective value correlations between neighboring candidate structures within a pseudobasin, as the perturbed candidate structure drops in objective value, its basin, highlighted in orange in each panel, progressively claims not only the immediately adjacent candidate structures (distance $k = 1$), but also those at distance $k = 2$ and then $k = 3$ and so on. In the toy search landscape shown, the number of neighbors at distance k from a given candidate structure does not grow with k. Hence the illustrated basin cardinality growth is approximately linear. In our search landscape, however, the number of candidate structures at Hamming distance k from a given candidate structure grows as $\binom{25}{k}$. Hence, under this model, analogous basin cardinality growth in our search landscape would be rapid, perhaps commensurate with our above observation of roughly exponential basin cardinality growth with basin depth.

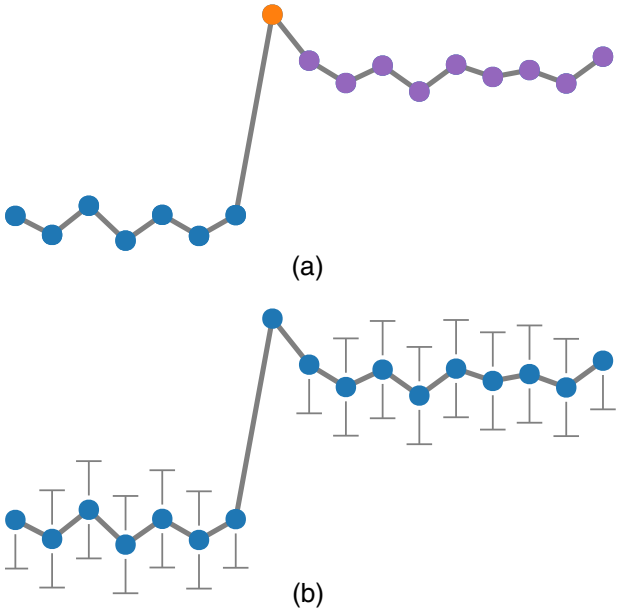

Fig. 16. (a) simplistically illustrates a pair of adjacent pseudobasins separated by an on-critical-path modification: The blue nodes on the left represent one pseudobasin, the purple nodes on the right represent the other pseudobasin, and the orange node separating the blue and purple node collections represents the on-critical path modification. (b) schematizes a simplistic model of the objective value correlations within a pseudobasin. The up and down tees extending from each node within the pseudobasins represent the hard constraints on relative position of neighboring candidate structures in our model.

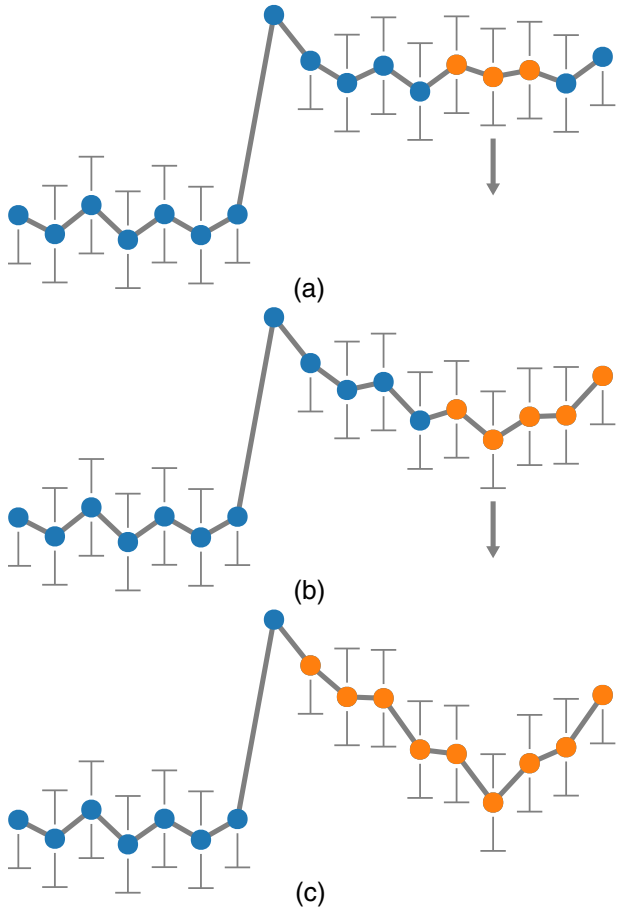

Fig. 17. Panels (a) through (c) illustrate the impact on basin structure of a sequence of perturbations to the objective value of the candidate structure identified by the arrow in (a).

