

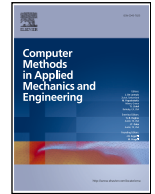


|              |                                                                                          |
|--------------|------------------------------------------------------------------------------------------|
| Title        | Wasserstein crossover for evolutionary algorithm-based topology optimization             |
| Author(s)    | Kii, Taisei; Yaji, Kentaro; Teramoto, Hiroshi et al.                                     |
| Citation     | Computer Methods in Applied Mechanics and Engineering. 2026, 451, p. 118713              |
| Version Type | VoR                                                                                      |
| URL          | <a href="https://hdl.handle.net/11094/104020">https://hdl.handle.net/11094/104020</a>    |
| rights       | This article is licensed under a Creative Commons Attribution 4.0 International License. |
| Note         |                                                                                          |

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka



# Wasserstein crossover for evolutionary algorithm-based topology optimization

Taisei Kii <sup>a</sup>, Kentaro Yaji <sup>a,\*</sup>, Hiroshi Teramoto <sup>b</sup>, Kikuo Fujita <sup>a</sup>

<sup>a</sup> Department of Mechanical Engineering, Graduate School of Engineering, The University of Osaka, 2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan

<sup>b</sup> Department of Mathematics, Faculty of Engineering Science, Kansai University, 3-3-35 Yamate-cho, Suita, Osaka, 564-8680, Japan

## ARTICLE INFO

### Keywords:

Topology optimization  
Evolutionary algorithm  
Wasserstein distance  
Maximum stress minimization  
Turbulent heat transfer

## ABSTRACT

Evolutionary algorithms (EAs) are promising approaches for non-differentiable or strongly multimodal topology optimization problems, but they often suffer from the curse of dimensionality, generally leading to low-resolution optimized results. This limitation stems in part from the difficulty of producing effective offspring through traditional crossover operators, which struggle to recombine complex parent design features in a meaningful way. In this paper, we propose a novel crossover operator for topology optimization, termed *Wasserstein crossover*, and develop a corresponding EA-based optimization framework. Our method leverages a morphing technique based on the Wasserstein distance—a distance metric between probability distributions derived from the optimal transport theory. Its key idea is to treat material distributions as probability distributions and generate offspring as Wasserstein barycenters, enabling smooth and interpretable interpolation between parent designs while preserving their structural features. The proposed framework incorporates Wasserstein crossover into an EA under a multifidelity design scheme, where low-fidelity optimized initial designs evolve through iterations of Wasserstein crossover and selection based on high-fidelity evaluation. We apply the proposed framework to three topology optimization problems: maximum stress minimization in two- and three-dimensional structural mechanics, and turbulent heat transfer in two-dimensional thermofluids. The results demonstrate that candidate solutions evolve iteratively toward high-performance designs through Wasserstein crossover, highlighting its potential as an effective crossover operator and validating the usefulness of the proposed framework for solving intractable topology optimization problems.

## 1. Introduction

Topology optimization, first proposed by Bendsoe and Kikuchi [1], offers a key advantage in its exceptionally high degree of design freedom. Its fundamental concept is to optimize the material distribution within a given design domain by mathematical programming under a computational model of physical phenomena. Typical topology optimization methods, such as the density-based method [2] and the level set method [3], update design variables sequentially based on the sensitivity to the evaluation functions. While these sensitivity-based approaches have high efficiency for convergence to reasonable optimized solutions, they inherently have significant challenges. One of the major limitations is the need for differentiability of the evaluation functions, which is often not satisfied in practical applications. For example, in stress-based topology optimization, the maximum stress is often approximated by  $p$ -norm or Kresselmeier-Steinhauser functions [4], which are differentiable and continuous ones for the sake of sensitivity analysis. Another

\* Corresponding author.

E-mail address: [yaji@mech.eng.osaka-u.ac.jp](mailto:yaji@mech.eng.osaka-u.ac.jp) (K. Yaji).

<https://doi.org/10.1016/j.cma.2025.118713>

Received 3 October 2025; Received in revised form 2 December 2025; Accepted 24 December 2025

Available online 3 January 2026

0045-7825/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

challenge is the optimization of strongly multimodal functions. A prominent example is turbulent flow in fluid topology optimization, where the complex physical phenomena lead to highly multimodal evaluation functions. In such cases, extensive parameter studies on optimization algorithms are often required to obtain a reasonable design, yet they still may be trapped in poor local optima. As a result, most studies in fluid topology optimization have focused on laminar flows under low Reynolds number conditions [5]. While such approximations have facilitated the application of topology optimization in a wide range of fields, there is potential to achieve even higher-performance designs through optimization approaches that can directly deal with the original evaluation functions and physics.

One of the promising approaches for solving such intractable optimization problems is the use of heuristic methods based on evolutionary algorithms (EAs) [6,7], which constitute a representative class of the broader field of evolutionary computation (EC). Genetic algorithms (GAs) [8], which are population-based methods inspired by the process of natural selection, are a typical example of EAs. In GAs, candidate solutions are encoded as chromosomes and handled as individuals within a population. The population evolves over successive generations through the repeated three genetic operators: selection, which favors individuals with higher fitness; crossover, which recombines features from selected parents to generate offspring; and mutation, which introduces random variations. In these optimization processes, the objective and constraint functions are evaluated as fitness values, without requiring their sensitivity information. Furthermore, compared to the local search behavior of sensitivity-based methods, the population-based emergent optimization mechanisms of EAs enable effective global search. As a result, they are well-suited for tackling intractable optimization problems that are non-differentiable or exhibit strong multimodality.

Recently, various EA-based topology optimization methods have been proposed, with innovations in both encoding material distributions as chromosomes and selecting appropriate algorithms, as summarized by Guirguis et al. [9]. The simplest approach, as developed by Chapman et al. [10], Jakiela et al. [11], and Wang and Tai [12], is GA-based topology optimization with bit array representation. As examples of GA-based approaches employing different encoding schemes, Madeira et al. [13] and Nimura and Oyama [14] proposed graph and quadtree representation-based methods, respectively. As other examples using different algorithms within the broader EC framework with bit array representation, Wu and Theng [15] and Luh et al. [16] employed differential evolution and particle swarm optimization, respectively. While these methods can also be applied to intractable problems due to the strength of EAs, EA-based topology optimization is typically constrained by a limitation commonly known as the curse of dimensionality, which refers to the exponential deterioration of search performance as the number of design variables increases. Due to this limitation, most EAs have been applied to optimization problems with relatively few design variables, typically on the order of a thousand or less, and EA-based topology optimization has likewise been carried out on coarsely discretized design domains. Sigmund [17] has emphasized that such coarse discretizations, imposed by the curse of dimensionality, inevitably lead to low-resolution optimized designs, and highlighted this limitation contrasting with the high-resolution designs achievable by the density-based method. Nevertheless, several pioneering works have sought to overcome this challenge by employing more sophisticated encoding schemes or advanced algorithms to achieve finer optimized designs. For example, Fujii [18] and Tanaka and Fujii [19] developed approaches that combine level set boundary representations with an advanced EA known as the covariance matrix adaptation evolution strategy (CMA-ES). In addition, Furuta et al. [20] incorporated a Karhunen-Loève expansion, which enables more compact representations of material distributions, into a differential evolution algorithm.

As discussed above, although a wide variety of EA-based topology optimization methods have been developed, most of them focus on how to represent material distributions as chromosomes and which algorithm to use. While these algorithmic aspects are important in EA-based approaches, it is also crucial to consider how effectively new candidate solutions can be generated within the algorithm. Especially, as crossover plays its central role in EA-based approaches, it is one of the key factors that largely determines the overall optimization performance of EAs. For example, the simplest crossover strategies, such as exchanging parts of material distributions represented as  $\{0, 1\}$  binary vectors, or linearly interpolating ones represented as  $[0, 1]$  continuous vectors in the manner of the density-based method, often lead to disconnected designs. In other words, in topology optimization, traditional crossover operators struggle to recombine parent design features in a meaningful way. Therefore, we believe that the design of effective crossover operators is essential for the successful development of EA-based topology optimization.

As one of the pioneering works aiming to overcome the limitations in EA-based topology optimization through the development of crossover strategies, Yamasaki et al. proposed a framework called data-driven topology design (DDTD) [21]. Its basic concept is to combine EAs with the idea of data-driven design [22] by employing a deep generative model to perform the role of crossover. Deep generative models, such as variational autoencoders (VAEs) [23] and generative adversarial networks (GANs) [24], are capable of generating an arbitrary dataset similar to those in the training dataset. DDTD takes advantage of this property by employing the VAE as a crossover operator to produce new candidate solutions that inherit features from a parent dataset, and iteratively updates the population based on the GA. Leveraging the strength of EAs for intractable problems, DDTD has recently been applied to a wide range of topology optimization problems such as structural mechanics [21,25–28], thermofluids [28–32], and electromagnetics [33], including more advanced applications like solid-porous infill structure design [34] and concurrent optimization of multiple design variable fields [35]. While the use of VAEs serves as a rational crossover operator for topology optimization, their inherent nature poses certain challenges. Specifically, the latent space constructed by compressing the material distributions of the population typically corresponds to a limited subspace of the vast overall design space in topology optimization. Consequently, generating new candidates through sampling from this subspace raises concerns regarding its suitability as an emergent crossover mechanism capable of global search. Moreover, it is also known that VAEs have limitations in the data dimensionality which can be effectively learned [27,28], which requires incorporating auxiliary techniques such as principal component analysis (PCA) [27] or image fragmented learning [28] so that DDTD can be applied to three-dimensional topology optimization problems. Thus, a crossover operator based on generative

models does not fully resolve the challenges in EA-based topology optimization, and the development of more effective crossover operators remains an important issue.

In this paper, we introduce a novel crossover operator for topology optimization, termed *Wasserstein crossover*, which applies a morphing technique based on the Wasserstein distance [36]. The Wasserstein distance [37], which is also known as the earth mover's distance [38], is a distance metric between probability distributions based on optimal transport theory [39]. Solomon et al. [36] developed a morphing technique for its application to graphics. This morphing approach involves treating pixel or voxel images as probability distributions and calculating their intermediate distribution based on optimal transport, i.e., the Wasserstein barycenter. Inspired by this idea, we adopt density-based representations of material distributions, following the density-based topology optimization method, and generate offspring by computing Wasserstein barycenters of the parent distributions. Conceptually, this amounts to interpolating between parent material distributions, yielding a simple yet distinctive crossover operator for EA-based topology optimization. Owing to the unique interpolative behavior induced by optimal transport, Wasserstein crossover can effectively recombine parent design features and generate diverse candidate solutions, which are particularly advantageous for EAs. We propose a novel EA-based topology optimization framework incorporating Wasserstein crossover into DDTD under the multifidelity design scheme [40], where initial designs derived by solving a low-fidelity problem are updated through iterative cycles of high-fidelity evaluation, selection, and Wasserstein crossover. This paper applies the proposed framework to several topology optimization problems, which are known to be intractable in structural mechanics and thermofluids, including a three-dimensional case, and demonstrates its effectiveness.

The rest of this paper is organized as follows. Section 2 provides background on the Wasserstein distance derived from optimal transport theory and its application to morphing, highlighting its potential as a crossover operator through comparison with other morphing approaches. Section 3 describes Wasserstein crossover and the proposed framework incorporating it into DDTD. Section 4 presents the detailed numerical implementation of the proposed framework, including the overall algorithm and Wasserstein crossover. Section 5 demonstrates the application of the proposed framework to several intractable topology optimization problems and discusses the results. Finally, Section 6 concludes the paper.

## 2. Preliminaries

In this section, we provide some prior knowledge on the Wasserstein distance and its application to graphical morphing. Subsequently, we compare it with other morphing methods, including a VAE-based one employed in DDTD, to highlight its potential as a crossover operator for topology optimization. We also review related works that have applied the idea of the Wasserstein distance to topology optimization and clarify the focus of this paper by highlighting the differences from those prior works.

### 2.1. Wasserstein distance

The fundamental idea of optimal transport is to measure the minimal cost required to transform one probability distribution  $\mu$  into another distribution  $\nu$  defined on  $\mathbb{R}^d$  (typically with  $d = 1, 2, 3$ ) [37], which is mathematically formulated as follows:

$$C(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}), \quad (1)$$

where  $c(\mathbf{x}, \mathbf{y})$  denotes the cost of transporting a unit mass from  $\mathbf{x} \in \mathbb{R}^d$  to  $\mathbf{y} \in \mathbb{R}^d$ , and  $\Pi(\mu, \nu)$  is the set of all transport plans with marginals  $\mu$  and  $\nu$ . The Wasserstein distance  $W_p(\mu, \nu)$  is defined as  $C(\mu, \nu)$  with a cost function based on the  $L^p$  norm  $\|\mathbf{x} - \mathbf{y}\|_p$ , where  $p \geq 1$ , to satisfy the axioms of a distance function, as follows:

$$W_p(\mu, \nu) = \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\|_p^p d\pi(\mathbf{x}, \mathbf{y}) \right)^{1/p}. \quad (2)$$

$W_p$  in Eq. (2) is the Wasserstein distance of order  $p$ , or simply the  $p$ -Wasserstein distance. In particular, the case of  $p = 1$  is also known as the earth mover's distance [38]. By explicitly incorporating the ground cost  $\|\mathbf{x} - \mathbf{y}\|_p^p$  into its definition, the Wasserstein distance reflects the geometric structure of the underlying space, providing an intuitive interpretation as the minimal effort required to morph one distribution into another. Unlike the simple measures such as the Kullback-Leibler divergence, which can become undefined or infinite when the distributions do not overlap, the Wasserstein distance remains well-defined and captures differences both in location and shape of the distributions. These properties make it a particularly expressive and versatile metric for comparing probability distributions.

While the Wasserstein distance is such a powerful metric, it is also known to be computationally expensive. To address this issue, a regularized formulation is often used by augmenting the objective function in Eq. (2) with an additional regularization term [41], as follows:

$$W_{p,\varepsilon}(\mu, \nu) = \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\|_p^p d\pi(\mathbf{x}, \mathbf{y}) - \varepsilon H(\pi) \right)^{1/p}, \quad (3)$$

where  $\varepsilon > 0$  is a regularization coefficient, and  $H(\pi)$  is the entropy of the transport plan  $\pi$ , defined as follows:

$$H(\pi) = - \int_{\mathbb{R}^d \times \mathbb{R}^d} (\log \pi(\mathbf{x}, \mathbf{y}) - 1) d\pi(\mathbf{x}, \mathbf{y}). \quad (4)$$



$W_{p,\varepsilon}$  in Eq. (3) is known as the entropy-regularized Wasserstein distance, and one of its advantages is that the regularization term  $H(\pi)$  renders the objective function strongly convex, which ensures smooth convergence to the optimal solution.

Another key advantage of the entropic regularization is that it enables much faster computation of the approximated optimal solution using the Sinkhorn algorithm [41,42]. To describe this, we consider the discretized case of the entropy-regularized Wasserstein distance, where the input probability distributions  $\mu$  and  $\nu$  are represented as discrete vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , and the distance between discrete points is given by a cost matrix  $\mathbf{C} \in \mathbb{R}^{n \times n}$ . The regularized optimal transport problem corresponding to the continuous case in Eq. (3) can then be formulated as follows:

$$\begin{aligned} & \underset{\mathbf{P} \in \mathbb{R}^{n \times n}}{\text{minimize}} && \sum_{i,j} P_{ij} C_{ij} + \varepsilon \sum_{i,j} P_{ij} (\log P_{ij} - 1) \\ & \text{subject to} && \mathbf{P}\mathbf{1} = \mathbf{a}, \\ & && \mathbf{P}^\top \mathbf{1} = \mathbf{b}, \\ & && P_{ij} \geq 0, \end{aligned} \tag{5}$$

where  $\mathbf{P}$  is the transport plan corresponding to  $\pi$  in the continuous formulation of Eq. (3), and  $\mathbf{1} \in \mathbb{R}^n$  is a vector of ones. For the optimization problem in Eq. (5), the pseudocode of the Sinkhorn algorithm, derived from its Lagrangian dual formulation, is presented in Algorithm 1. In this way, the Sinkhorn algorithm solves the entropy-regularized optimal transport problem by iteratively updating the scaling vectors  $\mathbf{u}$  and  $\mathbf{v}$ , with a computational complexity of  $\mathcal{O}(n^2)$  per iteration. In contrast, the original optimal transport problem formulated without the regularization term in Eq. (5) becomes a linear programming problem, whose computational complexity is  $\mathcal{O}(n^3)$ . Although the number of Sinkhorn iterations in Algorithm 1 depends on the strength of the regularization, in practice, convergence is achieved within tens or hundreds of iterations. Thus, the combination of the entropic regularization and Sinkhorn algorithm serves as an efficient approximation of the optimal transport solution with significantly reduced computational cost. Furthermore, as shown in Algorithm 1, the Sinkhorn algorithm essentially consists of simple matrix-vector computations, which makes it well-suited for GPU acceleration, as demonstrated by Ryu et al. [43]. In particular, when the cost function  $c(\mathbf{x}, \mathbf{y})$  is given by the squared Euclidean norm  $\|\mathbf{x} - \mathbf{y}\|_2^2$ , which corresponds to the case of the 2-Wasserstein distance  $W_{2,\varepsilon}$ , the matrix  $\mathbf{K}$  in Algorithm 1 becomes a Gaussian kernel matrix. As a result, the computations of  $\mathbf{K}\mathbf{v}$  and  $\mathbf{K}^\top \mathbf{u}$  are equivalent to convolutions with a Gaussian filter, which enables further acceleration through efficient convolution-based implementations. Thus, the Wasserstein distance, defined as a metric between probability distributions based on optimal transport, provides a theoretically grounded measure for which efficient computational methods have also been well established.

---

**Algorithm 1** Sinkhorn algorithm for computing the entropy-regularized optimal transport.

---

**Require:** Probability vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , cost matrix  $\mathbf{C} \in \mathbb{R}^{n \times n}$ , entropic regularization coefficient  $\varepsilon > 0$

- 1:  $\mathbf{K} \leftarrow \exp(-\mathbf{C}/\varepsilon)$  ▷ Element-wise exponential
  - 2:  $\mathbf{u} \leftarrow \mathbf{1}, \mathbf{v} \leftarrow \mathbf{1}$  ▷ Initialize with ones
  - 3: **repeat** ▷ Sinkhorn iterations
  - 4:    $\mathbf{u} \leftarrow \mathbf{a} \oslash (\mathbf{K}\mathbf{v})$  ▷  $\oslash$  denotes element-wise division
  - 5:    $\mathbf{v} \leftarrow \mathbf{b} \oslash (\mathbf{K}^\top \mathbf{u})$
  - 6:    $r_a \leftarrow \|\mathbf{u} \odot (\mathbf{K}\mathbf{v}) - \mathbf{a}\|_1$     $r_b \leftarrow \|\mathbf{v} \odot (\mathbf{K}^\top \mathbf{u}) - \mathbf{b}\|_1$  ▷ Marginal residuals,  $\odot$  denotes element-wise multiplication
  - 7: **until**  $\max(r_a, r_b) < \tau$  ▷ Convergence criterion with tolerance  $\tau$
  - 8:  $\mathbf{P} = \text{diag}(\mathbf{u})\mathbf{K}\text{diag}(\mathbf{v})$  ▷ Transport plan
  - 9: **return**  $\sum_{i,j} P_{ij} C_{ij}$  ▷ Wasserstein distance
- 

## 2.2. Wasserstein barycentric morphing

The concepts of the Wasserstein distance and optimal transport have been applied in various fields, including deep learning [44], image processing [38], and natural language processing [45]. Among these, one notable application is morphing proposed by Solomon et al. [36], where these techniques are used to interpolate between different shapes or distributions. It is based on the concept of the Wasserstein barycenter [46], which is defined as the weighted average of multiple probability distributions concerning the Wasserstein distance, given as follows:

$$\mu^* = \arg \min_{\mu} \sum_{i=1}^N \lambda_i W_p(\mu, \mu_i), \tag{6}$$

where  $\mu_i$  are the  $N$  input probability distributions,  $\lambda_i$  are the weights satisfying  $\sum_{i=1}^N \lambda_i = 1$ . The Wasserstein barycenter in Eq. (6) is formulated as a minimization problem for the multiple Wasserstein distances, which themselves require solving optimization problems in Eq. (2). This suggests that computing the Wasserstein barycenter would require solving a nested or two-level optimization problem. However, similar to the regularized Wasserstein distance in Eq. (3), entropic regularization can also be considered for the barycenter computation, allowing for an efficient approximation by a matrix-based iterative algorithm proposed by Benamou et al. [47], as outlined in Algorithm 2. Owing to the regularization, the approximated Wasserstein barycenter can be obtained without solving complex nested optimization problems, instead relying on simple iterative matrix operations. In addition, as described in

Section 2.1, these matrix operations can be further accelerated using GPU implementations, and in the case of the 2-Wasserstein distance, convolution with a Gaussian filter can also be used to speed up the computation of  $\mathbf{K}\mathbf{v}^{(i)}$  and  $\mathbf{K}^T\mathbf{u}^{(i)}$  in Algorithm 2. Note that the entropic regularization coefficient  $\varepsilon$  and the convergence tolerance  $\tau$  in Algorithm 2 affect the number of Sinkhorn iterations and should be chosen to balance computational cost and the quality of the approximated barycenter.

