Abstract—Black-box topology optimization (BBTO) uses evolutionary algorithms and other soft computing techniques to generate near-optimal topologies of mechanical structures. Although evolutionary algorithms are widely used to compensate the limited applicability of conventional gradient optimization techniques, methods based on BBTO have been criticized due to numerous drawbacks. In this paper, we discuss topology optimization as a black-box optimization problem. We review the main BBTO methods, discuss their challenges and present approaches to relax them. Dealing with these challenges effectively can lead to wider applicability of topology optimization, as well as the ability to tackle industrial, highly-constrained, nonlinear, many-objective and multimodal problems. Consequently, future research in this area may open the door for innovating new applications in science and engineering that may go beyond solving classical optimization problems of structural engineering. Furthermore, algorithms designed for BBTO can be added to existing software toolboxes and packages of topology optimization.

I. INTRODUCTION

STRUCTURAL OPTIMIZATION can be interpreted as the attempt to obtain mechanical structures to support specific load cases respecting a set of constraints. Within this context, topology optimization aims to obtain connectivity, shape, and location of voids inside a prescribed structural design domain [1].

Topology is driven from the ancient Greek word τόπος, which means place, location, domain or space [2]. Therefore, topology optimization can be defined as the study of the efficient placing of holes inside the domain. Although, the methodology has been applied with significance in the conceptual and early design stages of mechanical structures, it can be extended to solve any problem regardless of the type of structure by using Black-box topology optimization (BBTO).

In the case of solid mechanical structures, the structural design domain is usually discretized into finite elements [3]. In this case, the design variables represent the presence or density of material in each finite element directly, or, in boundary description methods, are computed from the shapes of the structural layout. Topology optimization can be classified into two main categories: gradient-based topology optimization (GTO) and BBTO. GTO relies on the use of the gradient information of the objective and constraint functions. This leads to controversy among practitioners. On one hand, when using gradient information, the optimization algorithm tends to quickly (few hundreds of iterations) converge to a solution. On the other hand, BBTO implies higher computational cost, but is more flexible in application, as it can also be applied to problems where gradient information not easily obtained or not available at all.

Hence, BBTO is supposed to fill that gap in the applicability of topology optimization by approximating solutions to the problems that cannot be solved using conventional GTO. However, there are challenges that limit its effectiveness and wide use. In this paper, we discuss these challenges, presenting techniques to relax or to avoid them. Breaking through these challenges will allow BBTO to take advantage of recent advances in evolutionary computation. Hence, allowing it to extend its effectiveness to general, industrial black-box, highly-constrained, nonlinear, multimodal and multiobjective problems. Note that the application of GTO approaches to black-box problems is also possible (although not advisable in the general case) under consideration of finite differentiation approaches. In this paper, however, we focus on BBTO in the context of evolutionary algorithms1. Recent reviews on GTO and multi-objective, metaheuristics-based structural optimization can be found in [6], [1] and [7], [8] respectively.

II. OVERVIEW

In this section, we provide a brief survey on the main methods proposed in the literature, local and global optimization algorithms, in addition to problems solved by utilizing BBTO.

1It is important to point out that approaches known under the name (Bi-directional) Evolutionary Structural Optimization or (B)ESO [4], [5] classify as GTO and not as evolutionary algorithms, since they use gradient information and lack any features of selection and variation required for an evolutionary process in the Darwinian sense.
Hybrid algorithms that combine sensitivity information with a black-box optimization approach are out of the scope of this paper.

A. Design domain representation

The design domain is the design space that is predetermined by the user to define the allowable space to be occupied by the optimum topological structure. Initially, the domain was being discretized into uniform spaced joints and the design was represented by the members which connect these joints [9], [10] as shown in Fig. 1(a). This discrete representation is limited to truss-like or frame structures [11]. As a natural step to relax the space limitation of the discrete representation, the continuum domain formulation has emerged as an alternative to overcome its limitation over the design domain [12], [13].

The traditional continuum representation is simply a binary image formulation where the domain consists of square cells (or hexagonal, e.g., [14]) and each cell is assigned zero (void) or one (material) as illustrated in Fig. 1(b) and Fig. 2(a).