3. Basin Extension

Additionally consider that, due to the similarity of objective values within a pseudobasin, the objective-value rank-ordering of a pseudobasin's candidate structures can change rapidly with perturbation of their objective values. Due to the relatively large differences in objective values *between* pseudobasins, however, changes in rank-ordering *between* different pseudobasins should happen much less readily. Figure 17 illustrates this point, as the objective-value rank-ordering of the right pseudobasin's candidate structures changes completely before any changes in objective-value rank-ordering *between* the pseudobasins occur. This insight provides a plausible origin for the extended rather than compact nature of our search landscape's basins: As a basin deepens and so grows, it accrues additional structures within its pseudobasin more rapidly than it accrues nearby structures lying outside of its pseudobasin. This growth behavior occurs even though those candidate structures within the pseudobasin may be significantly farther from the basin's local minimum than candidate structures lying outside the pseudobasin (as the latter potentially include candidate structures constituting even minor off-critical-path modifications of candidate structures in the pseudobasin).

8. DISCUSSION AND CONCLUSION

Though the search landscape analysis presented above addresses only a single frequency, results appear qualitatively similar in the band over which the candidate structures' waveguides simultaneously support the two spatial modes considered. Additionally, results are qualitatively similar for the multi-spatial-mode photonic crystal T-splitter discussed in [12]. Hence the specific search landscape features observed appear at least in a broader set of similarly constructed multi-spatial-mode photonic crystal devices.

As previously discussed, the general arguments regarding critical paths and resultant search landscape structure should generalize to a still broader array of optical systems, and the consequent general, statistical search landscape structure should accompany them. These arguments possess additional features worth noting.

In some systems, that presented being an example, we expect adding volume to the optimization region to decrease the typical relative extent of critical paths. Hence the expanded search landscape may possess additional exploitable structure, potentially enabling mitigation of search cost growth with search space expansion. Counterintuitively, in some cases search cost may even decline with search space expansion.

The extent of critical paths varies over a search space, perhaps somewhat predictably. As an example in our search space, candidate structures that exhibit strong reflection from cylinders impinging on the input waveguide's insertion should possess relatively small critical paths (primarily confined to the input waveguide). On the other hand, candidate structures that are structurally open should have relatively large critical paths. This feature may be exploitable.

Arguments regarding critical path extent beyond merely small relative size are also interesting. For example, knowing that critical paths are confined to or away from a particular region, independent of relative size, is informative and likely exploitable. Being able to argue for *large* relative size of critical paths is also informative.

Increasing the number of independent channels involved in the objective function—frequential or spatial—may homogenize the objective function. Hence search landscapes for spatially or frequentially broadband objective functions may be smoother or, more precisely, may exhibit more correlative structure than their narrowband counterparts. This additional structure may be exploitable, potentially mitigating the expected difficulty of design optimization of broadband systems relative to narrowband counterparts. We suspect a similar phenomenon should exist for robust objectives.

Restarted iterative best improvement's simplicity served this work's end—illustration of search landscape analysis's potential value in design optimization of optical structures—better than more complex algorithms. But certainly more sophisticated and effective algorithms for the presented problem exist.

Similarly, of necessity this work merely scratches the surface of search landscape analysis. A variety of techniques exist for various parts of the problem class and size spectrum. For a more complete and formal (yet lucid) exposition on search algorithms and search landscape analysis for combinatorial optimization, see [45].

Search landscape analysis can provide valuable insight into design optimization problems in optics. Such insight sheds light on the path to more effective design optimization algorithms and ultimately better designs.

APPENDIX A

Figure 18 shows the field behavior of a few high-performing candiate structures. Each row of panes corresponds to a candidate structure. The leftmost column in the row shows the candidate structure in isolation and provides the candidate structure's objective value. The middle and leftmost columns in the row respectively show the field patterns resulting from excitation of the first and second spatial modes on the candidate structure's input waveguide.

We selected these candidate structures uniformly at random from the subset of candidate structures with objective value below the objective function's discretization-error noise threshold (∼0.01). These five candidate structures constitute a diverse sample of our search space's high-performing candidate structures.

These structures exhibit remarkably varied behavior. For example, structures one through three exhibit strong, resonant field localizations, contrasting with the behavior of structures four and five.

In each pane, colors reflect field intensity relative to that pane's peak field intensity. We separately normalized each pane's intensity, as showing the fields on a common scale would unacceptably compress the effective dynamic range in some panes and saturate it in others. The necessity for separate normalization is due to the remarkably varied behavior of these structures as mentioned above.