---

**Algorithm 2** Sinkhorn algorithm for computing the Wasserstein barycenter.

---

**Require:** Probability vectors  $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(N)} \in \mathbb{R}^n$ , weights  $\lambda_1, \dots, \lambda_N$ , cost matrix  $\mathbf{C} \in \mathbb{R}^{n \times n}$ , entropic regularization coefficient  $\varepsilon > 0$

```

1:  $\mathbf{K} \leftarrow \exp(-\mathbf{C}/\varepsilon)$  ▷ Element-wise exponential
2:  $\mathbf{u}^{(i)} \leftarrow \mathbf{1}, \mathbf{v}^{(i)} \leftarrow \mathbf{1} \quad (\forall i = 1, \dots, N)$  ▷ Initialize with ones
3: repeat ▷ Sinkhorn iterations
4:   for  $i = 1, \dots, N$  do
5:      $\mathbf{u}^{(i)} \leftarrow \mathbf{a}^{(i)} \oslash (\mathbf{K}\mathbf{v}^{(i)})$ 
6:      $\mathbf{v}^{(i)} \leftarrow \left( \prod_j (\mathbf{K}^T\mathbf{u}^{(j)})^{\lambda_j} \right) \oslash (\mathbf{K}^T\mathbf{u}^{(i)})$ 
7:   end for
8:    $E \leftarrow \sum_{e=1}^n \text{Std}(\{\mathbf{v}^{(i)} \odot (\mathbf{K}^T\mathbf{u}^{(i)})\}_{i=1}^N)_e$  ▷ Marginal deviations
9: until  $E < \tau$  ▷ Convergence tolerance  $\tau$ 
10: return  $\mathbf{a}^* = \prod_i (\mathbf{K}^T\mathbf{u}^{(i)})^{\lambda_i}$  ▷ Compute barycenter

```

---

Any shape data represented in a pixel or voxel grid can be treated as a probability distribution by normalizing its sum to one. By gradually varying the weight  $\lambda_i$  in Eq. (6) for each input distribution  $\mu_i$ , the Wasserstein barycenters can be computed to interpolate between the original shapes, resulting in smooth morphing based on optimal transport. Benamou et al. [47] demonstrated this morphing technique using three 2D shapes (diamond, ring, and square), and Solomon et al. [36] extended it to 3D shapes.

### 2.3. Comparison with other morphing methods

We compare the Wasserstein barycentric morphing described in Section 2.2 with two alternative morphing methods to examine their interpolation behaviors and verify their potential as a crossover operator for topology optimization. Fig. 1 shows the interpolated results obtained by morphing two different sample flow channels for the heat transfer problem in Section 5.2, together with their performance values. As comparison methods, we consider:

- (1) linear interpolation, the simplest approach, which corresponds to Euclidean barycentric interpolation;
- (2) VAE-based interpolation, obtained by linearly interpolating between two latent vectors.
- (3) Wasserstein barycentric interpolation, proposed in this paper.

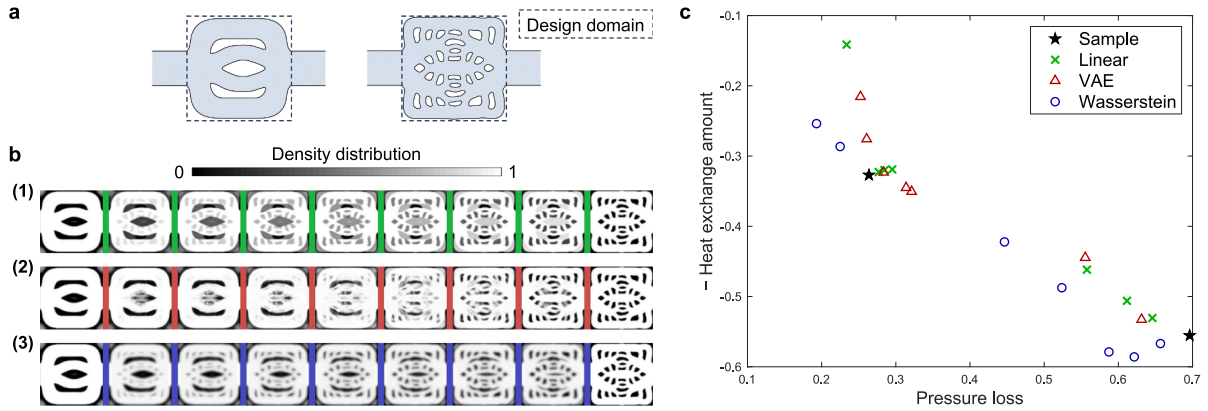
Comparing the three approaches in Fig. 1(b), linear interpolation results in a simple superposition of the two channels, whereas VAE-based interpolation exhibits a more complex and gradual transition through their intermediate shapes. This behavior stems from the nature of VAEs: the encoder and decoder networks are trained on a certain dataset, and interpolations in the constructed latent space implicitly incorporate the influence of the entire training dataset—100 different flow channels in this case—even when two specific designs are interpolated. As a result, the generated intermediate shapes reflect not only the two sample designs but also the characteristics of other flow channels in the training dataset. In contrast, the Wasserstein barycentric interpolation is derived from explicitly considering optimal transport between two sample designs. Compared to the grayscale-based transitions in the other two approaches, the Wasserstein barycentric morphing yields qualitatively different outcomes in that the fluid and solid regions (represented in white and black, respectively) remain relatively well distinguished throughout the transition. The resulting morphing involves a physically meaningful interpolation, in which the topology and branching patterns of flow channels are gradually transitioned.

Focusing on the performance values of the sample and interpolated channels shown in Fig. 1(c), we observe that the linear interpolated designs tend to exhibit mostly comparable performance values to those of the original samples. In contrast, the VAE-based interpolations are more scattered around the sample points, indicating slightly greater variation. Notably, however, the Wasserstein barycentric interpolations are more widespread in the objective space, including some designs that dominate the original samples in terms of Pareto optimality. These high-performance designs can be attributed to the distinctive interpolation behavior afforded by optimal transport, where the morphing in Fig. 1(b3) yields hybrid flow channels that effectively recombine the features of the samples.

These results demonstrate the potential of the Wasserstein barycentric morphing as a crossover operator, in that it enables the generation of offspring by shape interpolation between parents, some of which can have better performance compared to their parents. While the examples presented in Fig. 1 focus on two-dimensional flow channels for a heat transfer problem, we further investigate the performance of the proposed Wasserstein crossover through several numerical examples, including structural mechanics problems and three-dimensional cases, in Section 5.

### 2.4. Novelty in this paper compared to related works

Representative prior works that incorporate the concept of optimal transport and Wasserstein distance into topology optimization include the use of deep generative models based on the so-called Wasserstein GAN [44]. Wasserstein GANs are a variant of generative



**Fig. 1.** Comparison of different morphing methods between two sample data for a heat transfer problem: (a) two sample flow channels; (b) morphing results for design domain using (1) linear interpolation, (2) a variational autoencoder (VAE), and (3) Wasserstein barycentric interpolation; (c) objective values (pressure loss versus negative heat exchange amount) of sample and interpolated flow channels.

adversarial networks in which the earth mover's distance, i.e., the 1-Wasserstein distance, is used as the loss function. This formulation is known to improve learning stability and enhance the quality of the generated images. Recently, several topology optimization methods using Wasserstein GANs have been developed [48–52], but most of them aim to reduce computational cost by learning from a pre-topology-optimized dataset to predict optimized designs and eliminate iterations in topology optimization. This objective differs fundamentally from the aim of this paper, which focuses on developing a crossover operator within an EA-based topology optimization framework.

As an example of incorporating the idea of optimal transport into EAs, Sow et al. [53] proposed crossover and mutation operators based on Wasserstein barycenters. While their work focuses on the optimization of point clouds and considers wind-farm layout design as an engineering application, the design variables in their setting consist of the positions of only a few dozen points. This is fundamentally different from topology optimization, where the design variables typically involve material distributions over grids with thousands or more degrees of freedom.

Another example of applying Wasserstein barycenters to shape design is the framework proposed by Ma et al. [54] for the design of underwater swimming robots. They focused on the co-design of both the geometry and controller of swimmers, and successfully utilized Wasserstein barycentric interpolation to generate candidate designs of robots that are represented with tens of thousands of voxel resolutions. However, their approach remains centered on repeated simulation and interpolation between base designs given by the user, which is fundamentally different from the idea of EAs, where a population of candidate designs is evolved through genetic operations.

In this paper, we propose an EA-based topology optimization framework that introduces Wasserstein barycentric interpolation as an effective crossover operator for material distributions. The originality and novelty of this paper, compared with those related works, can be summarized as follows:

- Introduction of Wasserstein crossover as a dedicated and effective crossover operator for material distributions, based on the Wasserstein barycentric morphing.
- Development of an EA-based topology optimization framework capable of handling finely discretized continuum representations of material distributions with reasonable computational cost.

### 3. Framework

In this section, we describe the proposed framework incorporating Wasserstein crossover. Our proposed framework is based on DDTD [21,29], and focuses on solving the general multi-objective topology optimization problem, which is formulated as follows:

$$\begin{aligned}
 & \underset{\gamma(\mathbf{x})}{\text{minimize}} && [J_1(\gamma), J_2(\gamma) \dots, J_{n_o}(\gamma)], \\
 & \text{subject to} && G_j(\gamma) \leq 0 \quad (j = 1, 2, \dots, n_c), \\
 & && \gamma(\mathbf{x}) \in \{0, 1\}, \quad \forall \mathbf{x} \in D,
 \end{aligned} \tag{7}$$

where  $\gamma(\mathbf{x})$  is the binary material distribution defined on the design domain  $D \subset \mathbb{R}^d$  ( $d = 2$  or  $3$ ),  $J_i(\gamma)$  and  $G_j(\gamma)$  are the objective and constraint functions, respectively. To solve the optimization problem of Eq. (7) through the evolutionary process incorporating Wasserstein crossover, it is necessary to prepare a diverse set of initial candidate designs. Several approaches could be used for this purpose, such as parametrically generating designs based on predefined rules or heuristics [32], leveraging a database [21,55], or coarse optimization based on the multilevel mesh idea [56]. In this paper, we employ a multifidelity design approach [40] as a reasonable way to tackle the challenging topology optimization problems, although the proposed evolutionary process itself does not

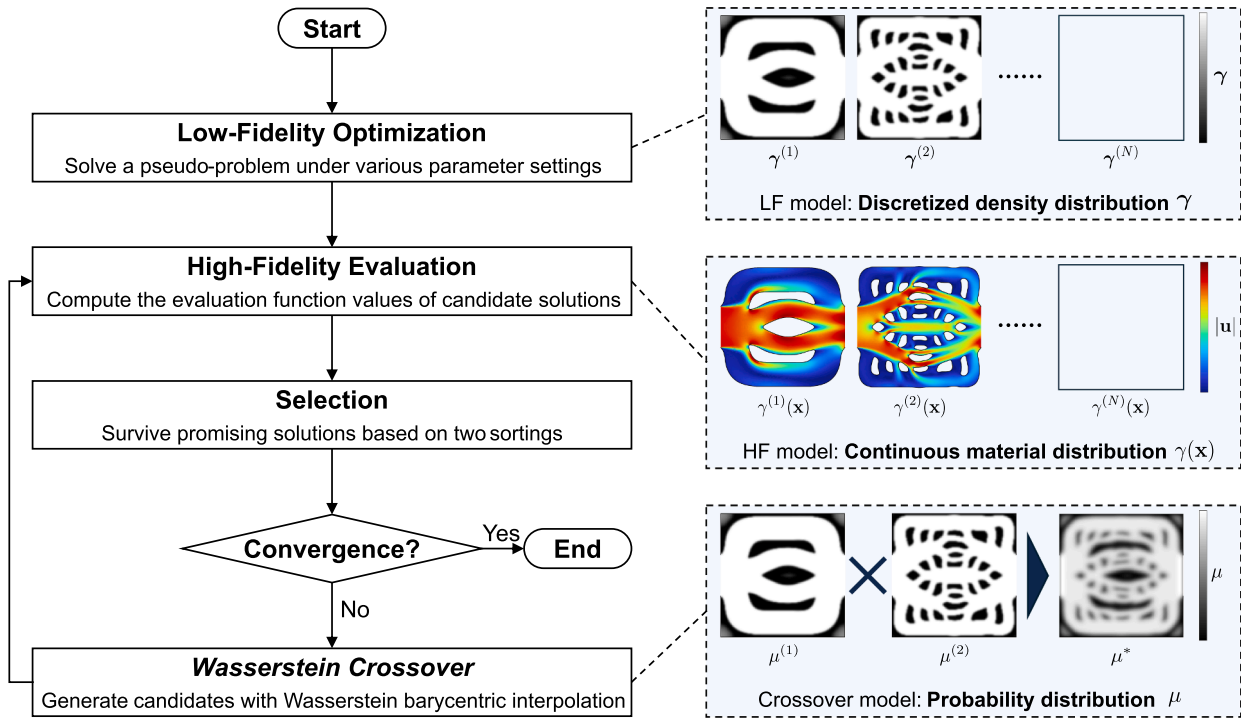


Fig. 2. Schematic flowchart of the proposed framework and representation of material distributions in each process.

depend on this particular choice. Specifically, we treat the original formulation of Eq. (7) as a high-fidelity (HF) model and construct a low-fidelity (LF) model, which is formulated as a simplified pseudo-problem as follows:

$$\begin{aligned}
 & \underset{\boldsymbol{\gamma}^{(k)}}{\text{minimize}} && \tilde{J}(\boldsymbol{\gamma}^{(k)}, \mathbf{s}^{(k)}), \\
 & \text{subject to} && \tilde{G}_j(\boldsymbol{\gamma}^{(k)}, \mathbf{s}^{(k)}) \leq 0 \quad (j = 1, 2, \dots, \tilde{n}_c), \\
 & && \boldsymbol{\gamma}^{(k)} \in [0, 1]^n, \\
 & \text{for given} && \mathbf{s}^{(k)},
 \end{aligned} \tag{8}$$

where  $\tilde{J}(\boldsymbol{\gamma}^{(k)}, \mathbf{s}^{(k)})$  and  $\tilde{G}_j(\boldsymbol{\gamma}^{(k)}, \mathbf{s}^{(k)})$  are the objective and constraint functions of the LF model, respectively. Solving the optimization problem of Eq. (8) is referred to as LF optimization, whose primary aim is to derive candidate solutions for the original problem of Eq. (7). The vector composed of artificial parameters  $\mathbf{s} = [s_1, s_2, \dots, s_{n_{sd}}]^\top$  is called a seeding parameter, and  $\mathbf{s}^{(k)}$  with  $k = 1, 2, \dots, N_{\text{lf}}$  represents the sample point of  $\mathbf{s}$ , which serves to induce diverse candidate designs. Taking into account the compatibility with Wasserstein crossover, which interpolates between candidate solutions represented as probability distributions, the LF optimization problem is assumed to be solved using the density-based method [2]. Thus, the objective and constraint functions,  $\tilde{J}(\boldsymbol{\gamma}^{(k)}, \mathbf{s}^{(k)})$  and  $\tilde{G}_j(\boldsymbol{\gamma}^{(k)}, \mathbf{s}^{(k)})$ , are assumed to be differentiable with respect to the design variable  $\boldsymbol{\gamma}^{(k)}$ , LF optimization problem of Eq. (8) should be easily solvable by gradient-based optimizers. Accordingly, the design variable  $\boldsymbol{\gamma}^{(k)}$  is defined as a continuous vector representing a material density distribution with values in the range  $[0, 1]$ , in contrast to the original binary material distribution  $\gamma(\mathbf{x})$  defined in the continuous domain. In addition, the LF optimization problem of Eq. (8) is converted to a single-objective problem from the original multi-objective problem of Eq. (7) using the weighted sum method [57],  $\epsilon$ -constraint method [58], and so on.

Based on the above formulations, the overall flowchart of the proposed framework is illustrated in Fig. 2. In this framework, Wasserstein crossover plays a central role in the iterative evolutionary process, as new candidate designs are generated exclusively through this operation at every generation. Note that the multifidelity formulations of Eq. (7) and Eq. (8) play a supplementary role in solving intractable topology optimization problems efficiently, and the roles of LF optimization and HF evaluation are clearly separated. LF optimization is used only to generate an initial population, and the subsequent evolutionary process exclusively consists of HF evaluation, selection, convergence check, and Wasserstein crossover. In the following, we describe the details of each procedure in the proposed framework, including LF optimization, HF evaluation, selection, convergence check, and Wasserstein crossover. It should also be noted that mutation, which is implemented with LF optimization in DDTD proposed by Yaji et al. [29], is not incorporated into the proposed framework because its impact in DDTD has been observed to be limited, while it also incurs additional computational cost. The development of effective mutation strategies is left as a subject for future work.

### Low-fidelity optimization

The first step of the proposed framework is to solve the LF optimization problem of Eq. (8). The primary aim of this step is to derive a diverse and promising set of density distributions, which serves as the initial population for the subsequent evolutionary process. The key component of its role is the seeding parameter  $s$  in Eq. (8). For example, by varying optimization algorithmic parameters such as filter radius or projection smoothness and threshold, the LF optimization problem can be solved multiple times to derive a diverse set of optimized density distributions. In addition, employing different weightings in the weighted sum method or varying constraint thresholds in the  $\varepsilon$ -constraint method enables the derivation of certain results that reflect different trade-offs between the objectives, thereby providing a promising set of candidate solutions.

While the LF optimization step relies on a density-based topology optimization approach, the associated parameter setting differs substantially from the trial-and-error parameter calibration often required when applying density-based methods directly to the original HF problem. Because the goal of this step is to generate a diverse set of initial candidate solutions rather than a single high-performance optimized design, uniformly sampled seeding parameters can be used without the need for fine manual tuning. This strategy avoids the extensive parameter studies typically associated with the density-based topology optimization for highly challenging problems. It should also be noted that parameters regarding LF optimization are not iteratively adjusted; once  $N_{lf}$  candidates are generated, subsequent HF evaluation and selection yield  $N_{pop}$  of them as the initial population, where  $N_{pop}$  is the population size.

### High-fidelity evaluation

In this step, each candidate solution is evaluated on the HF model to compute the objective and constraint values in the original optimization problem of Eq. (7). A crucial aspect of this step is that it only requires the numerical evaluation of those functions, without the need for sensitivity analysis concerning the design variables. This aspect allows the proposed framework to accommodate arbitrary evaluation functions, including non-differentiable and multimodal ones, thereby enabling the handling of minimax formulations and other complex optimization problems. Moreover, this property allows each candidate represented as a discretized density distribution  $\gamma$  to be converted into a continuous material distribution  $\gamma(\mathbf{x})$ , remeshed, and then evaluated on the HF model.