![Discrete representation (a) in comparison to continuum structural representation (b).](image)

Although the binary coded representation can represent the topological details without the limitation of the discrete representation, it is associated with two main drawbacks. First, it results in stairs-like boundaries. Second, the number of decision variables is the total finite elements inside the domain. For instance, to get a fine-meshed equilateral domain of 100 elements in each direction, we would have 10,000 variables for a 2D structure and 1,000,000 variables for a 3D structure. These large-scale problems are computational prohibitive, and may need millions of function evaluations, e.g. [15]; as the total of possible solutions can be determined by the following formula:

\[ \frac{N!}{M!(N-M)!} \]  

where \( N \) is the total number of elements and \( M \) is the number of elements that could be filled with material according to the structure’s specifications, satisfying the volume constraint. Towards relaxation of these limitations, implicit representations have been introduced to reduce the problem’s scale and to allow the representation of finer topological details. The main continuum domain representations proposed are illustrated in Fig. 2. We will define them with detailed discussions in Section III.

![Illustration of main representations proposed for the structural topology:](image)

B. Search space exploration

Topology optimization has been intensively studied in the literature as a continuous differentiable mathematical problem. However, real-world applications can be complex, uncertain, discrete, and with conflicting objectives. The fitness functions can be of discontinuous landscapes, multimodal landscapes, contain ruggedness or flatten space [148]. The optimization problem can be straightforward, difficult or misleading [149].

Moreover, at the stage of product development, generating alternative designs and exploring their objective trade-offs are of crucial importance to reach an informative decision [150]. There are two search spaces that can be explored: the objective space (phenotype) and the space of the decision variables (genotype) [89], [151]. Gradient methods proceed with directional moves, that makes them very sensitive to the initial solution and the nature of the fitness landscape. On the other hand, black-box methods search the landscape without directed moves based on the gradients. The initial solutions and the fitness landscape impact the performance of the black-box methods as well [148], [152]. However, that influence is of a very little impact in comparison to the gradient methods and can be limited by using different mechanisms [153], [154]. The following approaches have been utilized in the literature of BBTO:

- Enhancement of the initial population [84], [25], [155], [142], [77]
- Multi-stage and multi-population evolutionary algorithms [67], [79]
• Clustering at the objective space [84], [85], [23], [156]
• Sensitivity predictions and clustering in the state space [61], [157], [62], [63]
• Fitness sharing and speciation [21], [158], [79]
• Custom phenotype genetic operators [19], [50], [159], [32], [64], [82], [112], [27], [130]
• Generative designs [160], [161]
• Simple-to-complex evolutionary strategy [79]

C. Optimization algorithms

The selection of the optimization algorithm is crucial to achieve satisfactory results. Many factors should be considered, e.g., the nature of the fitness landscape, the correlation between the decision variables that are determined by the design domain representation, target accuracy, and the computational budget. A wide range of algorithms have been proposed in the literature starting from a population-based stochastic method e.g., [18] to a single-solution deterministic-based method e.g., [59]. Most commonly adopted optimization algorithms can be found in Table I.

For optimization of monolithic or multi-component structures, the structural connectivity is one of the non-linear constraints that should be taken into consideration [162], [81]. In addition, a structure’s specifications and manufacturability are constraints that should be satisfied [81], [80], [147], [8], [163], [164], [165], [166]. Nonlinear constraints can be handled in two ways: (1) implicitly by using repair/filtering mechanisms, e.g., [42], [80], [81], [21], [68], [73], [46], [112], [101], [37] or assigning a penalty function to the objective function value, e.g., [81], [59], [79], [25], [19], and (2) coupled explicitly with the optimization algorithm [25], [35], [95], [102], [82], [52], [44] incorporating constraint-handling techniques, e.g., [167], [154], [168], [169], [170], [171], [172].

D. Applications

To facilitate the application of mathematical programming and the usage of gradient methods, efforts have been exerted to devise differentiable objective functions for the problems on hand. On the other hand, the gradient-free optimization methods deal with the objective functions as black-boxes, i.e., the solver does not require the objective’s gradients to proceed. Table I lists main design domain representations, optimization algorithms and applications.

Challenges of BBTO include high dimensionality, structural disconnectivity, coarse meshing, high computational cost, non-smooth boundaries, poor topological attainability, checkerboard patterns, poor quality of optimized solutions, and incapability to reproduce the results. A BBTO method is a set of elements and procedures, where a single element may affect the performance of the whole method.

The design domain representation determines the number of variables and the correlation between them. In addition, it plays a vital role in the topological attainability and the features that can be represented in the optimized solutions [74], [97]. Most of these challenges are associated with the traditional bit-array representation. Recent BBTO methods that are based on parametric level-set methods have succeeded to reduce the problem’s dimensionality, generate topologies with smooth boundaries, overcoming drawbacks of checkerboard patterns, poor quality of optimized solutions, see Fig. 3.