Structures with differences in objective value below the objective function's discretization-error noise threshold should be considered equally well performing. Given that the shown candidate structures' objective values fall below that threshold, they

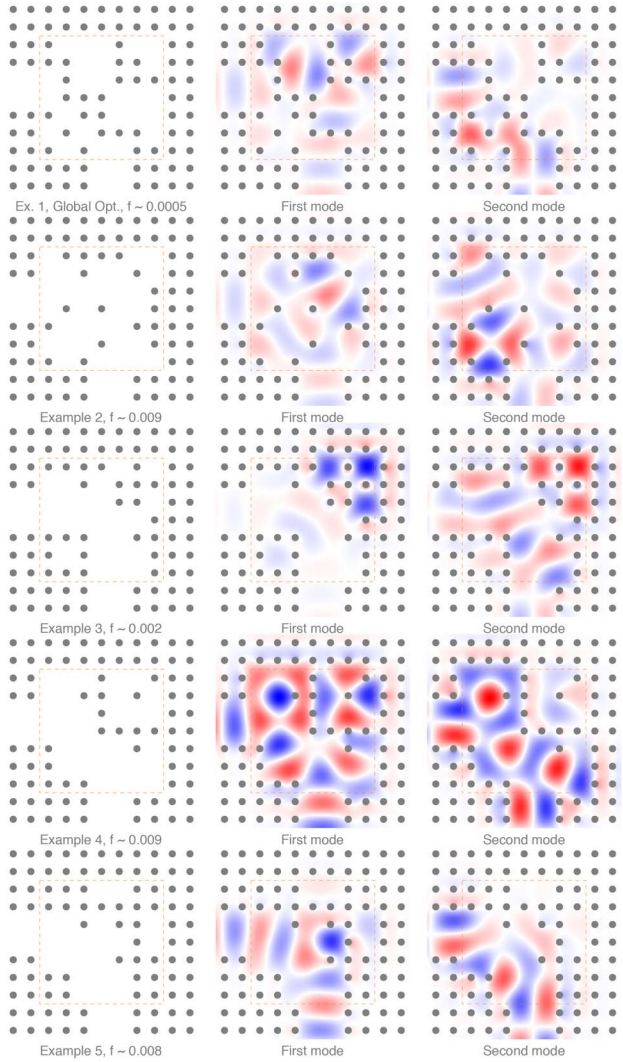

Fig. 18. Field behavior of a few high-performing candidate structures.

should be considered equally well performing to within our calculation's resolution. We show these objective values simply to indicate the high-performing nature of these structures (and truncate them after a single nontrivial digit as even that digit's significance is questionable).

Funding. Department of Energy Office of Science (SC) (DE-AC05-06OR23100); National Science Foundation (NSF); Air Force Office of Scientific Research (AFOSR) (FA9550-15-1-0335).

Acknowledgment. This research was supported in part by the Department of Energy Office of Science Graduate Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract DE-AC05-06OR23100. This research was also supported in part by the National Science Foundation through XSEDE resources provided by the XSEDE Science Gateways program, and the United States Air Force Office of Scientific Research (USAFOSR) grant FA9550-15-1-0335. We thank J. E. Herriman for suggestions that considerably improved the paper. We thank an anonymous reviewer of [12] for sparking our interest in this topic.