### Selection

In this step, promising candidate solutions are retained for the next generation, while the others are eliminated from the population based on their performance values obtained in the previous step. Selection plays a crucial role in EAs, as it helps maintain population diversity and facilitates effective global search. Since the proposed framework targets the multi-objective optimization problem defined in Eq. (7), ensuring sufficient diversity becomes even more challenging. To address this, we adopt the selection strategy in DDTD proposed by Kii et al. [59], which is designed to preserve the intrinsic diversity of material distributions. This procedure is based on the non-dominated sorting genetic algorithm II (NSGA-II) [60,61], which is one of the representative multi-objective GAs, and ranks candidates according to a two-stage sorting strategy. The first stage is the non-dominated sorting, which assigns ranks to candidates based on their Pareto dominance relations. The second stage, unique to their approach, further sorts candidates within the same rank using the Wasserstein distance sorting. This is achieved by employing a topological data analysis method called persistent homology [62,63] to extract topological features from the material distributions, which are then quantified using the Wasserstein distance [64]. While the detailed explanation, including the implementation of these procedures, is left to the original paper [59], a predefined number of higher-ranked candidates are selected based on these two types of sorting, and the population for the current generation is formed accordingly.

### Convergence check

In this step, the convergence of the optimization process is checked based on the hypervolume indicator [65], which is a representative performance metric for multi-objective optimization. The hypervolume indicator quantifies the area or volume formed by the solution set  $\mathcal{A}$  and a given reference point  $\mathbf{r} \in \mathbb{R}^{n_o}$ , defined as follows:

$$HV(\mathcal{A}, \mathbf{r}) = \mathcal{L} \left( \bigcup_{\mathbf{a} \in \mathcal{A}} \{\mathbf{b} \mid \mathbf{a} \leq \mathbf{b} \leq \mathbf{r}\} \right), \quad (9)$$

where  $\mathcal{L}(\cdot)$  denotes the Lebesgue measure. In this paper, we set the reference point  $\mathbf{r}$  as 1.1 times the worst objective values among all the candidates in the initial population, which is a common practice in the literature [65]. The optimization is terminated either when the increase in hypervolume is deemed sufficiently converged or when a predefined number of iterations is reached.

### Wasserstein crossover

This step constitutes the crossover operation proposed in this paper for integration into DDTD, where new offspring are generated from parent individuals using interpolation based on the Wasserstein barycenter. A key consideration here is how to interpret the discretized density distribution  $\gamma$  on the LF model in Eq. (8) as a probability distribution to compute the

Wasserstein barycenter in Eq. (6). To this end, the density vector  $\gamma$  is first normalized into a probability vector  $\mathbf{p} \in \mathbb{R}^n$  as follows:

$$\mathbf{p} = \frac{\gamma}{\sum_{e=1}^n \gamma_e}. \quad (10)$$

Then, the corresponding discrete probability measure  $\mu$  is defined by

$$\mu = \sum_{e=1}^n p_e \delta_{\mathbf{x}_e}, \quad (11)$$

where  $\delta_{\mathbf{x}_e}$  denotes the Dirac measure at the  $i$ -th grid point  $\mathbf{x}_e$ , and  $p_e$  represents the probability mass assigned to  $\mathbf{x}_e$ , i.e.,  $\mu(\{\mathbf{x}_e\}) = p_e$ . Based on this method of interpreting the discretized density distribution as a probability distribution, the proposed Wasserstein crossover takes the following steps:

1. Randomly select two parents  $\gamma^{(i)}, \gamma^{(j)} \in \Theta^{(t)}$ , where  $\Theta^{(t)}$  is the current population at iteration  $t$ .
2. Convert the parents into probability distributions  $\mu^{(i)}$  and  $\mu^{(j)}$  using Eqs. (10) and (11).
3. Compute the Wasserstein barycenter  $\mu^*$  of the two parental distributions  $\mu^{(i)}$  and  $\mu^{(j)}$  as follows:

$$\mu^* = \arg \min_{\mu} \{ \lambda W_{p,\varepsilon}(\mu, \mu^{(i)}) + (1 - \lambda) W_{p,\varepsilon}(\mu, \mu^{(j)}) \}, \quad (12)$$

where  $\lambda \in [0, 1]$  is a randomly sampled weight parameter.

4. Convert the barycenter  $\mu^* = \sum_{e=1}^n p_e^* \delta_{\mathbf{x}_e}$  back into a density distribution  $\gamma^*$  by applying min-max scaling to the probability vector  $\mathbf{p}^* = [p_1^*, p_2^*, \dots, p_n^*]^T$  as follows:

$$\gamma_e^* = \frac{p_e^* - \min_{e'} p_{e'}^*}{\max_{e'} p_{e'}^* - \min_{e'} p_{e'}^*}, \quad \text{for } e = 1, 2, \dots, n, \quad (13)$$

where  $p_e^* = \mu^*(\{\mathbf{x}_e\})$  denotes the probability mass at the  $e$ -th grid point in the barycenter  $\mu^*$ .

These sequential steps are repeated for a predefined number of offspring  $N_{\text{xo}}$ , each time adding the resulting density distribution  $\gamma^*$  as a new candidate solution to the population.

## 4. Numerical implementation

In this section, we describe the detailed implementation of the proposed framework. The overall procedure of the proposed framework, shown in Fig. 2, is summarized in Algorithm 3. Note that the LF optimization and HF evaluation steps in Algorithm 3 are independent of the others, allowing for acceleration through parallel implementation as computational resources permit. The entire algorithm was implemented using MATLAB (version 2023b). In the following, we describe the implementation details of each operation in Algorithm 3.

### 4.1. Low-fidelity optimization

The LF optimization problem of Eq. (8) is solved using the density-based method [2], which is a representative approach for topology optimization. To ensure the smoothness of the design variable  $\gamma^{(k)}$  in Eq. (8), we employ a density filter [66,67] defined as follows:

$$\tilde{\gamma}_e^{(k)} = \frac{\sum_{e' \in \mathcal{N}_e} w_{ee'} \gamma_{e'}^{(k)}}{\sum_{e' \in \mathcal{N}_e} w_{ee'}}, \quad \text{for } e = 1, 2, \dots, n, \quad (14)$$

where  $\mathcal{N}_e$  is the set of neighboring elements of the  $e$ -th element defined by:

$$\mathcal{N}_e = \{e' \mid \|\mathbf{x}_e - \mathbf{x}_{e'}\|_2 \leq R\}, \quad (15)$$

and  $w_{ee'}$  is the weight assigned to element  $e'$  with respect to element  $e$ , given by:

$$w_{ee'} = 1 - \frac{\|\mathbf{x}_e - \mathbf{x}_{e'}\|_2}{R}, \quad (16)$$

where  $R$  is the filter radius. We use the method of moving asymptotes (MMA) [68] for updating the design variables, and the move limit is set to 0.05 for all calculations of LF optimization in this paper. The entire LF optimization process was implemented in MATLAB, while the finite element analysis and sensitivity analysis, based on the discrete adjoint method, were performed using COMSOL Multiphysics (version 6.3).



**Algorithm 3** Overall procedure of the proposed framework.

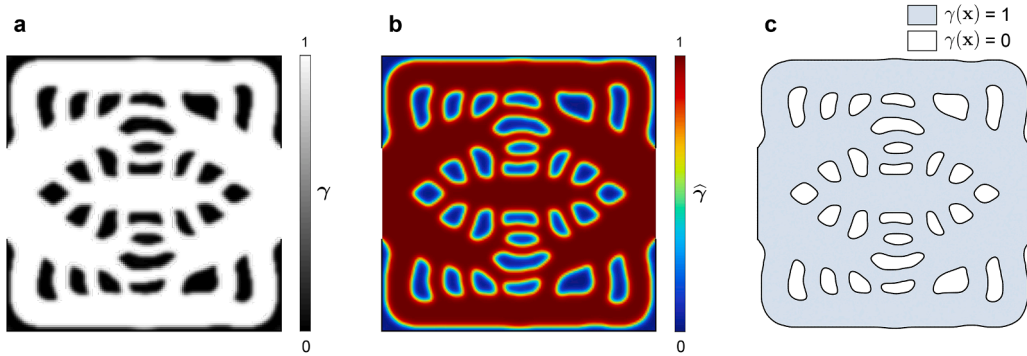
---

```

1: for  $k = 1$  to  $N_{lf}$  do
2:   Solve the LF optimization problem of Eq. (8) for  $\gamma^{(k)}$  on  $s^{(k)}$ 
3: end for
4: Assemble a temporary dataset of the LF optimized density distributions,  $\Theta_{tmp} \leftarrow \{\gamma^{(1)}, \dots, \gamma^{(N_{lf})}\}$ 
5: for  $t = 0$  to  $t_{max}$  do
6:   for  $k = 1$  to  $|\Theta_{tmp}|$  do
7:     Calculate HF evaluation function values in Eq. (7),  $\{J_1^{(k)}, \dots, J_{n_o}^{(k)}\}$  and  $\{G_1^{(k)}, \dots, G_{n_c}^{(k)}\}$ 
8:   end for
9: Remove candidates from  $\Theta_{tmp}$  that violate the constraints  $G_j^{(k)} \leq 0$  for all  $j = 1, \dots, n_c$ 
10: if  $t = 0$  then
11:   Set the current candidates,  $\Theta^{(t)} \leftarrow \Theta_{tmp}$ 
12: else
13:   Augment the current candidates,  $\Theta^{(t)} \leftarrow \Theta^{(t-1)} \cup \Theta_{tmp}$ 
14: end if
15: Update  $\Theta^{(t)}$  using the selection algorithm based on  $J(\Theta^{(t)}) \leftarrow \bigcup_{k=1}^{|\Theta^{(t)}|} \{J_1^{(k)}, \dots, J_{n_o}^{(k)}\}$ 
16: if HV in Eq. (9) regarding  $J(\Theta^{(t)})$  is converged then
17:   break
18: end if
19: Initialize a temporary dataset,  $\Theta_{tmp} \leftarrow \emptyset$ 
20: for  $k = 1$  to  $N_{xo}$  do
21:   Randomly select two parents  $\gamma^{(i)}, \gamma^{(j)} \in \Theta^{(t)}$ 
22:   Generate an offspring  $\gamma^{(k)}$  as the Wasserstein barycenter of  $\gamma^{(i)}$  and  $\gamma^{(j)}$ 
23:   Add the offspring to the temporary dataset,  $\Theta_{tmp} \leftarrow \Theta_{tmp} \cup \{\gamma^{(k)}\}$ 
24: end for
25: end for
26: return the optimized solutions  $\Theta^{(t)}$  and their performance values  $J(\Theta^{(t)})$ 

```

---



**Fig. 3.** Pre-process for high-fidelity evaluation: (a) the density distribution  $\gamma$  on the low-fidelity model; (b) the filtered density distribution  $\hat{\gamma}$ ; (c) the material distribution  $\gamma(x)$  on the high-fidelity model.

#### 4.2. High-fidelity evaluation

To compute the evaluation function values in the HF model of Eq. (7), it is necessary to convert the discretized density distribution  $\gamma$  on the LF model into a continuous material distribution  $\gamma(x)$  on the HF model as a pre-processing step. The pre-processing procedure adopted in this paper is illustrated in Fig. 3. First, we compute the filtered density distribution  $\hat{\gamma}$  using a partial differential equation (PDE)-based filter [69] defined as follows:

$$-R_h^2 \nabla^2 \hat{\gamma} + \hat{\gamma} = \gamma, \quad (17)$$

where  $R_h$  is the filter radius. Note that when solving the PDE in Eq. (17), Dirichlet boundary conditions are imposed on specific regions as necessary: for example, fixed and loaded boundaries in structural mechanics problems, or inlet and outlet boundaries in thermofluid problems, where  $\hat{\gamma}_e = 1$  or  $\hat{\gamma}_e = 0$  is prescribed accordingly. An isosurface at  $\hat{\gamma}_e = 0.5$  is then extracted from the filtered distribution field. This process yields a smooth boundary, from which a continuous material distribution  $\gamma(x)$  is obtained. The resulting material distributions are then discretized using appropriate meshes based on the problem settings, and the HF evaluation function values are subsequently evaluated. All of these HF evaluation processes were implemented in COMSOL Multiphysics.

### 4.3. Wasserstein crossover

In this paper, we use the 2-Wasserstein distance with entropic regularization in Eq. (12) for the barycentric computation, and replace certain matrix operations in Algorithm 2 with convolutions using a Gaussian filter to accelerate the computation as described in Section 2.2. In addition, the entropic regularization coefficient  $\varepsilon$ , which controls the trade-off between the accuracy of optimal transport and computational cost, is adaptively adjusted based on the pair of parent distributions selected for crossover. Specifically, prior to the crossover step described in Section 3, we compute a distance matrix  $\mathbf{D} \in \mathbb{R}^{N_{\text{pop}} \times N_{\text{pop}}}$ , where  $N_{\text{pop}}$  is the population size, defined as follows:

$$D_{ij} = \|\gamma^{(i)} - \gamma^{(j)}\|_2 = \sqrt{\sum_{e=1}^n \left( \gamma_e^{(i)} - \gamma_e^{(j)} \right)^2}. \quad (18)$$

Using the maximum and minimum value  $D_{\max}$ ,  $D_{\min}$  of the distance matrix  $\mathbf{D}$ , the entropic regularization coefficient  $\varepsilon$  for crossover between two parents  $\gamma^{(i)}$  and  $\gamma^{(j)}$  is adaptively determined as follows:

$$\varepsilon = \varepsilon_{\min} + (\varepsilon_{\max} - \varepsilon_{\min}) \frac{D_{ij} - D_{\min}}{D_{\max} - D_{\min}}, \quad (19)$$

where  $\varepsilon_{\max}$  and  $\varepsilon_{\min}$  are the maximum and minimum values of the entropic regularization coefficient, respectively. In this paper, we determine  $\varepsilon_{\max}$  and  $\varepsilon_{\min}$ , and the convergence tolerance  $\tau$  empirically based on preliminary numerical experiments for each problem setting, taking into account the trade-off between the degree of blurring introduced by regularization, the validity of the interpolated results, and the computational cost. By adaptively adjusting the entropic regularization coefficient in this manner, the Wasserstein barycenter can be computed more accurately when the parents are relatively similar within the population, while maintaining computational efficiency when they are dissimilar, which enhances the crossover performance. The Wasserstein crossover procedure was entirely implemented in Python (version 3.10.16). To accelerate the computation of the algorithm described in Algorithm 2, we employed JAX (version 0.6.1) [70] and implemented it based on POT (version 0.9.5) [71], a Python library for numerical computation of optimal transport.

## 5. Numerical examples

In this section, we demonstrate the effectiveness of the proposed framework for intractable topology optimization problems through three numerical examples in structural mechanics and thermofluids. The first example is a two-dimensional stress minimization problem that includes a non-differentiable objective function in the formulation. We verify the effectiveness of the framework for structural mechanics problems and highlight the advantages of the proposed Wasserstein crossover by comparing it with the VAE-based crossover employed in conventional DDTD. In addition, we investigate the usefulness of the proposed framework by comparing its optimized design with multiple designs optimized by the conventional density-based method under different parameter settings. The second example is a two-dimensional turbulent heat transfer problem, where evaluation functions exhibit strong multimodality due to the complex physics of turbulent flow. This example illustrates the applicability of the proposed framework beyond simple structural mechanics problems, extending it to complex thermofluid problems. We tackle a three-dimensional stress minimization problem as the final example, demonstrating the suitability for three-dimensional topology optimization problems. Finally, we discuss the computational cost for each example to evaluate the scalability of the proposed framework.

All numerical examples presented in this paper were computed on a Linux workstation with a 2.7 GHz AMD Ryzen Threadripper PRO 3995WX 64-core processor with 512 GB RAM and an NVIDIA RTX A6000 GPU with 48 GB memory. For the two-dimensional cases, the LF optimization and HF evaluation processes were executed in parallel using 50 CPU cores, whereas 30 CPU cores were used for the three-dimensional case. In all examples, the Wasserstein crossover process was computed on the GPU.

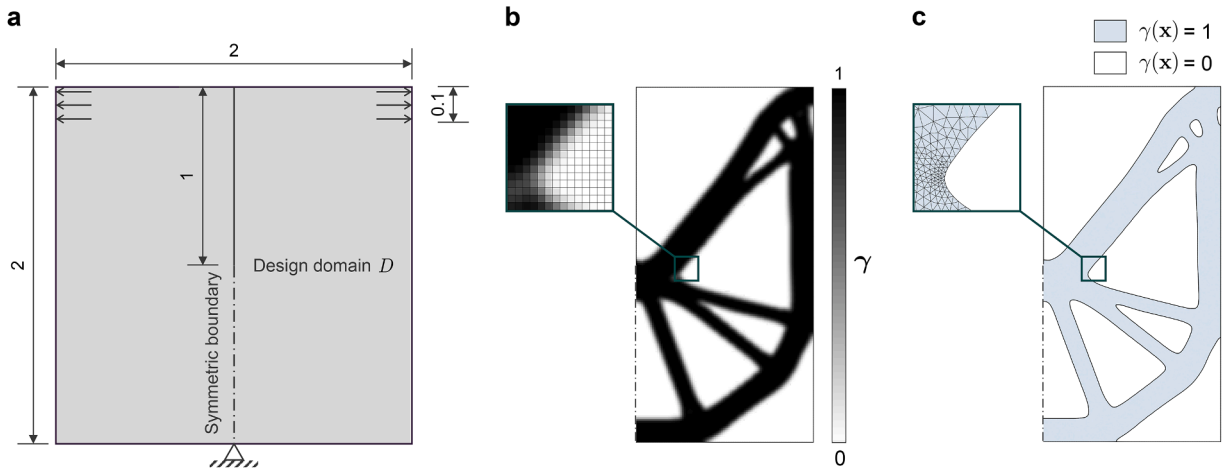
### 5.1. 2D Stress minimization problem

#### 5.1.1. Problem settings

Let us consider a multi-objective topology optimization problem of a two-dimensional cracked plate design, which is often used as a benchmark for stress-based topology optimization [73–77]. The optimization problem is formulated as follows:

$$\begin{aligned} \underset{\gamma(\mathbf{x})}{\text{minimize}} \quad & J_1 = \max_{\mathbf{x} \in D} \{\sigma(\mathbf{x})\}, \\ & J_2 = \frac{\int_D \gamma(\mathbf{x}) d\Omega}{\int_D d\Omega}, \\ \text{subject to} \quad & \gamma(\mathbf{x}) \in \{0, 1\}, \quad \forall \mathbf{x} \in D. \end{aligned} \quad (20)$$

Herein,  $\sigma(\mathbf{x})$  is the von Mises stress at the point  $\mathbf{x} \in D \subset \mathbb{R}^2$ . It is formulated as a two-objective optimization problem of minimizing the maximum stress  $J_1$  and volume fraction  $J_2$ . While the formulation of Eq. (20) is considered as the HF model corresponding to the



**Fig. 4.** Problem settings in the two-dimensional cracked plate design problem: (a) the design domain and boundary conditions; (b, c) a topology optimized configuration with (b) a density distribution  $\gamma$  discretized with structured meshes on the low-fidelity model, and (c) a material distribution  $\gamma(\mathbf{x})$  discretized with body-fitted meshes on the high-fidelity model.

formulation of Eq. (7), the LF optimization problem corresponding to Eq. (8) is formulated as follows:

$$\begin{aligned}
 & \underset{\gamma^{(k)}}{\text{minimize}} && \tilde{J} = \left( \sum_{e=1}^n \sigma_e^P \right)^{1/P}, \\
 & \text{subject to} && \tilde{G} = \sum_{e=1}^n v_e \gamma_e^{(k)} - V \sum_{e=1}^n v_e \leq 0, \\
 & && \gamma^{(k)} \in [0, 1]^n, \\
 & \text{for given} && \mathbf{s}^{(k)} = [s_1, s_2]^T.
 \end{aligned} \tag{21}$$

Herein,  $\sigma_e$  and  $v_e$  are the elemental von Mises stress and volume, respectively. Two seeding parameters,  $s_1$  and  $s_2$ , are used to determine the filter radius  $R$  in Eq. (14) and volume constraint  $V$ , respectively. Each parameter is sampled uniformly from the interval  $[0, 1]$  with  $n_{s_1}$  and  $n_{s_2}$  divisions. The filter radius  $R$  is obtained by linearly interpolating between  $R_{\min}$  and  $R_{\max}$  according to  $s_1$ , while the volume constraint  $V$  is similarly interpolated between  $V_{\min}$  and  $V_{\max}$  using  $s_2$ , resulting in a total of  $N_{\text{lf}} = n_{s_1} \times n_{s_2}$  parameter combinations. The objective function  $\tilde{J}$  in the LF formulation of Eq. (21) is defined as the  $P$ -norm stress, which is a differentiable approximation of the maximum stress. For the norm parameter  $P$ , we set a constant value of  $P = 8$ , which is often used from the perspective of the trade-off between accuracy and numerical stability [4]. The LF optimization problem of Eq. (21) is solved  $N_{\text{lf}}$  times independently using the solid isotropic material with penalty (SIMP) method [78].

The design domain and boundary conditions of the cracked plate are illustrated in Fig. 4(a). Due to the symmetry, we define the right half of a  $2 \times 2$  plate as the design domain  $D$ , where a symmetric boundary condition is imposed only on the lower half of the central boundary to represent a crack in the upper half. In the LF model shown in Fig. 4(b), the design variable field  $\gamma^{(k)}$  is discretized using structured meshes with  $200 \times 100$  quadrilateral elements, i.e.,  $n = 20000$ . In contrast, body-fitted meshes are employed in the HF model shown in Fig. 4(c), where the boundary between material and void regions is defined by an isosurface extracted through the pre-processing procedure described in Section 4.2. The maximum and minimum element sizes for the HF model are set to 0.04 and  $1.5 \times 10^{-4}$ , respectively, and body-fitted meshes with quadrilateral elements are automatically built for each candidate solution. In the LF model, the design variable  $\gamma^{(k)}$  is discretized using  $\mathbb{P}^0$  Lagrange finite elements, whereas the HF model employs  $\mathbb{P}^2$  Lagrange finite elements for structural analysis.

The parameters regarding overall procedures of the proposed framework are listed in Table 1. Note that the convergence check with the hypervolume indicator is not performed in this example, and the optimization is terminated after a predefined number of iterations  $t_{\max}$ .

Using this numerical example, we compare the VAE-based crossover employed in conventional DDTD and the proposed Wasserstein crossover. The settings for the VAE in this comparison are listed in Table 2. We determined these settings following the previous studies [25,26,59] where DDTD was applied to stress-based topology optimization problems similar to this example.

### 5.1.2. Results and discussion

Fig. 5 shows the initial population obtained from the LF optimization problem of Eq. (21). Owing to the effect of the seeding parameters  $s_1$  and  $s_2$  in Eq. (21), density distributions on the LF model shown in Fig. 5(a) exhibit a variety of topologies under different volume fractions. These density distributions are then converted into binarized material distributions on the HF model, as shown in Fig. 5(b). Some disconnected structures are observed in Fig. 5(b), which results from the fact that the Heaviside projection

**Table 1**  
Overall parameters in the two-dimensional stress minimization problem.

| Description                                      | Symbol            | Value                |
|--------------------------------------------------|-------------------|----------------------|
| Maximum number of iterations                     | $t_{\max}$        | 100                  |
| Population size                                  | $N_{\text{pop}}$  | 100                  |
| Number of offspring                              | $N_{\text{xo}}$   | 100                  |
| Number of LF optimized initial designs           | $N_{\text{lf}}$   | 100                  |
| Number of seeding parameters for filter radius   | $n_{s_1}$         | 4                    |
| Number of seeding parameters for volume fraction | $n_{s_2}$         | 25                   |
| Minimum filter radius                            | $R_{\min}$        | 0.03                 |
| Maximum filter radius                            | $R_{\max}$        | 0.12                 |
| Minimum volume fraction                          | $V_{\min}$        | 0.30                 |
| Maximum volume fraction                          | $V_{\max}$        | 0.60                 |
| Filter radius for HF evaluation                  | $R_h$             | 0.01                 |
| Minimum entropic regularization coefficient      | $\epsilon_{\min}$ | $1.0 \times 10^{-6}$ |
| Maximum entropic regularization coefficient      | $\epsilon_{\max}$ | $1.0 \times 10^{-4}$ |
| Convergence tolerance in Sinkhorn algorithm      | $\tau$            | $1.0 \times 10^{-9}$ |

**Table 2**  
Parameters for the VAE-based crossover in the two-dimensional stress minimization problem.

| Description                                                | Value                                         |
|------------------------------------------------------------|-----------------------------------------------|
| Size of input and output data                              | 20,000                                        |
| Size of latent space                                       | 8                                             |
| Architecture of encoder                                    | [20,000, 512, 8] (fully connected)            |
| Architecture of decoder                                    | [8, 512, 20,000] (fully connected)            |
| Activation function                                        | Sigmoid (output layer)<br>ReLU (other layers) |
| Reconstruction loss function                               | Mean squared error                            |
| Weighting coefficient for Kullback-Leibler (KL) divergence | 0.01                                          |
| Optimizer                                                  | Adam [72]                                     |
| Learning rate                                              | 0.001                                         |
| Batch size                                                 | 10                                            |
| Number of epochs                                           | 500                                           |

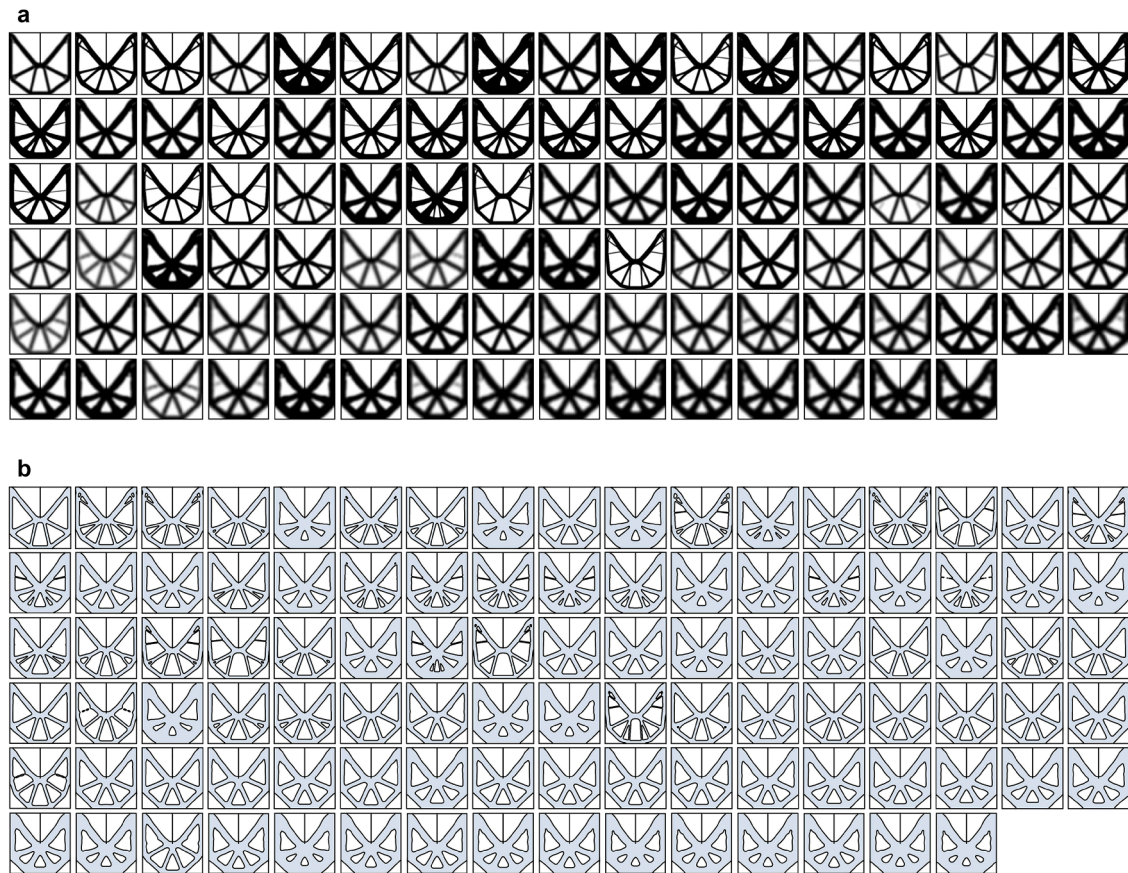
technique [79], commonly used to promote binarization in the density-based method, was deliberately not applied to prioritize stability in the LF optimization.

Fig. 6 shows the optimization results comparing the VAE-based crossover and the proposed Wasserstein crossover. The initial designs shown in Fig. 5 are progressively evolved through iterations using the VAE-based and Wasserstein crossovers. As shown in Fig. 6(a), Wasserstein crossover yields a significantly greater performance improvement. Specifically, the VAE-based crossover achieves an approximate 31% improvement from the initial value, whereas the Wasserstein crossover achieves an about 80% improvement. Fig. 6(b) further illustrates that the Pareto front obtained via the Wasserstein crossover exhibits a substantial advancement from the initial solutions and completely dominates the Pareto front via the VAE-based crossover.

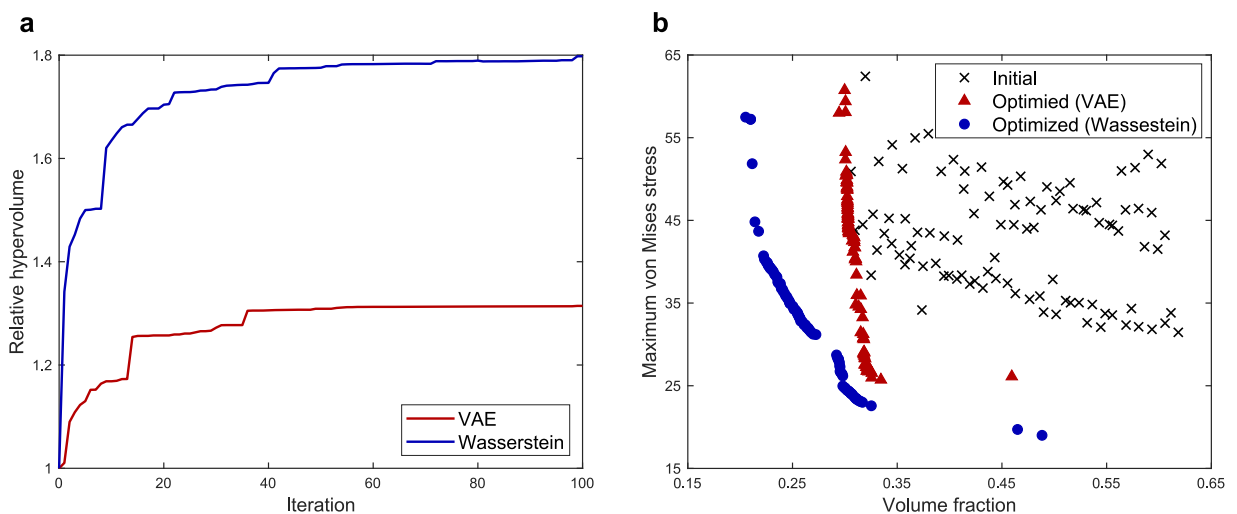
Fig. 7 compares the optimized designs obtained via the VAE-based and Wasserstein crossovers. A common feature observed in both cases is that the population has converged to material distributions with almost the same topology, which also exists in the initial population in Fig. 5(b). As can be seen in Fig. 7(a), the optimized designs via Wasserstein crossover achieve a more significant reduction in volume.

To enable a more detailed comparison, Fig. 8 presents the comparison results of the initial, optimized via VAE-based crossover, and optimized via Wasserstein crossover designs under almost identical volume fraction conditions. While these designs indeed exhibit the same topology, a comparison of the maximum stress value  $J_1$  reveals that the optimized design via VAE-based crossover achieves an approximate 28% reduction from the initial design, whereas the optimized design via Wasserstein crossover further improves upon this, achieving a 40% reduction. This improvement can be attributed to subtle geometric modifications, such as changes in the outer shape and the rounding of holes, which lead to more uniform stress distributions. Notably, the optimized design via Wasserstein crossover particularly exhibits a more uniform stress distribution around the tip of the crack, which is typically prone to a high stress concentration. Fig. 9 next presents a comparison under approximately identical conditions in terms of maximum stress  $J_1$ . Although all three designs exhibit similar outer shapes, it can be seen that the thin structure, which is barely continuous in the initial design, is removed in both the optimized designs. Focusing on the crack tip in Fig. 9(b), stress concentration is evident in the initial design and optimized design via VAE-based crossover. In contrast, the optimized design via Wasserstein crossover effectively redistributes the stress throughout the surrounding thin structures, resulting in a reduction in the volume fraction  $J_2$  by approximately 30% compared to the initial design.

In this numerical example, the LF optimization employs a conventional approach for stress-based topology optimization using a gradient-based method. Thus, comparing the initial designs with those optimized by the proposed method implicitly highlights its advantages over the conventional approach. However, it is well known that, in the formulation using the  $P$ -norm stress in Eq. (21),



**Fig. 5.** Low-fidelity optimized initial designs in the two-dimensional cracked plate design problem: (a) density distributions on the low-fidelity model; (b) material distributions on the high-fidelity model.



**Fig. 6.** Optimization results in the two-dimensional cracked plate design problem: (a) convergence history of the hypervolume indicator normalized by the initial value; (b) objective space where the performance values of the initial and optimized designs by VAE-based and Wasserstein crossover are plotted.





Fig. 7. Optimized designs in the two-dimensional cracked plate design problem: (a) optimized via VAE-based crossover; (b) optimized via Wasserstein crossover.

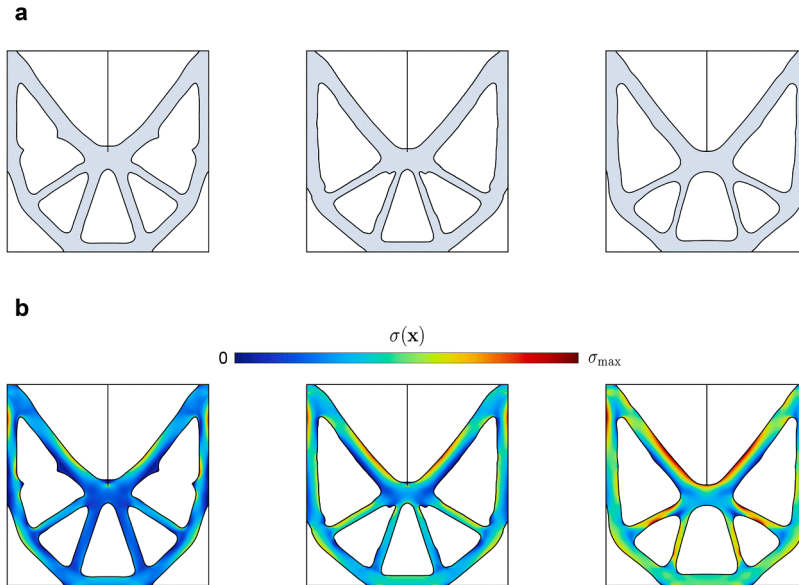
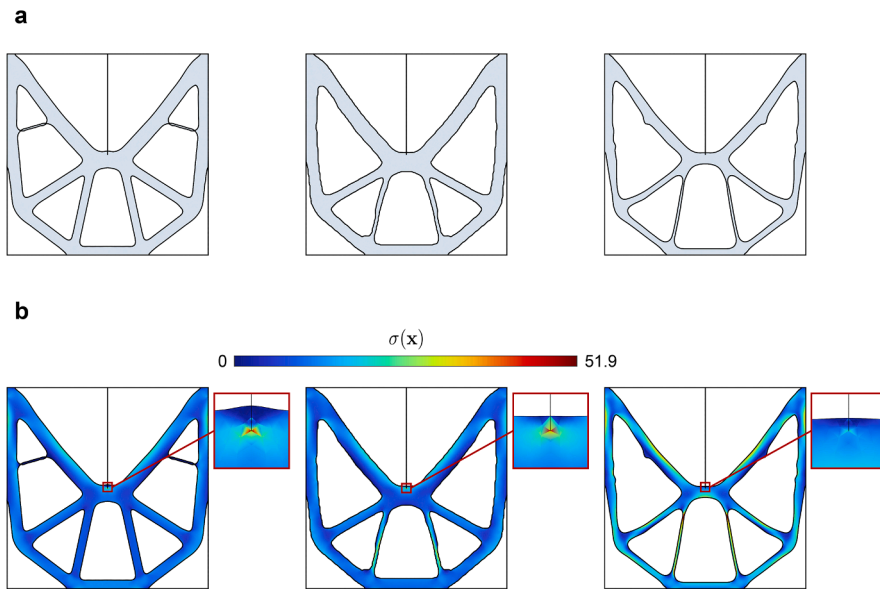


Fig. 8. Results comparison of initial (left), optimized via VAE-based crossover (middle), and optimized via Wasserstein crossover (right) designs under nearly identical volume fraction conditions in the two-dimensional cracked plate design problem: (a) material distributions; (b) stress distributions  $\sigma(\mathbf{x})$ . Herein, these objective values are:  $J_1 = 38.4$ ,  $J_2 = 0.325$  (left);  $J_1 = 27.5$ ,  $J_2 = 0.320$  (middle);  $J_1 = 23.0$ ,  $J_2 = 0.317$  (right), where  $J_1$  and  $J_2$  are the maximum stress and volume fraction, respectively. Note that the color bar is normalized with respect to the maximum stress  $\sigma_{\max}$  of each design. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)





**Fig. 9.** Results comparison of initial (left), optimized via VAE-based crossover (middle), and optimized via Wasserstein crossover (right) designs under nearly identical maximum stress conditions in the two-dimensional cracked plate design problem: (a) material distributions; (b) stress distributions  $\sigma(x)$ . Herein, these objective values are:  $J_1 = 50.9$ ,  $J_2 = 0.306$  (left);  $J_1 = 50.9$ ,  $J_2 = 0.301$  (middle);  $J_1 = 51.9$ ,  $J_2 = 0.212$  (right), where  $J_1$  and  $J_2$  are the maximum stress and volume fraction, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

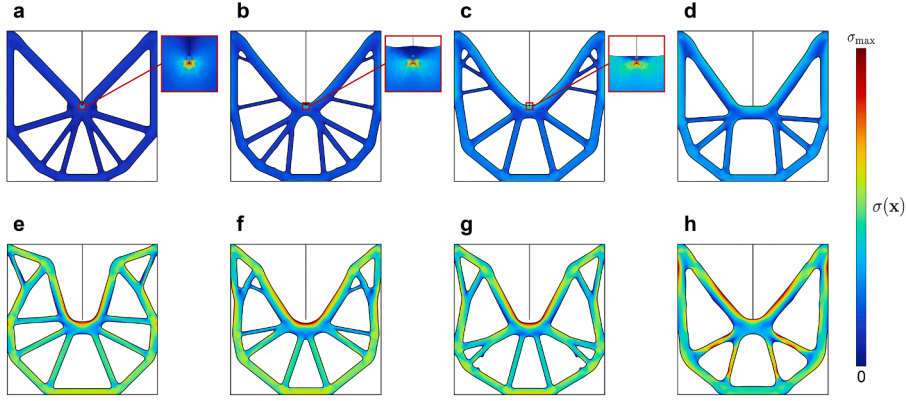
**Table 3**

Comparison of the objective values optimized using the density-based method and the proposed framework with Wasserstein crossover.