REFERENCES

1. J. Jensen and O. Sigmund, "Systematic design of photonic crystal structures using topology optimization: Low-loss waveguide bends," Appl. Phys. Lett. **84**, 2022–2024 (2004).
2. J. Jensen and O. Sigmund, "Topology optimization of photonic crystal structures: a high-bandwidth low-loss T-junction waveguide," J. Opt. Soc. Am. B **22**, 1191–1198 (2005).
3. N. Ikeda, Y. Sugimoto, Y. Watanabe, N. Ozaki, A. Mizutani, Y. Takata, J. Jensen, O. Sigmund, P. Borel, M. Kristensen, and K. Asakawa, "Topology optimised photonic crystal waveguide intersections with high-transmittance and low crosstalk," Electron. Lett. **42**, 1031–1032 (2006).
4. L. Frandsen, Y. Elesin, O. Sigmund, J. Jensen, and K. Yvind, "Wavelength selective 3D topology optimized photonic crystal devices," in *CLEO* (Optical Society of America, 2013), paper CTh4L-6.
5. Y. Elesin, B. Lazarov, J. Jensen, and O. Sigmund, "Time domain topology optimization of 3D nanophotonic devices," Photon. Nanostr. Fundam. Appl. **12**, 23–33 (2014).
6. A. Piggott, J. Lu, T. Babinec, K. Lagoudakis, J. Petykiewicz, and J. Vučković, "Inverse design and implementation of a wavelength demultiplexing grating coupler," Sci. Rep. **4**, 7210 (2014).
7. A. Piggott, J. Lu, K. Lagoudakis, J. Petykiewicz, T. Babinec, and J. Vučković, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," Nat. Photonics **9**, 374–377 (2015).
8. J. Lu and J. Vučković, "Objective-first design of high-efficiency, small-footprint couplers between arbitrary nanophotonic waveguide modes," Opt. Express **20**, 7221–7236 (2012).
9. L. Frandsen, Y. Elesin, L. Frellsen, M. Mitrovic, Y. Ding, O. Sigmund, and K. Yvind, "Topology optimized mode conversion in a photonic crystal waveguide fabricated in silicon-on-insulator material," Opt. Express **22**, 8525–8532 (2014).
10. Y. Jiao, S. Fan, and D. Miller, "Systematic photonic crystal device design: global and local optimization and sensitivity analysis," IEEE J. Quantum Electron. **42**, 266–279 (2006).
11. V. Liu and S. Fan, "Compact bends for multi-mode photonic crystal waveguides with high transmission and suppressed modal crosstalk," Opt. Express **21**, 8069–8075 (2013).
12. S. Verweij, V. Liu, and S. Fan, "Accelerating simulation of ensembles of locally differing optical structures via a Schur complement domain decomposition," Opt. Lett. **39**, 6458–6461 (2014).
13. V. Liu, Y. Jiao, D. Miller, and S. Fan, "Design methodology for compact photonic-crystal-based wavelength division multiplexers," Opt. Lett. **36**, 591–593 (2011).
14. X. Sheng, S. Johnson, J. Michel, and L. Kimerling, "Optimization-based design of surface textures for thin-film Si solar cells," Opt. Express **19**, A841–A850 (2011).
15. C. Lin and M. Povinelli, "Optimal design of aperiodic, vertical silicon nanowire structures for photovoltaics," Opt. Express **19**, A1148–A1154 (2011).
16. V. Ganapati, O. Miller, and E. Yablonovitch, "Light trapping textures designed by electromagnetic optimization for subwavelength thick solar cells," IEEE J. Photovolt. **4**, 175–182 (2013).
17. N. Sergeant, O. Pincon, M. Agrawal, and P. Peumans, "Design of wide-angle solar-selective absorbers using aperiodic metal-dielectric stacks," Opt. Express **17**, 22800–22812 (2009).
18. P. Bermel, M. Ghebrebrhan, W. Chan, Y. Yeng, M. Araghchini, R. Hamam, C. Marton, K. Jensen, M. Soljačić, J. Joannopoulos, S. Johnson, and I. Celanovic, "Design and global optimization of high-efficiency thermophotovoltaic systems," Opt. Express **18**, A314–A334 (2010).