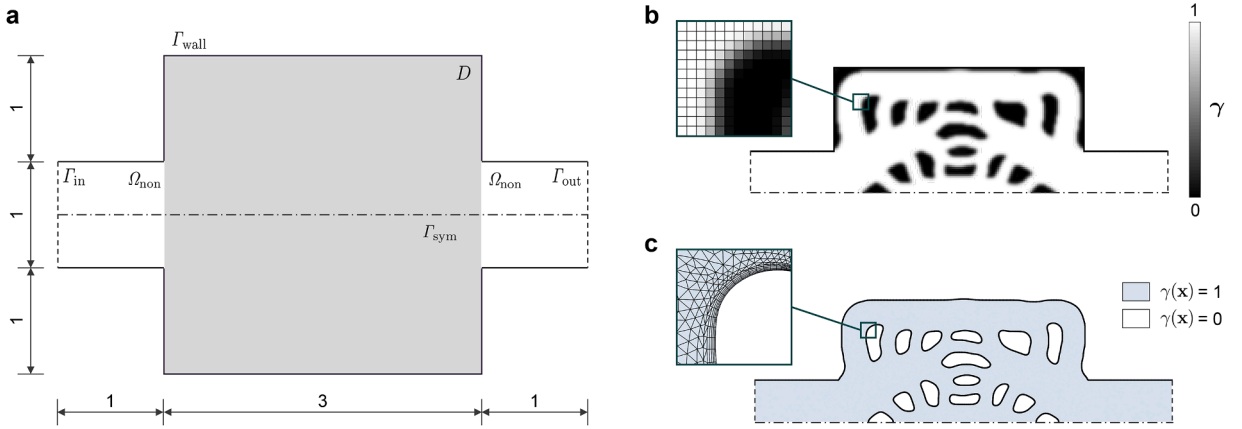
| Index in Fig. 10 | Optimization method  | Norm parameter $P$ | Maximum stress $J_1$ | Volume fraction $J_2$ |
|------------------|----------------------|--------------------|----------------------|-----------------------|
| (a)              | Density-based method | 4                  | 91.4                 | 0.304                 |
| (b)              |                      | 6                  | 65.8                 | 0.303                 |
| (c)              |                      | 8                  | 46.9                 | 0.306                 |
| (d)              |                      | 10                 | 44.2                 | 0.305                 |
| (e)              |                      | 12                 | 26.3                 | 0.304                 |
| (f)              |                      | 14                 | 25.1                 | 0.308                 |
| (g)              |                      | 16                 | 25.9                 | 0.306                 |
| (h)              | Proposed framework   | –                  | 24.4                 | 0.303                 |

both the resulting design and its structural performance are highly sensitive to the choice of the norm parameter  $P$ . Therefore, to further validate the effectiveness of the proposed framework, we solved the optimization problem in Eq. (21) with a fixed volume fraction constraint of  $V = 0.30$  using the density-based method for several different values of  $P$  ranging from 4 to 16, and the resulting designs were then binarized and evaluated on the HF model following the same procedure described in Section 4.2. Fig. 10 and Table 3 compare these results with the optimized design by the proposed framework at approximately the same volume fraction. Due to the nature of the norm, the  $P$ -norm stress tends to well approximate the maximum stress as  $P$  increases, resulting in lower maximum stress values  $J_1$  in Table 3 and more distributed stress distributions around the crack tip in Figs. 10(e)–(g). The deterioration observed at  $P = 16$  is likely due to the optimization being trapped in a poor local optimum. Nevertheless, as seen in Fig. 10(h), the optimized design by the proposed framework effectively redistributes the stress not only around the crack tip but also throughout other structural members, leading to the lowest maximum stress  $J_1$  among all designs despite its smallest volume fraction  $J_2$ . This result demonstrates the advantage of the proposed framework, in which the maximum stress evaluated on the HF model using the body-fitted meshes is directly minimized through an EA, rather than relying on approximations such as the  $P$ -norm stress. It should be noted that although we employed  $P = 8$  in this example for the LF optimization with the primary aim of stably generating diverse initial designs in the proposed framework, using a larger value or varying it through an additional seeding parameter could potentially yield even better optimized results.

These results suggest that the proposed framework can be effectively applied to stress-based topology optimization problems, particularly in handling non-differentiable objective functions such as the maximum stress. Furthermore, the proposed Wasserstein crossover appears capable of generating physically reasonable material distributions not only in fluid problems, as shown in Fig. 1, but also in structural mechanics problems. Its performance as a crossover operator seems to be superior to that of the VAE-based crossover, which is used in conventional DDTD.



**Fig. 10.** Comparison of stress distributions optimized using the density-based method (a–g) and the proposed framework with Wasserstein crossover (h) in the two-dimensional cracked plate design problem. In the results of the density-based method, each design was optimized under a different parameter setting in the  $P$ -norm stress function: (a)  $P = 4$ ; (b)  $P = 6$ ; (c)  $P = 8$ ; (d)  $P = 10$ ; (e)  $P = 12$ ; (f)  $P = 14$ ; (g)  $P = 16$ . The corresponding objective values are listed in Table 3. Note that the color bar is normalized with respect to the maximum stress  $\sigma_{\max}$  of each design. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 11.** Problem settings in the two-dimensional heat sink design problem: (a) the analysis domain composed of the design domain  $D$  and non-design domain  $\Omega_{\text{non}}$ ; (b, c) a topology optimized configuration with (b) a density distribution  $\gamma$  discretized with structured meshes on the low-fidelity model, and (c) a material distribution  $\gamma(\mathbf{x})$  discretized with body-fitted meshes and boundary layer meshes on the high-fidelity model. The fluid comes from the inlet  $\Gamma_{\text{in}}$  to the outlet  $\Gamma_{\text{out}}$ . The boundary  $\Gamma_{\text{sym}}$  is defined as a symmetric boundary. The remaining external boundary  $\Gamma_{\text{wall}}$  is defined as a wall boundary.

## 5.2. 2D Turbulent heat transfer problem

Let us consider a heat sink design problem considering two-dimensional turbulent heat transfer. Fig. 11(a) illustrates the design domain  $D \subset \mathbb{R}^2$  and an analysis domain  $\mathcal{O} = D \cup \Omega_{\text{non}}$  with  $D \cap \Omega_{\text{non}} \neq \emptyset$ , where  $\Omega_{\text{non}}$  is a non-design domain. Examples of the optimized configuration on the LF and HF models are shown in Figs. 11(b) and (c), respectively.

### 5.2.1. Governing equations

First, we provide the governing equations and boundary conditions for both the HF and LF models.

#### High-fidelity model

The analysis domain  $\mathcal{O}$  on the HF model is composed of the fluid domain  $\Omega_f$  and the solid domain  $\Omega_s = \mathcal{O} \setminus \Omega_f$ . Within the design domain  $D$ , the fluid domain  $\Omega_f$  and solid domain  $\Omega_s$  are indicated by  $\gamma(\mathbf{x}) = 1$  and  $\gamma(\mathbf{x}) = 0$ , respectively, where  $\gamma(\mathbf{x})$  is the design variable field of a material distribution. The evaluation functions are given by the governing equations for fluid velocity  $\mathbf{u} : \Omega_f \rightarrow \mathbb{R}^2$ , pressure  $p : \Omega_f \rightarrow \mathbb{R}$ , and temperature  $T : \Omega_f \rightarrow \mathbb{R}$ .

In this paper, we model the turbulent flow using the steady-state incompressible Reynolds-averaged Navier-Stokes (RANS) equations with the standard  $k - \epsilon$  model, which is a widely used approximate model for turbulent flow analysis. The governing equations for the fluid velocity  $\mathbf{u}$  and pressure  $p$  in  $\Omega_f$  are given by the following dimensionless RANS equations:

$$\nabla \cdot \mathbf{u} = 0, \quad (22)$$

$$(\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p + \nabla \cdot \left( \frac{2}{Re} \mathbf{S} \right) + \nabla \cdot \mathbf{T}, \quad (23)$$

where  $Re$  is the Reynolds number.  $\mathbf{S} = (\nabla \mathbf{u} + (\nabla \mathbf{u})^\top)/2$  is the mean strain-rate tensor and  $\mathbf{T} = 2\nu_t \mathbf{S} - 2k\mathbf{I}/3$  is the Reynolds stress tensor based on the Boussinesq approximation. Here,  $k$  is the turbulent kinetic energy and  $\nu_t$  is the eddy viscosity, which is defined as follows:

$$\nu_t = C_\mu \frac{k^2}{\epsilon}, \quad (24)$$

where  $C_\mu$  is a empirical constant, and  $\epsilon$  is the energy dissipation rate.  $k$  and  $\epsilon$  are governed by the following equations:

$$\mathbf{u} \cdot \nabla k = \nabla \cdot \left( \frac{\nu_t}{\sigma_k} \nabla k \right) + 2\nu_t \mathbf{S} : \mathbf{S} - \epsilon, \quad (25)$$

$$\mathbf{u} \cdot \nabla \epsilon = \nabla \cdot \left( \frac{\nu_t}{\sigma_\epsilon} \nabla \epsilon \right) + C_{1\epsilon} \frac{\epsilon}{k} 2\nu_t \mathbf{S} : \mathbf{S} - C_{2\epsilon} \frac{\epsilon^2}{k}, \quad (26)$$

where the empirical constants are set as  $C_\mu = 0.09$ ,  $C_{1\epsilon} = 1.44$ ,  $C_{2\epsilon} = 1.92$ ,  $\sigma_k = 1.0$ , and  $\sigma_\epsilon = 1.3$ . The temperature  $T : \Omega_f \rightarrow \mathbb{R}$  is governed by the following conjugate heat transfer equation in a dimensionless form:

$$\mathbf{u} \cdot \nabla T = \nabla \cdot \left( \left( \frac{1}{Pe} + \frac{\nu_t}{Pr_t} \right) \nabla T \right), \quad (27)$$

where  $Pe$  is the Péclet number defined as  $Re \cdot Pr$  where  $Pr$  is the Prandtl number, and  $Pr_t$  is the turbulent Prandtl number. The parameters regarding these governing equations for the HF model are listed in Table 4.

The boundary conditions in the analysis domain shown in Fig. 11(a) are defined in dimensionless forms as follows:

$$\mathbf{u} = -\mathbf{n}, \quad T = 0 \quad \text{on } \Gamma_{\text{in}}, \quad (28)$$

$$p = 0, \quad \mathbf{n} \cdot \nabla T = 0 \quad \text{on } \Gamma_{\text{out}}, \quad (29)$$

$$\mathbf{u} \cdot \mathbf{n} = 0, \quad \mathbf{n} \cdot \nabla T = 0 \quad \text{on } \Gamma_{\text{wall}} \cup \Gamma_{\text{sym}}, \quad (30)$$

$$-\mathbf{n} \cdot \mathbf{q} = Bi(1 - T) \quad \text{on } \partial\Omega_f \setminus (\Gamma_{\text{wall}} \cup \Gamma_{\text{sym}}), \quad (31)$$

where  $\mathbf{n}$  is the outward unit normal vector and  $Bi$  is the Biot number. The heat flux  $\mathbf{q}$  in a dimensionless form is defined as follows:

$$\mathbf{q} = -\left( \frac{1}{Pe} + \frac{\nu_t}{Pr_t} \right) \nabla T. \quad (32)$$

**Table 4**

Parameters regarding the high-fidelity model in the two-dimensional turbulent heat transfer problem.

| Description                     | Symbol | Value |
|---------------------------------|--------|-------|
| Reynolds number                 | $Re$   | 5000  |
| Péclet number                   | $Pe$   | 500   |
| Turbulent Prandtl number        | $Pr_t$ | 0.1   |
| Biot number                     | $Bi$   | 0.1   |
| Filter radius for HF evaluation | $R_h$  | 0.021 |

**Table 5**

Parameters regarding the low-fidelity model in the two-dimensional turbulent heat transfer problem.

| Description                                                           | Symbol         | Value             |
|-----------------------------------------------------------------------|----------------|-------------------|
| Inverse permeability                                                  | $\alpha$       | $1.0 \times 10^4$ |
| Prandtl number of fluid                                               | $Pr_f$         | 0.1               |
| Prandtl number of solid                                               | $Pr_s$         | 1                 |
| Tuning parameter for inverse permeability                             | $q_a$          | 100               |
| Tuning parameter for Prandtl number                                   | $q_{Pr}$       | 10                |
| Minimum Reynolds number                                               | $Re_{\min}$    | 10                |
| Maximum Reynolds number                                               | $Re_{\max}$    | 80                |
| Minimum volumetric heat transfer coefficient                          | $\beta_{\min}$ | 0.1               |
| Maximum volumetric heat transfer coefficient                          | $\beta_{\max}$ | 0.5               |
| Number of seeding parameters for Reynolds number                      | $n_{s_1}$      | 4                 |
| Number of seeding parameters for volumetric heat transfer coefficient | $n_{s_2}$      | 25                |
| Filter radius                                                         | $R$            | 0.06              |

**Table 6**

Parameters regarding the overall algorithm in the two-dimensional turbulent heat transfer problem.

| Description                                 | Symbol            | Value                |
|---------------------------------------------|-------------------|----------------------|
| Maximum number of iterations                | $t_{\max}$        | 100                  |
| Population size                             | $N_{\text{pop}}$  | 100                  |
| Number of offspring                         | $N_{\text{so}}$   | 100                  |
| Number of LF optimized initial designs      | $N_{\text{lf}}$   | 100                  |
| Minimum entropic regularization coefficient | $\epsilon_{\min}$ | $2.0 \times 10^{-6}$ |
| Maximum entropic regularization coefficient | $\epsilon_{\max}$ | $2.0 \times 10^{-4}$ |
| Convergence tolerance in Sinkhorn algorithm | $\tau$            | $1.0 \times 10^{-9}$ |

### Low-fidelity model

In order to stably solve the LF optimization problem and derive diverse and promising initial designs, we employ a laminar flow model under low Reynolds number conditions as the LF model, which is a common approach in the literature [5]. To enable density-based topology optimization by representing both fluid and solid domains using a single design variable, we consider the following governing equations for the state variable fields  $\tilde{\mathbf{u}} : \mathcal{O} \rightarrow \mathbb{R}^2$ ,  $\tilde{p} : \mathcal{O} \rightarrow \mathbb{R}$ , and  $\tilde{T} : \mathcal{O} \rightarrow \mathbb{R}$  in the LF model:

$$\nabla \cdot \tilde{\mathbf{u}} = 0, \quad (33)$$

$$(\tilde{\mathbf{u}} \cdot \nabla) \tilde{\mathbf{u}} = -\nabla \tilde{p} + \frac{1}{\tilde{Re}} \nabla^2 \tilde{\mathbf{u}} - \alpha_{\tilde{\gamma}} \tilde{\mathbf{u}}, \quad (34)$$

$$\tilde{\mathbf{u}} \cdot \nabla \tilde{T} = \frac{1}{\tilde{Re} \tilde{Pr}_{\tilde{\gamma}}} \nabla^2 \tilde{T} + \beta_{\tilde{\gamma}} (1 - \tilde{T}), \quad (35)$$

where  $\tilde{Re}$  is the Reynolds number for the laminar flow model. Note that the governing equations in Eqs. (33), (34), and (35) are formulated in dimensionless forms. The design variable-dependent parameters  $\alpha_{\tilde{\gamma}}$ ,  $\beta_{\tilde{\gamma}}$ , and  $\tilde{Pr}_{\tilde{\gamma}}$  are defined as follows:

$$\alpha_{\tilde{\gamma}} = \alpha \frac{1 - \tilde{\gamma}}{\tilde{\gamma} + q_{\alpha}}, \quad (36)$$

$$\beta_{\tilde{\gamma}} = \beta (1 - \tilde{\gamma}), \quad (37)$$

$$\tilde{Pr}_{\tilde{\gamma}} = Pr_f + (Pr_s - Pr_f) \frac{1 - \tilde{\gamma}}{1 + q_{Pr} \tilde{\gamma}}, \quad (38)$$

where  $\tilde{\gamma} = \sum_{e=1}^n \gamma_e^{(k)} \chi_e(\mathbf{x})$  is the design variable field converted from  $\gamma^{(k)}$  to a continuous field, where  $\chi_e(\mathbf{x})$  is the indicator function of the  $e$ -th element, i.e.,  $\chi_e(\mathbf{x}) = 1$  if  $\mathbf{x}$  is in the  $e$ -th element and  $\chi_e(\mathbf{x}) = 0$  otherwise. The constants  $\alpha$  and  $\beta$  denote the inverse permeability and volumetric heat transfer coefficient, respectively.  $Pr_f$  and  $Pr_s$  are the Prandtl numbers of the fluid and solid domains, respectively, and  $q_{\alpha}$  and  $q_{Pr}$  are tuning parameters controlling the convexity of the interpolation functions.  $\alpha_{\tilde{\gamma}}$  appears in the Navier-Stokes equation in Eq. (34) to model the solid domain [80], while  $\beta_{\tilde{\gamma}}$  is used in the energy equation in Eq. (35) to represent the heat source in the solid domain. The parameters regarding these governing equations for the LF model are listed in Table 5.

The boundary conditions in the analysis domain shown in Fig. 11(a) are defined as follows:

$$\tilde{p} = 1, \quad \tilde{T} = 0 \quad \text{on } \Gamma_{\text{in}}, \quad (39)$$

$$\tilde{p} = 0, \quad \mathbf{n} \cdot \nabla \tilde{T} = 0 \quad \text{on } \Gamma_{\text{out}}, \quad (40)$$

$$\tilde{\mathbf{u}} = 0, \quad \mathbf{n} \cdot \nabla \tilde{T} = 0 \quad \text{on } \Gamma_{\text{wall}}, \quad (41)$$

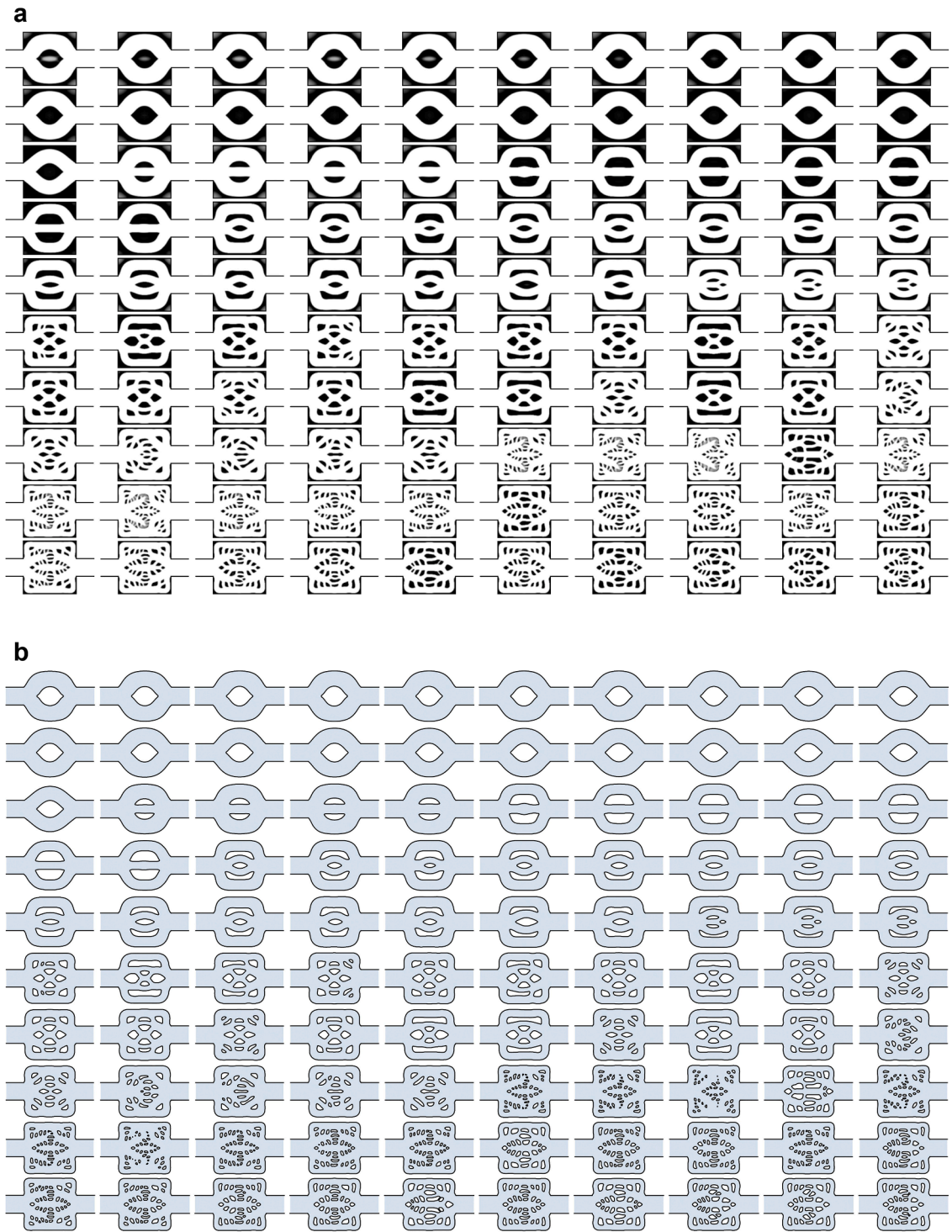
$$\tilde{\mathbf{u}} \cdot \mathbf{n} = 0, \quad \mathbf{n} \cdot \nabla \tilde{T} = 0 \quad \text{on } \Gamma_{\text{sym}}. \quad (42)$$

### 5.2.2. Formulations

The optimization problem is formulated as follows:

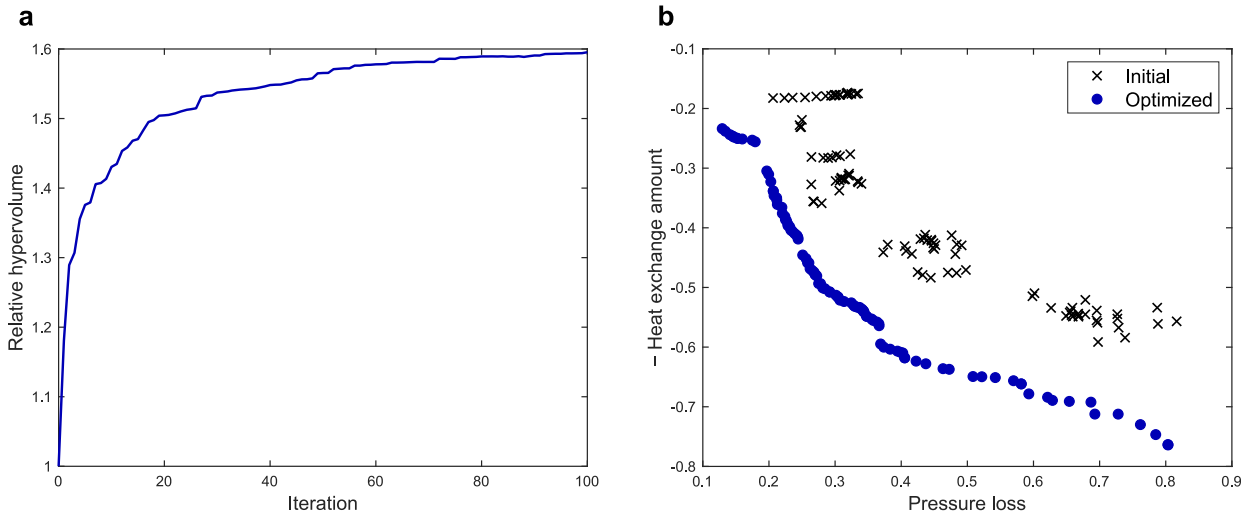
$$\begin{aligned} \underset{\gamma(\mathbf{x})}{\text{minimize}} \quad & J_1 = -\frac{\int_{\Gamma_{\text{out}}} T \, d\Gamma}{\int_{\Gamma_{\text{out}}} d\Gamma}, \\ & J_2 = \frac{\int_{\Gamma_{\text{in}}} P \, d\Gamma}{\int_{\Gamma_{\text{in}}} d\Gamma}, \\ \text{subject to} \quad & \gamma(\mathbf{x}) \in \{0, 1\}, \quad \forall \mathbf{x} \in D. \end{aligned} \quad (43)$$

Herein, setting  $J_1$  and  $J_2$  as the objective functions corresponds to maximizing the heat exchange amount between the fluid and solid domains, and minimizing the pressure loss in the fluid domain, respectively. For given material distribution  $\gamma(\mathbf{x})$ , these objective values are evaluated on the HF model by solving the governing equations in Eqs. (22)–(27) and the boundary conditions in Eqs. (28)–(31). As shown in Fig. 11(c), in the HF model, the analysis domain  $\mathcal{O}$  is discretized using body-fitted meshes with triangular elements and five boundary layers with quadrilateral elements, which are automatically built with the maximum and minimum element sizes of  $6.6 \times 10^{-2}$  and  $2.9 \times 10^{-3}$ , respectively. All state variable fields are discretized using  $\mathbb{P}^1$  Lagrange finite elements.

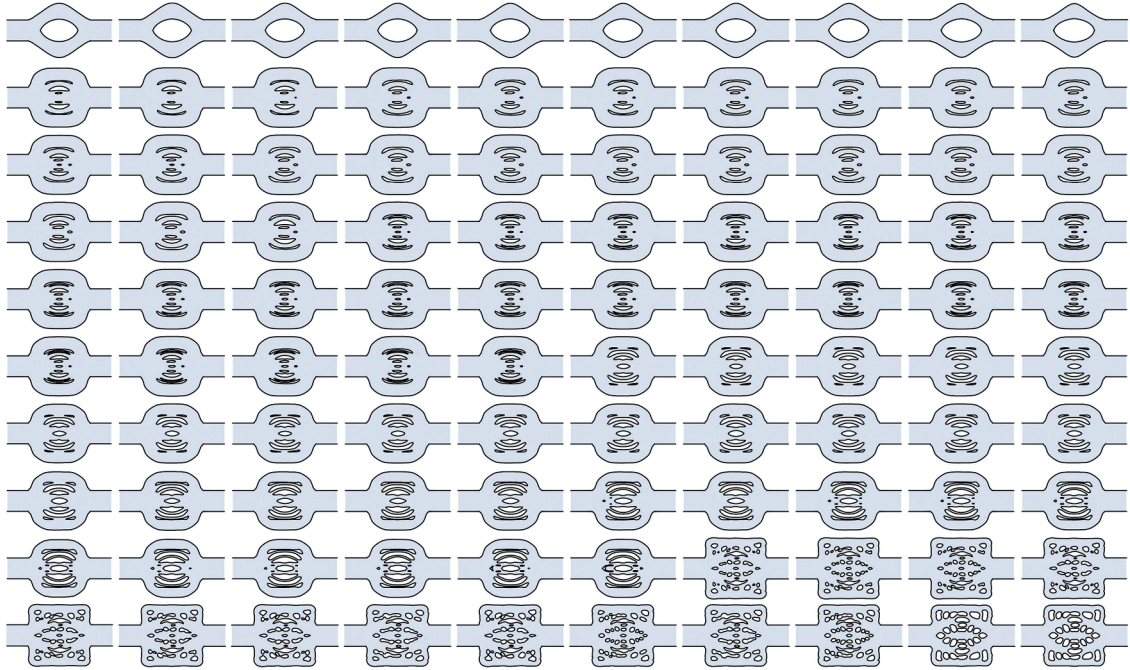


**Fig. 12.** Low-fidelity optimized initial designs in the two-dimensional heat sink design problem: (a) density distributions on the low-fidelity model; (b) material distributions on the high-fidelity model.





**Fig. 13.** Optimization results in the two-dimensional heat sink design problem: (a) convergence history of the hypervolume indicator normalized by the initial value; (b) objective space where the performance values of the initial and optimized designs are plotted.

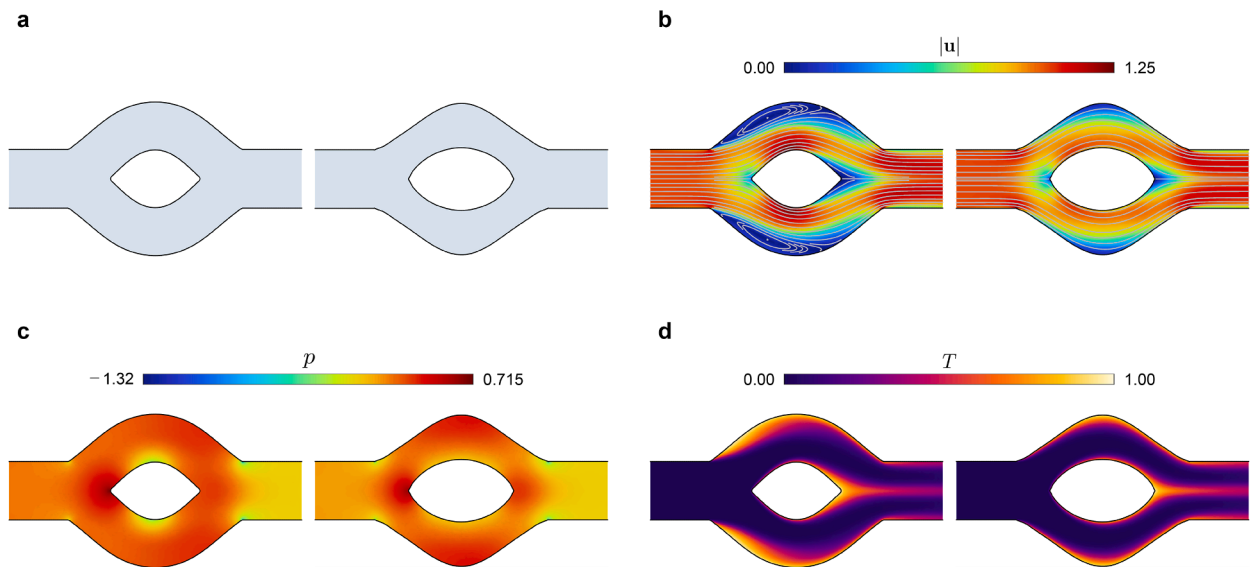


**Fig. 14.** Optimized designs in the two-dimensional heat sink design problem.

In contrast, the LF optimization problem is formulated as follows:

$$\begin{aligned}
 &\underset{\boldsymbol{\gamma}^{(k)}}{\text{minimize}} && \tilde{J} = - \int_{\Gamma_{\text{out}}} (\tilde{\mathbf{u}} \cdot \mathbf{n}) \tilde{T} \, d\Gamma \\
 &\text{subject to} && \boldsymbol{\gamma}^{(k)} \in [0, 1]^n, \\
 &\text{for given} && \mathbf{s}^{(k)} = [s_1, s_2]^\top.
 \end{aligned} \tag{44}$$

Herein, two seeding parameters,  $s_1$  and  $s_2$ , are employed to seed the Reynolds number  $\tilde{Re}$  in Eqs. (34) and (35), and the volumetric heat transfer coefficient  $\beta$  in Eq. (37), respectively. The Reynolds number  $\tilde{Re}$  is obtained by linearly interpolating between  $Re_{\min}$  and  $Re_{\max}$  according to  $s_1$ , while  $\beta$  is interpolated between  $\beta_{\min}$  and  $\beta_{\max}$  using  $s_2$ . Both  $s_1$  and  $s_2$  are sampled uniformly from the interval  $[0, 1]$  with  $n_{s_1}$  and  $n_{s_2}$  divisions, respectively, as in the numerical example in Section 5.1. Consequently, the LF optimization problem of Eq. (44) is solved independently  $N_{\text{lf}} = n_{s_1} \times n_{s_2}$  times. Compared to the original problem with a turbulent flow model under high



**Fig. 15.** Results comparison of initial (left) and optimized (right) designs with similar shapes in the two-dimensional heat sink design problem: (a) channel configurations; (b) velocity magnitude  $|u|$  with stream lines; (c) pressure  $p$ ; (d) temperature  $T$ . Herein, these objective values are:  $J_1 = -0.183$ ,  $J_2 = 0.206$  (left);  $J_1 = -0.234$ ,  $J_2 = 0.129$  (right), where  $J_1$  and  $J_2$  are the negative heat exchange amount and the pressure loss, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Reynolds number conditions, the LF optimization problem is formulated as a simplified pseudo-problem under the governing equations in Eqs. (33)–(35) and the boundary conditions in Eqs. (39)–(42), which can be stably solved under low Reynolds number conditions. Solving the LF optimization under these stably solvable conditions with seeding parameters allows us to efficiently derive diverse and promising initial designs for the original problem. It should be emphasized that the LF model is used solely for this initialization stage; the actual optimization is driven by the repeated cycle of HF evaluations under the turbulent flow model, selection, and Wasserstein crossover. Thus, although the LF model employs laminar flow conditions, the optimization itself progresses based on the thermofluidic performance under turbulent flow. As shown in Fig. 11(b), on the LF model, the design variable  $\gamma^{(k)}$  is discretized using structured meshes with 10,368 quadrilateral elements, i.e.,  $n = 10368$ .

The parameters regarding the overall algorithm are listed in Table 6. As in the numerical example in Section 5.1, the convergence check with the hypervolume indicator is not performed in this example.

### 5.2.3. Results and discussion

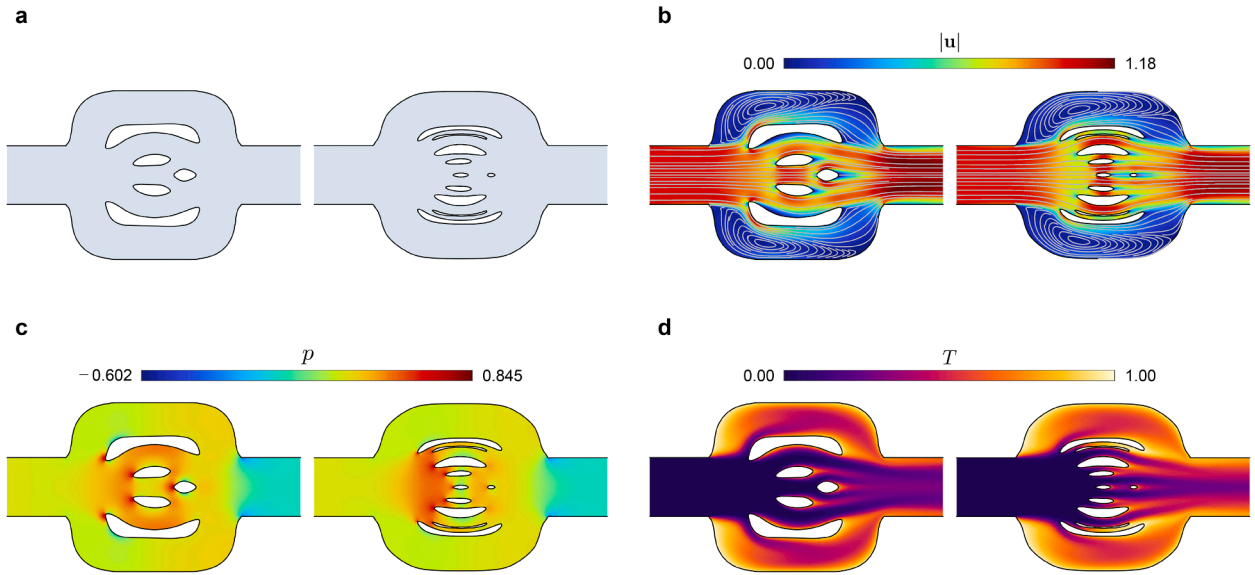
Fig. 12 shows the initial population obtained by solving the LF optimization problem of Eq. (44). As in the example in Section 5.1, seeding parameters  $s_1$  and  $s_2$  in Eqs. (44) yield diverse density distributions on the LF model shown in Fig. 12(a), which are then appropriately converted into flow channels on the HF model as shown in Fig. 12(b).

Fig. 13 shows the optimization results, including the hypervolume convergence history and the objective space. In Fig. 13(a), the steady increase in hypervolume over iterations indicates that the initial designs are progressively evolved by Wasserstein crossover, finally achieving a 60% improvement from the initial value. As shown in Fig. 13(b), both objective values have significantly improved from the initial designs, and the optimized solutions form a well-aligned Pareto front. It should be noted that the comparison between the initial and optimized designs in this numerical example implicitly serves as a baseline against gradient-based approaches with the density-based method, since the initial population is derived from LF optimization using a simplified laminar flow model, which has been widely used in the literature [5].

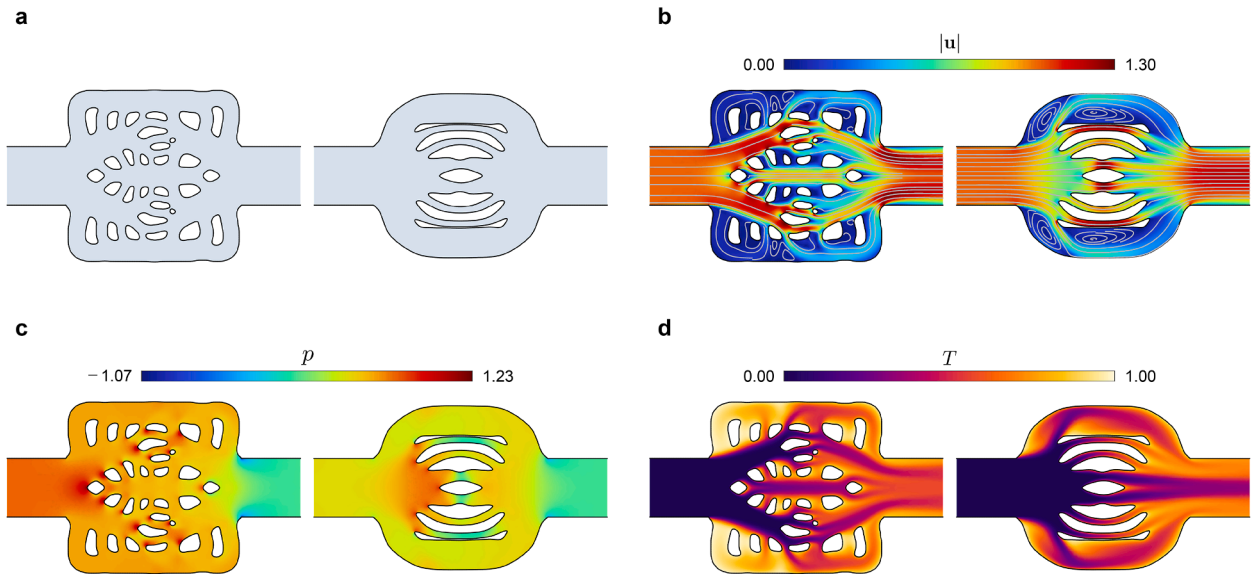
Fig. 14 shows the optimized population. Compared with the initial population in Fig. 12(b), the optimized population includes both designs that resemble the initial ones and others with entirely novel channel configurations. To further investigate their differences, Fig. 15 compares the initial and optimized channel configurations that share similar overall shapes. A particularly notable difference is in the streamlines shown in Fig. 15(b): while the initial design exhibits characteristic turbulent vortices near the top and bottom ends of the channel, the optimized design suppresses such vortices by narrowing these regions. As a result, there is a significant difference in pressure loss  $J_2$  between them.

Fig. 16 compares the initial and optimized designs under nearly identical pressure loss conditions. Focusing on the solid domains given as heat sources in Fig. 16(a), the initial design contains five such domains, whereas the optimized one has ten. As a result, the optimized configuration achieves approximately 40% better heat exchange amount  $J_1$  than the initial one. Nevertheless, as shown in Fig. 16(b), the pressure loss  $J_2$  remains comparable between them by arranging solid domains so as not to obstruct the flow.