19. P. Borel, A. Harpøth, L. Frandsen, M. Kristensen, P. Shi, J. Jensen, and O. Sigmund, "Topology optimization and fabrication of photonic crystal structures," Opt. Express **12**, 1996–2001 (2004).
20. J. Jensen and O. Sigmund, "Topology optimization for nano-photonics," Laser Photon. Rev. **5**, 308–321 (2011).
21. C. Y. Kao, S. Osher, and E. Yablonovitch, "Maximizing band gaps in two-dimensional photonic crystals by using level set methods," Appl. Phys. B **81**, 235–244 (2005).
22. C. Lalau-Keraly, S. Bhargava, O. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," Opt. Express **21**, 21693–21701 (2013).
23. J. Lu and J. Vučković, "Inverse design of nanophotonic structures using complementary convex optimization," Opt. Express **18**, 3793–3804 (2010).
24. J. Lu and J. Vučković, "Nanophotonic computational design," Opt. Express **21**, 13351–13367 (2013).
25. J. Andkjaer, V. Johansen, K. Friis, and O. Sigmund, "Inverse design of nanostructured surfaces for color effects," J. Opt. Soc. Am. B **31**, 164–174 (2014).
26. D. Liu, L. Gabrielli, M. Lipson, and S. Johnson, "Transformation inverse design," Opt. Express **21**, 14223–14243 (2013).
27. F. Poletti, V. Finazzi, T. Monro, N. Broderick, V. Tse, and D. Richardson, "Inverse design and fabrication tolerances of ultra-flattened dispersion holey fibers," Opt. Express **13**, 3728–3736 (2005).
28. A. Oskooi, A. Mutapcic, S. Noda, J. Joannopoulos, S. Boyd, and S. Johnson, "Robust optimization of adiabatic tapers for coupling to slow-light photonic-crystal waveguides," Opt. Express **20**, 21558–21575 (2012).
29. S. Bhargava and E. Yablonovitch, "Lowering HAMR near-field transducer temperature via inverse electromagnetic design," IEEE Trans. Magn. **51**, 1–7 (2015).
30. A. Gondarenko and M. Lipson, "Low modal volume dipole-like dielectric slab resonator," Opt. Express **16**, 17689–17694 (2008).
31. P. Seliger, M. Mahvash, C. Wang, and A. Levi, "Optimization of aperiodic dielectric structures," J. Appl. Phys. **100**, 034310 (2006).
32. W. Frei, D. Tortorelli, and H. Johnson, "Geometry projection method for optimizing photonic nanostructures," Opt. Lett. **32**, 77–79 (2007).
33. A. Hakansson, P. Sanchis, J. Sanchez-Dehesa, and J. Marti, "High-efficiency defect-based photonic-crystal tapers designed by a genetic algorithm," J. Lightwave Technol. **23**, 3881–3888 (2005).
34. E. Kerrinckx, L. Bigot, M. Douay, and Y. Quiquempois, "Photonic crystal fiber design by means of a genetic algorithm," Opt. Express **12**, 1990–1995 (2004).
35. V. Liu, D. Miller, and S. Fan, "Highly tailored computational electromagnetics methods for nanophotonic design and discovery," Proc. IEEE **101**, 484–493 (2013).
36. M. Kumar, S. Menabde, S. Yu, and N. Park, "Directional emission from photonic crystal waveguide terminations using particle swarm optimization," J. Opt. Soc. Am. B **27**, 343–349 (2010).
37. J. Jiang, J. Cai, G. Nordin, and L. Li, "Parallel microgenetic algorithm design for photonic crystal and waveguide structures," Opt. Lett. **28**, 2381–2383 (2003).
38. W. Bogaerts, V. Wiaux, D. Taillaert, S. Beckx, B. Luyssaert, P. Bienstman, and R. Baets, "Fabrication of photonic crystals in silicon-on-insulator using 248-nm deep UV lithography," IEEE J. Sel. Top. Quantum Electron. **8**, 928–934 (2002).
39. W. Bogaerts, R. Baets, P. Dumon, V. Wiaux, S. Beckx, D. Taillaert, B. Luyssaert, J. Van Campenhout, P. Bienstman, and D. Van Thourhout, "Nanophotonic waveguides in silicon-on-insulator fabricated with CMOS technology," J. Lightwave Technol. **23**, 401–412 (2005).
40. S. Selvaraja, P. Jaenen, W. Bogaerts, D. Van Thourhout, P. Dumon, and R. Baets, "Fabrication of photonic wire and crystal circuits in silicon-on-insulator using 193-nm optical lithography," J. Lightwave Technol. **27**, 4076–4083 (2009).
41. C. Poulton, X. Zeng, M. Wade, J. Shainline, J. Orcutt, and M. Popović, "Photonic crystal microcavities in a microelectronics 45-nm SOI CMOS technology," IEEE Photon. Technol. Lett. **27**, 665–668 (2015).
42. J. Orcutt and R. Ram, "Photonic device layout within the foundry CMOS design environment," IEEE Photon. Technol. Lett. **22**, 544–546 (2010).
43. J. Orcutt, A. Khilo, C. Holzwarth, M. Popović, H. Li, J. Sun, T. Bonifield, R. Hollingsworth, F. Kärtner, H. Smith, and V. Stojanović, "Nanophotonic integration in state-of-the-art CMOS foundries," Opt. Express **19**, 2335–2346 (2011).
44. Z. Hu and Y. Lu, "Efficient analysis of photonic crystal devices by Dirichlet-to-Neumann maps," Opt. Express **16**, 17383–17399 (2008).
45. H. Hoos and T. Stützle, *Stochastic Local Search: Foundations & Applications* (Morgan Kaufmann, 2004).
46. S. Kauffman and S. Levin, "Towards a general theory of adaptive walks on rugged landscapes," J. Theor. Biol **128**, 11–45 (1987).
47. P. Stadler, "Landscapes and their correlation functions," J. Math. Chem. **20**, 1–45 (1996).
48. R. Fisher, *Statistical Methods for Research Workers* (Oliver & Boyd, 1925).