Fig. 17 presents a comparison under approximately identical heat exchange amount conditions. As shown in Fig. 17(a), the initial and optimized designs exhibit entirely different channel configurations: the initial one resembles a pin-fin heat sink, while

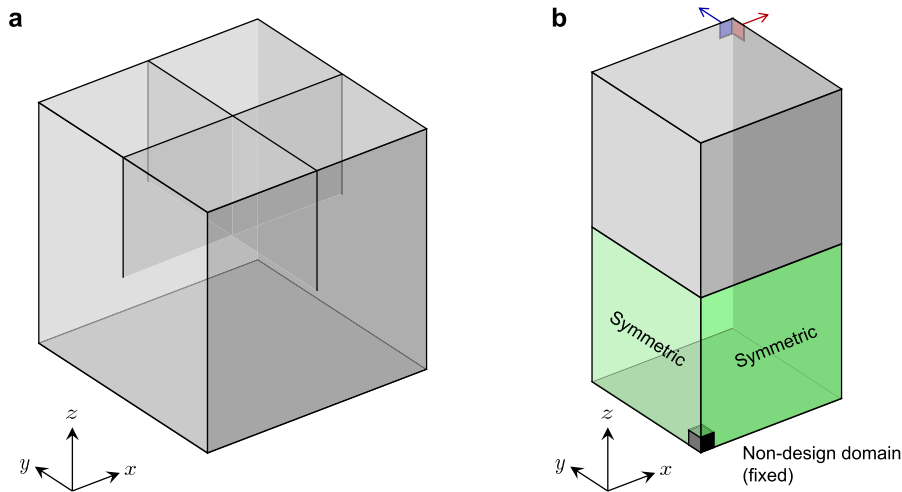


**Fig. 16.** Results comparison of initial (left) and optimized (right) designs under nearly identical pressure loss conditions in the two-dimensional heat sink design problem: (a) channel configurations; (b) velocity magnitude  $|u|$  with stream lines; (c) pressure  $p$ ; (d) temperature  $T$ . Herein, these objective values are:  $J_1 = -0.359$ ,  $J_2 = 0.280$  (left);  $J_1 = -0.501$ ,  $J_2 = 0.281$  (right), where  $J_1$  and  $J_2$  are the negative heat exchange amount and the pressure loss, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 17.** Results comparison of initial (left) and optimized (right) designs under nearly identical heat exchange amount conditions in the two-dimensional heat sink design problem: (a) channel configurations; (b) velocity magnitude  $|u|$  with stream lines; (c) pressure  $p$ ; (d) temperature  $T$ . Herein, these objective values are:  $J_1 = -0.591$ ,  $J_2 = 0.670$  (left);  $J_1 = -0.595$ ,  $J_2 = 0.269$  (right), where  $J_1$  and  $J_2$  are the negative heat exchange amount and the pressure loss, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the optimized one takes on a plate-fin-like configuration with an adjusted outer shape. This unique curved plate-fin configuration reduces the pressure loss  $J_2$  by approximately 60% compared to the initial design. In Fig. 17(b), the initial flow channel, originally optimized under a laminar flow model, shows minimal flow near the solid domains close to the inlet under the turbulent flow model. In contrast, the optimized flow channel maintains high flow velocity between all fins, indicating that effective heat exchange has been achieved with fewer solid domains than the initial one.



**Fig. 18.** Problem settings in the three-dimensional cracked box design problem: (a) the whole domain; (b) the design domain and boundary conditions defined in the quarter domain.

**Table 7**

Overall parameters in the three-dimensional stress minimization problem.

| Description                                      | Symbol            | Value                |
|--------------------------------------------------|-------------------|----------------------|
| Maximum number of iterations                     | $l_{\max}$        | 150                  |
| Population size                                  | $N_{\text{pop}}$  | 60                   |
| Number of offspring                              | $N_{\text{xo}}$   | 60                   |
| Number of LF optimized initial designs           | $N_{\text{lf}}$   | 120                  |
| Number of seeding parameters for filter radius   | $n_{s_1}$         | 3                    |
| Number of seeding parameters for volume fraction | $n_{s_2}$         | 20                   |
| Number of seeding parameters for norm parameter  | $n_{s_3}$         | 2                    |
| Minimum filter radius                            | $R_{\min}$        | 0.04                 |
| Maximum filter radius                            | $R_{\max}$        | 0.12                 |
| Minimum volume fraction                          | $V_{\min}$        | 0.10                 |
| Maximum volume fraction                          | $V_{\max}$        | 0.30                 |
| Minimum norm parameter for LF model              | $P_{\min}$        | 6                    |
| Maximum norm parameter for LF model              | $P_{\max}$        | 8                    |
| Filter radius for HF evaluation                  | $R_h$             | 0.025                |
| Minimum entropic regularization coefficient      | $\epsilon_{\min}$ | $1.0 \times 10^{-4}$ |
| Maximum entropic regularization coefficient      | $\epsilon_{\max}$ | $1.0 \times 10^{-3}$ |
| Convergence tolerance in Sinkhorn algorithm      | $\tau$            | $1.0 \times 10^{-5}$ |

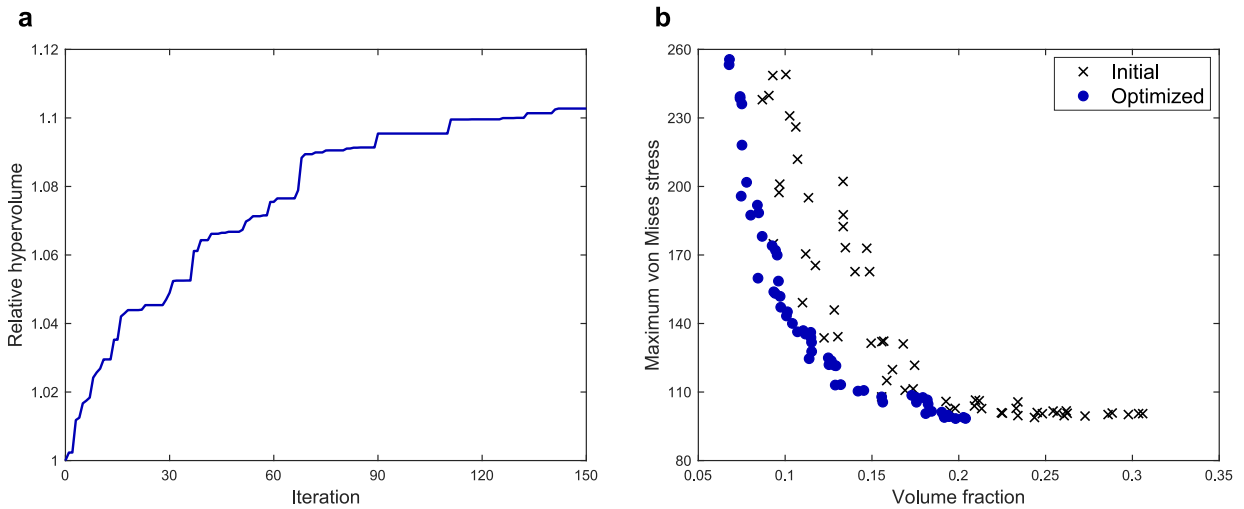
These results indicate the potential of the proposed framework to handle complex physical models such as turbulent flow. In particular, the emergence of novel flow channel configurations—absent from the initial population but introduced through Wasserstein crossover during the optimization process—appears to be a key factor in achieving significant performance improvements.

### 5.3. 3D Stress minimization problem

#### 5.3.1. Problem settings

Let us consider a three-dimensional stress minimization problem. Fig. 18 illustrates the analysis domain of the three-dimensional cracked box. In Fig. 18(a), two cracks are present in both the  $x$  and  $y$  directions, and the design domain  $D \subset \mathbb{R}^3$  is defined as a quarter of the whole domain shown in Fig. 18(b), where the cracks are represented by applying symmetric boundary conditions on the bottom part. This problem can be regarded as a three-dimensional extension of the cracked plate design problem in Section 5.1. The formulations of HF and LF models are basically identical to Eqs. (20) and (21), respectively, except that an additional seeding parameter  $s_3$  is introduced to change the norm parameter  $P$  in the  $P$ -norm stress in Eq. (21). The norm parameter  $P$  is obtained by linearly interpolating between  $P_{\min}$  and  $P_{\max}$  according to  $s_3$ , where  $P_{\min}$  and  $P_{\max}$  are the minimum and maximum values of the norm parameter  $P$ , respectively. The parameters regarding the overall algorithm are listed in Table 7.

In the LF model, the design variable field  $\gamma^{(k)}$  is discretized using structured meshes with  $40 \times 40 \times 80$  hexahedral elements, i.e.,  $n = 128000$ , whereas the HF model employs body-fitted meshes with tetrahedral elements whose maximum and minimum element sizes are set to 0.11 and  $8.5 \times 10^{-3}$ , respectively. As in the two-dimensional case, the design variable  $\gamma^{(k)}$  is discretized using  $\mathbb{P}^0$  Lagrange finite elements for the LF model, while the HF model employs  $\mathbb{P}^2$  Lagrange finite elements for structural analysis.



**Fig. 19.** Optimization results in the three-dimensional cracked box design problem: (a) convergence history of the hypervolume indicator normalized by the initial value; (b) objective space where the performance values of the initial and optimized designs are plotted.

**Table 8**

Breakdown of computational time for each numerical example (hours).

| Case                  | 2D stress minimization | 2D turbulent heat transfer | 3D stress minimization |
|-----------------------|------------------------|----------------------------|------------------------|
| LF optimization       | 0.39                   | 0.46                       | 19.01                  |
| HF evaluation         | 6.36                   | 19.58                      | 80.02                  |
| Selection             | 3.23                   | 1.18                       | 7.51                   |
| Wasserstein crossover | 21.76                  | 21.52                      | 24.54                  |
| Total                 | 31.74                  | 42.74                      | 131.08                 |

### 5.3.2. Results and discussion

Fig. 19 shows the hypervolume convergence history and the objective space as the optimization results. The hypervolume improves over iterations in Fig. 19(a), finally increasing by about 10% from its initial value. In Fig. 19(b), the Pareto front has also progressed from the initial solutions. These results suggest that even in three-dimensional cases, Wasserstein crossover can potentially lead to optimized designs with improved physical performance.

As in the previous numerical examples, we compare the representative initial and optimized designs under nearly identical volume fraction conditions in Fig. 20. The optimized designs shown in Fig. 20(a) and (b) are structurally characterized by smoother geometries that include rounded corners and edges, in contrast to the initial designs. Focusing on the stress distributions in Fig. 20(c), it can be seen that the optimized design removes the material at the intersection point of the two cracks, where stress concentration occurs in the initial design. As a result, the maximum stress  $J_1$  is reduced by approximately 13% in the optimized design compared to the initial one, while the volume fraction  $J_2$  remains nearly unchanged.

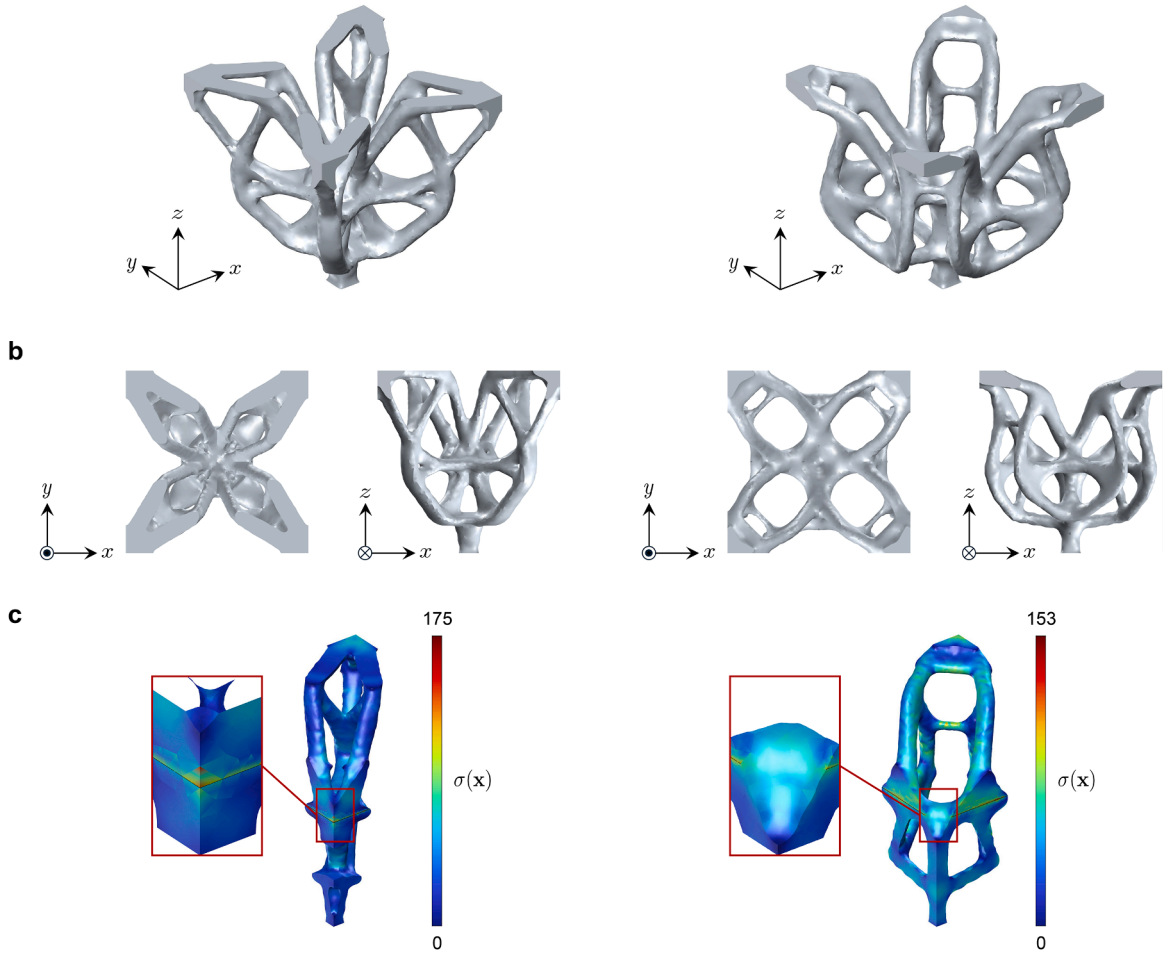
These results indicate the potential applicability of the proposed framework to three-dimensional topology optimization problems. As with the two-dimensional cases in Sections 5.1 and 5.2, the proposed Wasserstein crossover seems to produce physically reasonable material distributions that may lead to improved performance.

### 5.4. Computational cost analysis

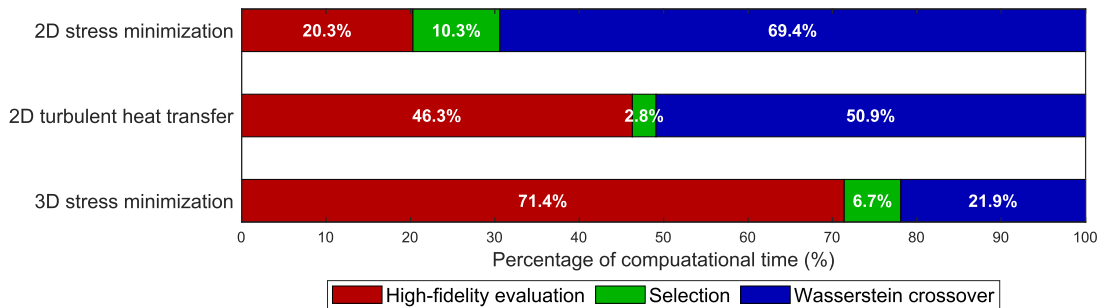
To assess the scalability of the proposed framework, we measured the computational time for each numerical example. Table 8 summarizes the breakdown across the procedures described in Section 3. First, for the two-dimensional stress minimization problem in Section 5.1, the proposed framework required 31.74 hours, whereas the DDTD framework took 16.84 hours, indicating that Wasserstein crossover incurred approximately 1.9 times higher computational cost compared to VAE-based crossover. This result suggests that, under the conditions of Section 5.1, repeatedly computing the Wasserstein barycenter is more computationally expensive than the training and sampling process of the VAE. Nevertheless, in light of the striking difference in search performance, as demonstrated by the approximately 2.6 times improvement in hypervolume shown in Fig. 6(a), the increased computational cost of Wasserstein crossover appears to be acceptable. It should also be noted that further acceleration can be expected through parallel computing techniques with multiple GPUs, as the computation of the Wasserstein barycenter for each offspring is independent.

Next, comparing the computational times across the three numerical examples in Table 8 reveals that the total costs are highly dependent on LF optimization and HF evaluation. In other words, the computational load of the finite element analysis significantly influences the overall cost of the proposed framework. Regarding the Wasserstein crossover, its computational complexity primarily





**Fig. 20.** Results comparison of initial (left) and optimized (right) designs under nearly identical volume fraction conditions in the three-dimensional cracked box design problem: (a) material distributions; (b) projections from top and side views; (c) stress distributions  $\sigma(\mathbf{x})$ . Herein, these objective values are:  $J_1 = 175$ ,  $J_2 = 9.30 \times 10^{-2}$  (left);  $J_1 = 153$ ,  $J_2 = 9.43 \times 10^{-2}$  (right), where  $J_1$  and  $J_2$  are the maximum stress and volume fraction, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 21.** Percentage of computational time within the main loop of the proposed framework for three numerical examples.

depends on the number of design variables  $n$  and the number of offspring  $N_{\text{xo}}$ . Consequently, the computational times for Wasserstein crossover in the two-dimensional problems are nearly identical. For the three-dimensional stress minimization problem, despite the quite larger number of design variables, the computational time for Wasserstein crossover is not substantially higher, which can be partly attributed to the smaller number of offspring  $N_{\text{xo}}$  compared to the two-dimensional cases. In addition, the stronger entropic regularization achieved by larger values of  $\epsilon_{\min}$  and  $\epsilon_{\max}$ , together with a relaxed convergence tolerance  $\tau$  in the Sinkhorn algorithm in Algorithm 2, led to a decrease in the number of Sinkhorn iterations, thereby reducing the computational cost.

For a clearer comparison, Fig. 21 illustrates the proportion of computational time spent on each procedure within the main loop, consisting of HF evaluation, selection, and Wasserstein crossover, excluding LF optimization with relatively low computational costs. In all cases, the majority of the computational time was consumed by HF evaluation and Wasserstein crossover, while the time for selection was negligible. The relative contribution of Wasserstein crossover becomes particularly noticeable in relatively simple problems, such as the two-dimensional solid mechanics case in Section 5.1, where the finite element analysis is computationally inexpensive. In contrast, when HF evaluation becomes more computationally demanding, as in the turbulent heat transfer problem in Section 5.2 or the three-dimensional stress minimization problem in Section 5.3, its cost dominates the main loop, and the relative impact of Wasserstein crossover diminishes. This trend suggests that the proposed framework is especially advantageous for intractable topology optimization problems in which HF evaluations are substantially more expensive. For extending the proposed framework to such problems, it will be helpful to reduce the number of evaluations through adjustments in population size or the use of HF surrogate models, which will be explored in future work. At the same time, even if HF evaluations are inexpensive, it is worth emphasizing the advantages of the proposed framework for optimization problems involving non-differentiable or strongly multimodal evaluation functions. In such cases, the EA-based search mechanism, enhanced by Wasserstein crossover, remains beneficial despite its computational cost.

## 6. Conclusions

In this paper, we introduced a Wasserstein crossover operator for material distributions in topology optimization and developed an EA-based framework. In the proposed framework, material distributions are represented as density fields following a density-based topology optimization approach. During the Wasserstein crossover, selected parent candidate solutions are interpreted as probability distributions, and their Wasserstein barycenter is computed to generate offspring. This approach allows for smooth and meaningful interpolation between parents based on the optimal transport theory, leading to a more effective crossover operator compared to conventional ones, such as the VAE-based crossover employed in DDTD. The proposed framework also incorporates the idea of multifidelity design: LF optimization, which is formulated as an easily solvable pseudo-problem, is used to derive initial designs, followed by iterative cycles of HF evaluation, selection, and Wasserstein crossover.

As numerical examples, we considered three topology optimization problems: a two-dimensional stress minimization problem, a two-dimensional turbulent heat transfer problem, and a three-dimensional stress minimization problem. In the first example, a comparison with the VAE-based crossover used in DDTD highlighted the superiority of the proposed Wasserstein crossover. Across all examples, the proposed framework effectively handled non-differentiable or strongly multimodal evaluation functions and successfully derived physically reasonable optimized designs with a computationally feasible effort. These results highlight the potential of Wasserstein crossover as a powerful crossover operator for topology optimization. Thus, a natural future direction is to further investigate its applicability to more complex topology optimization problems, such as those involving geometric nonlinearity or buckling, unsteady thermofluid problems, as well as other physics-related problems. Three-dimensional extensions of these problems will also be explored in future work, as they require substantially greater computational resources than those used in this paper.

One of the major challenges of Wasserstein crossover lies in its current limitation to problems with rectangular design domains. Although the computation of optimal transport is theoretically possible [81,82], the acceleration achieved via convolutional operations becomes restricted, leading to impractically high computational costs. To address this issue, normalization techniques such as design domain mapping [55] are expected to be beneficial, which will be explored in future work. Additionally, another challenge in the optimization framework is the need to formulate a pseudo-optimization problem that can be easily solved using the density-based method. Developing a systematic approach to construct the initial population in a problem-independent manner is a crucial topic, which will also be investigated in our subsequent studies.

## CRedit authorship contribution statement

**Taisei Kii:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis; **Kentaro Yaji:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition; **Hiroshi Teramoto:** Writing – review & editing, Project administration, Methodology, Conceptualization; **Kikuo Fujita:** Writing – review & editing, Supervision.

## Data availability

Data will be made available on request.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by JSPS KAKENHI, Japan, Grant Numbers 23H03799 and 24KJ1640.

## References

- [1] M.P. Bendsoe, N. Kikuchi, Generating optimal topologies in structural design using a homogenization method, *Comput. Methods Appl. Mech. Eng.* 71 (2) (1988) 197–224. [https://doi.org/10.1016/0045-7825\(88\)90086-2](https://doi.org/10.1016/0045-7825(88)90086-2)
- [2] M.P. Bendsoe, Optimal shape design as a material distribution problem, *Struct. Optim.* 1 (1989) 193–202. <https://doi.org/10.1007/BF01650949>
- [3] J.A. Sethian, A. Wiegmann, Structural boundary design via level set and immersed interface methods, *J. Comput. Phys.* 163 (2) (2000) 489–528. <https://doi.org/10.1006/jcph.2000.6581>
- [4] C. Le, J. Norato, T. Bruns, C. Ha, D. Tortorelli, Stress-based topology optimization for continua, *Struct. Multidiscip. Optim.* 41 (4) (2010) 605–620. <https://doi.org/10.1007/s00158-009-0440-y>
- [5] J. Alexandersen, C.S. Andreasen, A review of topology optimisation for fluid-based problems, *Fluids* 5 (1) (2020) 29. <https://doi.org/10.3390/fluids5010029>
- [6] T. Bäck, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*, Oxford University Press, 1996. <https://doi.org/10.1093/oso/9780195099713.001.0001>
- [7] X. Yu, M. Gen, *Introduction to Evolutionary Algorithms*, Springer, 2010. <https://doi.org/10.1007/978-1-84996-129-5>
- [8] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, 1989.
- [9] D. Guirguis, N. Aulig, R. Picelli, B. Zhu, Y. Zhou, W. Vicente, F. Iorio, M. Olhofer, W. Matusik, C.A. Coello Coello, K. Saitou, Evolutionary black-box topology optimization: challenges and promises, *IEEE Trans. Evol. Comput.* 24 (4) (2020) 613–633. <https://doi.org/10.1109/TEVC.2019.2954411>
- [10] C.D. Chapman, K. Saitou, M.J. Jakiela, Genetic algorithms as an approach to configuration and topology design, *J. Mech. Des.* 116 (4) (1994) 1005–1012. <https://doi.org/10.1115/1.2919480>
- [11] M.J. Jakiela, C. Chapman, J. Duda, A. Adewuyi, K. Saitou, Continuum structural topology design with genetic algorithms, *Comput. Methods Appl. Mech. Eng.* 186 (2–4) (2000) 339–356. [https://doi.org/10.1016/S0045-7825\(99\)00390-4](https://doi.org/10.1016/S0045-7825(99)00390-4)
- [12] S.Y. Wang, K. Tai, Structural topology design optimization using genetic algorithms with a bit-array representation, *Comput. Methods Appl. Mech. Eng.* 194 (36–38) (2005) 3749–3770. <https://doi.org/10.1016/j.cma.2004.09.003>
- [13] J.F.A. Madeira, H.L. Pina, H.C. Rodrigues, GA Topology optimization using random keys for tree encoding of structures, *Struct. Multidiscip. Optim.* 40 (2010) 227–240. <https://doi.org/10.1007/s00158-008-0353-1>
- [14] N. Nimura, A. Oyama, Multiobjective evolutionary topology optimization algorithm using quadtree encoding, *IEEE Access* 12 (2024) 73839–73848. <https://doi.org/10.1109/ACCESS.2024.3404594>
- [15] C.-Y. Wu, K.-Y. Tseng, Topology optimization of structures using modified binary differential evolution, *Struct. Multidiscip. Optim.* 42 (2010) 939–953. <https://doi.org/10.1007/s00158-010-0523-9>
- [16] G.-C. Luh, C.-Y. Lin, Y.-S. Lin, A binary particle swarm optimization for continuum structural topology optimization, *Appl. Soft Comput.* 11 (2) (2011) 2833–2844. <https://doi.org/10.1016/j.asoc.2010.11.013>
- [17] O. Sigmund, On the usefulness of non-gradient approaches in topology optimization, *Struct. Multidiscip. Optim.* 43 (2011) 589–596. <https://doi.org/10.1007/s00158-011-0638-7>
- [18] G. Fujii, M. Takahashi, Y. Akimoto, CMA-ES-based structural topology optimization using a level set boundary expression—application to optical and carpet cloaks, *Comput. Methods Appl. Mech. Eng.* 332 (2018) 624–643. <https://doi.org/10.1016/j.cma.2018.01.008>
- [19] S. Tanaka, G. Fujii, CMA-ES-based topology optimization accelerated by spectral level-set-boundary modeling, *Comput. Methods Appl. Mech. Eng.* 432 (2024) 117331. <https://doi.org/10.1016/j.cma.2024.117331>
- [20] K. Furuta, Y. Tsukuda, T. Sasaki, N. Ishida, K. Izui, S. Nishiwaki, S. Watanabe, Differential evolution-based topology optimization using Karhunen-Loève expansion, *Eng. Optim.* 57 (9) (2024) 2419–2445. <https://doi.org/10.1080/0305215X.2024.2400558>
- [21] S. Yamasaki, K. Yaji, K. Fujita, Data-driven topology design using a deep generative model, *Struct. Multidiscip. Optim.* 64 (3) (2021) 1401–1420. <https://doi.org/10.1007/s00158-021-02926-y>
- [22] R.V. Woldseth, N. Aage, J.A. Barentzen, O. Sigmund, On the use of artificial neural networks in topology optimization, *Struct. Multidiscip. Optim.* 65 (10) (2022) 294. <https://doi.org/10.1007/s00158-022-03347-1>
- [23] D.P. Kingma, M. Welling, Auto-encoding variational bayes, 2013, (arXiv). <https://doi.org/10.48550/arXiv.1312.6114>
- [24] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, 2014, (arXiv). <https://doi.org/10.48550/arXiv.1406.2661>
- [25] T. Kii, K. Yaji, K. Fujita, Z. Sha, C.C. Seepersad, Latent crossover for data-driven multifidelity topology design, *J. Mech. Des.* 146 (5) (2024) 051713. <https://doi.org/10.1115/1.4064979>
- [26] M. Kato, T. Kii, K. Yaji, K. Fujita, Maximum stress minimization via data-driven multifidelity topology design, *J. Mech. Des.* 147 (8) (2025) 081702. <https://doi.org/10.1115/1.4067750>
- [27] J. Yang, K. Yaji, S. Yamasaki, Data-driven topology design based on principal component analysis for 3D structural design problems, *Struct. Multidiscip. Optim.* 68 (5) (2025) 103. <https://doi.org/10.1007/s00158-025-04036-5>
- [28] Y. Yang, K. Yaji, S. Yamasaki, K. Fujita, Image fragmented learning for data-driven topology design, *Struct. Multidiscip. Optim.* 68 (8) (2025) 148. <https://doi.org/10.1007/s00158-025-04054-3>
- [29] K. Yaji, S. Yamasaki, K. Fujita, Data-driven multifidelity topology design using a deep generative model: application to forced convection heat transfer problems, *Comput. Methods Appl. Mech. Eng.* 388 (2022) 114284. <https://doi.org/10.1016/j.cma.2021.114284>
- [30] J.-W. Luo, K. Yaji, L. Chen, W.-Q. Tao, Data-driven multi-fidelity topology design of fin structures for latent heat thermal energy storage, *Appl. Energy* 377 (2025) 124596. <https://doi.org/10.1016/j.apenergy.2024.124596>
- [31] J.-W. Luo, K. Yaji, L. Chen, W.-Q. Tao, Data-driven multifidelity topology design for enhancing turbulent natural convection cooling, *Int. J. Heat Mass Transf.* 240 (2025) 126659. <https://doi.org/10.1016/j.ijheatmasstransfer.2024.126659>
- [32] K. Urata, R. Tsumoto, K. Yaji, K. Fujita, Data-driven morphological exploration and shape optimization for turbulent pipe systems, *J. Mech. Des.* 147 (9) (2025) 091704. <https://doi.org/10.1115/1.4068984>
- [33] D. Zhou, K. Nomura, S. Yamasaki, Data-driven topology design for conductor layout problem of electromagnetic interference filter, *IEEE Trans. Electromagn. Compat.* 67 (3) (2025) 872–883. <https://doi.org/10.1109/TEMC.2025.3558260>
- [34] S. Xu, H. Kawabe, K. Yaji, Evolutionary de-homogenization using a generative model for optimizing solid-porous infill structures considering the stress concentration issue, *Mater. Des.* 257 (2025) 114380. <https://doi.org/10.1016/j.matdes.2025.114380>
- [35] H. Kawabe, K. Yaji, Y. Aoki, Data-driven multifidelity topology design with multi-channel variational auto-encoder for concurrent optimization of multiple design variable fields, *Comput. Methods Appl. Mech. Eng.* 437 (2025) 117772. <https://doi.org/10.1016/j.cma.2025.117772>
- [36] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, L. Guibas, Convolutional Wasserstein distances: efficient optimal transportation on geometric domains, *ACM Trans. Graph.* 34 (4) (2015) 1–11. <https://doi.org/10.1145/2766963>
- [37] C. Villani, *Optimal Transport: Old and New*, Springer, 2009. <https://doi.org/10.1007/978-3-540-71050-9>
- [38] Y. Rubner, C. Tomasi, L.J. Guibas, The Earth Mover's distance as a metric for image retrieval, *Int. J. Comput. Vis.* 40 (2) (2000) 99–121. <https://doi.org/10.1023/A:1026543900054>
- [39] G. Peyré, M. Cuturi, *Computational Optimal Transport: With Applications to Data Science*, Now Foundations and Trends, 2019. <https://doi.org/10.1561/22000000073>
- [40] K. Yaji, S. Yamasaki, K. Fujita, Multifidelity design guided by topology optimization, *Struct. Multidiscip. Optim.* 61 (3) (2020) 1071–1085. <https://doi.org/10.1007/s00158-019-02406-4>
- [41] M. Cuturi, Sinkhorn distances: lightspeed computation of optimal transport, *Adv. Neural Inf. Process. Syst.* 26 (2013) 2292–2300. <https://doi.org/10.48550/arXiv.1406.2661>

- [42] R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums, *Am. Math. Month.* 74 (4) (1967) 402–405. <https://doi.org/10.2307/2314570>
- [43] E.K. Ryu, Y. Chen, W. Li, S. Osher, Vector and matrix optimal mass transport: theory, algorithm, and applications, *SIAM J. Sci. Comput.* 40 (5) (2018). <https://doi.org/10.1137/17M1163396>
- [44] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein GAN, 2017, (arXiv). <https://doi.org/10.48550/arXiv.1701.07875>
- [45] M.J. Kusner, Y. Sun, N.I. Kolkun, K.Q. Weinberger, From word embeddings to document distances, in: *Proceedings of the 32nd International Conference on Machine Learning*, 37, 2015, pp. 957–966. <https://doi.org/10.5555/3045118.3045221>
- [46] M. Agueh, G. Carlier, Barycenters in the Wasserstein space, *SIAM J. Math. Anal.* 43 (2) (2011) 904–924. <https://doi.org/10.1137/100805741>
- [47] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyré, Iterative bregman projections for regularized transportation problems, *SIAM J. Sci. Comput.* 37 (2) (2015) A1111–A1138. <https://doi.org/10.1137/141000439>
- [48] S. Rawat, M.-H.H. Shen, A novel topology optimization approach using conditional deep learning, 2019, (arXiv). <https://doi.org/10.48550/arXiv.1901.04859>
- [49] S. Rawat, M.-H.H. Shen, Application of adversarial networks for 3D structural topology optimization, *SAE Technical Paper* (2019). <https://doi.org/10.4271/2019-01-0829>
- [50] Z. Wang, S. Melkote, D.W. Rosen, Generative design by embedding topology optimization into conditional generative adversarial network, *J. Mech. Des.* 145 (11) (2023) 111702. <https://doi.org/10.1115/1.4062980>
- [51] L. Pereira, L. Driemeier, Wasserstein generative adversarial networks for topology optimization, *Structures* 67 (2024) 106924. <https://doi.org/10.1016/j.istruc.2024.106924>
- [52] Q. Zeng, X. Liu, X. Zhu, X. Zhang, P. Hu, Data-driven structural topology optimization method using conditional Wasserstein generative adversarial networks with gradient penalty, *Comput. Model. Eng. Sci.* 141 (3) (2024) 2065–2085. <https://doi.org/10.32604/cmescs.2024.052620>
- [53] B. Sow, R. Le Riche, J. Pelamatti, M. Keller, S. Zannane, Wasserstein-based evolutionary operators for optimizing sets of points: application to wind-farm layout design, *Appl. Sci.* 14 (17) (2024) 7916. <https://doi.org/10.3390/app14177916>
- [54] P. Ma, T. Du, J.Z. Zhang, K. Wu, A. Spielberg, R.K. Katzschmann, W. Matusik, Diffaqua: a differentiable computational design pipeline for soft underwater swimmers with shape interpolation, *ACM Trans. Graph.* 40 (4) (2021) 1–14. <https://doi.org/10.1145/3450626.3459832>
- [55] S. Yamasaki, K. Yaji, K. Fujita, Knowledge discovery in databases for determining formulation in topology optimization, *Struct. Multidiscip. Optim.* 59 (2) (2019) 595–611. <https://doi.org/10.1007/s00158-018-2086-0>
- [56] J. Yang, S. Yamasaki, Enhanced data-driven topology design with multilevel mesh and correlation-based mutation for stress-related multi-objective optimization, *Struct. Multidiscip. Optim.* 68 (5) (2025) 1–20. <https://doi.org/10.1007/s00158-025-04029-4>
- [57] L. Zadeh, Optimality and non-scalar-valued performance criteria, *IEEE Trans. Automat. Contr.* 8 (1) (1963) 59–60. <https://doi.org/10.1109/TAC.1963.1105511>
- [58] Y. Haimes, On a bicriterion formulation of the problems of integrated system identification and system optimization, *IEEE Trans. Syst. Man Cybern.* 1 (3) (1971) 296–297. <https://doi.org/10.1109/TSMC.1971.4308298>
- [59] T. Kii, K. Yaji, H. Teramoto, K. Fujita, Data-driven topology design with persistent homology for enhancing population diversity, *Int. J. Mech. Sci.* 301 (2025) 110493. <https://doi.org/10.1016/j.ijmecsci.2025.110493>
- [60] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197. <https://doi.org/10.1109/4235.996017>
- [61] S. Verma, M. Pant, V. Snael, A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems, *IEEE Access* 9 (2021) 57757–57791. <https://doi.org/10.1109/ACCESS.2021.3070634>
- [62] H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification, *Discr. Comput. Geometr.* 28 (4) (2002) 511–533. <https://doi.org/10.1007/s00454-002-2885-2>
- [63] A. Zomorodian, G. Carlsson, Computing persistent homology, *Discr. Comput. Geometr.* 33 (2) (2005) 249–274. <https://doi.org/10.1007/s00454-004-1146-y>
- [64] Y. Mileyko, S. Mukherjee, J. Harer, Probability measures on the space of persistence diagrams, *Inverse Probl.* 27 (12) (2011) 124007. <https://doi.org/10.1088/0266-5611/27/12/124007>
- [65] K. Shang, H. Ishibuchi, L. He, L.M. Pang, A survey on the hypervolume indicator in evolutionary multiobjective optimization, *IEEE Trans. Evol. Comput.* 25 (1) (2020) 1–20. <https://doi.org/10.1109/TEVC.2020.3013290>
- [66] T.E. Bruns, D.A. Tortorelli, Topology optimization of non-linear elastic structures and compliant mechanisms, *Comput. Methods Appl. Mech. Eng.* 190 (26–27) (2001) 3443–3459. [https://doi.org/10.1016/S0045-7825\(00\)00278-4](https://doi.org/10.1016/S0045-7825(00)00278-4)
- [67] B. Bourdin, Filters in topology optimization, *Int. J. Numer. Methods Eng.* 50 (9) (2001) 2143–2158. <https://doi.org/10.1002/nme.116>
- [68] K. Svanberg, The method of moving asymptotes—a new method for structural optimization, *Int. J. Numer. Methods Eng.* 24 (2) (1987) 359–373. <https://doi.org/10.1002/nme.1620240207>
- [69] B.S. Lazarov, O. Sigmund, Filters in topology optimization based on Helmholtz-type differential equations, *Int. J. Numer. Methods Eng.* 86 (6) (2011) 765–781. <https://doi.org/10.1002/nme.3072>
- [70] J. Bradbury, R. Frostig, P. Hawkins, M.J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, Q. Zhang, JAX: composable transformations of Python + NumPy programs, 2018, <http://github.com/google/jax>
- [71] R. Flamary, N. Courty, A. Gramfort, M.Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N.T.H. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D.J. Sutherland, R. Tavenard, A. Tong, T. Vayer, POT: Python optimal transport, *J. Mach. Learn. Res.* 22 (78) (2021) 1–8. <http://jmlr.org/papers/v22/20-451.HTML>
- [72] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, (arXiv). <https://doi.org/10.48550/arXiv.1412.6980>
- [73] H. Emmendoerfer, Jr., E.A. Fancello, A level set approach for topology optimization with local stress constraints, *Int. J. Numer. Methods Eng.* 99 (2) (2014) 129–156. <https://doi.org/10.1002/nme.4676>
- [74] H. Emmendoerfer, Jr., E.A. Fancello, Topology optimization with local stress constraint based on level set evolution via reaction-diffusion, *Comput. Methods Appl. Mech. Eng.* 305 (2016) 62–88. <https://doi.org/10.1016/j.cma.2016.02.024>
- [75] S. Chu, L. Gao, M. Xiao, Z. Luo, H. Li, X. Gui, A new method based on adaptive volume constraint and stress penalty for stress-constrained topology optimization, *Struct. Multidiscip. Optim.* 57 (3) (2018) 1163–1185. <https://doi.org/10.1007/s00158-017-1803-4>
- [76] O. Giraldo-Londoño, G.H. Paulino, Polystress: a Matlab implementation for local stress-constrained topology optimization using the augmented Lagrangian method, *Struct. Multidiscip. Optim.* 63 (4) (2021) 2065–2097. <https://doi.org/10.1007/s00158-020-02760-8>
- [77] J.A. Norato, H.A. Smith, J.D. Deaton, R.M. Kolonay, A maximum-rectifier-function approach to stress-constrained topology optimization, *Struct. Multidiscip. Optim.* 65 (10) (2022) 286. <https://doi.org/10.1007/s00158-022-03357-z>
- [78] M.P. Bendsoe, O. Sigmund, *Topology Optimization: Theory, Methods, and Applications*, Springer, 2003.
- [79] F. Wang, B.S. Lazarov, O. Sigmund, On projection methods, convergence and robust formulations in topology optimization, *Struct. Multidiscip. Optim.* 43 (6) (2011) 767–784. <https://doi.org/10.1007/s00158-010-0602-y>
- [80] T. Borrvall, J. Petersson, Topology optimization of fluids in Stokes flow, *Int. J. Numer. Methods Fluids* 41 (1) (2003) 77–107. <https://doi.org/10.1002/flid.426>
- [81] T. Séjourné, G. Peyré, F.-X. Vialard, Unbalanced optimal transport, from theory to numerics, 2022, (arXiv). <https://doi.org/10.48550/arXiv.2211.08775>
- [82] M. Bauer, N. Charon, T. Needham, M. Nishino, Path constrained unbalanced optimal transport, 2024, (arXiv). <https://doi.org/10.48550/arXiv.2402.15860>