<table>
<thead>
<tr>
<th>Design domain representations</th>
<th>Optimization algorithms</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit bit array representation [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]</td>
<td>Binary-coded genetic algorithms (GAs) [19], [20], [64], [21], [22], [52], [65], [66], [67], [68], [69], [70]</td>
<td>Structural stiffness and/or strength [75], [18], [21], [110], [23], [67], [98], [68], [69], [111], [112]</td>
</tr>
<tr>
<td>Voronoi representation [26], [27], [28], [29], [30]</td>
<td>Branch and cut [71]</td>
<td>Vehicle crashworthiness [62], [63], [44], [113]</td>
</tr>
<tr>
<td>Dipole representation [26]</td>
<td>Real-coded GAs [72], [34], [46], [73], [74], [75], [31], [45]</td>
<td>Eigenfrequency [75], [114]</td>
</tr>
<tr>
<td>Parametric curves and graph [31], [32], [33], [34], [35], [36]</td>
<td>Multi-objective evolutionary algorithms [77], [78], [79], [80], [81], [42], [82], [23], [83], [84], [85], [86], [87]</td>
<td>Frequency selective surfaces [22], [115], [116], [117]</td>
</tr>
<tr>
<td>Map Lyndenmayer systems [37], [38], [39], [40]</td>
<td>Population-based incremental learning [51], [53], [52], [54]</td>
<td>Magnetics [95], [118], [96]</td>
</tr>
<tr>
<td>Moving morphable components and solid geometries [26], [41], [42], [43], [44], [45]</td>
<td>Evolutionary algorithms with custom operators [64], [19], [68], [69], [89], [20]</td>
<td>Bandpass Filters [119]</td>
</tr>
<tr>
<td>Material mask overlay [46], [47], [48], [49], [50]</td>
<td>Differential evolution [90], [91], [92]</td>
<td>Composite structures [120], [121], [122]</td>
</tr>
<tr>
<td>Ground element filtering and multi-resolution design variables [51], [52], [53], [54]</td>
<td>Artificial immune systems [93], [94], [95], [96]</td>
<td>Heat transfer [102], [15], [86], [123], [124], [97]</td>
</tr>
<tr>
<td>Spectral level-set formulation [55], [56]</td>
<td>Hill climbing [48], [50]</td>
<td>Fluid flow channels [125]</td>
</tr>
<tr>
<td>Interpolated level-sets [57], [58], [59]</td>
<td>Pattern search [59], [97]</td>
<td>Satellite systems [126]</td>
</tr>
<tr>
<td>Parameterized B-spline surface [60]</td>
<td>Particle swarm optimization [47], [98], [99], [100]</td>
<td>Multi-material and multi-component structures [127], [128], [129], [80], [130], [81], [79], [131], [97]</td>
</tr>
<tr>
<td>State-based representation [61], [62], [63]</td>
<td>Simulated annealing [101], [52], [102], [53], [51]</td>
<td>Optics and photonics [132], [133], [134], [135], [136], [137], [104], [105], [87], [138], [139]</td>
</tr>
<tr>
<td></td>
<td>Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [61], [62], [63], [64], [103], [104], [105]</td>
<td>Piezoelectric actuators [140]</td>
</tr>
<tr>
<td></td>
<td>Ant colony optimization [106], [107], [108]</td>
<td>compliant mechanisms [47], [141], [82], [31], [46], [49], [127], [84], [142], [85], [143], [36]</td>
</tr>
<tr>
<td></td>
<td>Constrained optimization by linear approximations (COBYLA) [55], [56], [109]</td>
<td>Biomaterials [65]</td>
</tr>
<tr>
<td></td>
<td>Random search [56], [109]</td>
<td>Protein design [144], [145]</td>
</tr>
<tr>
<td></td>
<td>Structural stiffness and/or strength [75], [18], [21], [110], [23], [67], [98], [68], [69], [111], [112]</td>
<td>Stereology [146]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Resin transfer molding [147]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Helmet facemask [83]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PV panels [76]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vibration control [40]</td>
</tr>
</tbody>
</table>
Fig. 3. Level-set function representation using (a) moving morphable components and (b) Level-set function representation of interpolated fixed knots.

Furthermore, the attainability of high-resolution topologies with complex features is demonstrated in [97]. Some examples are given in the Appendix.

The role of the optimization algorithms comes next. The optimization algorithm can be selected based on the design domain formulation, the correlation between the variables, the number of conflicting objectives, the nature of constraints and the landscape of the problem’s objective domain. Although the dimensionality can be significantly reduced by using effective design representations, hundreds of variables are needed to attain detailed and complex features [97]. Thus, development of optimization algorithms that are capable to deal with such a large-scale problem is crucial to advance BBTO. Advancement in hardware acceleration, e.g., GPUs and multiprocessing, can be utilized to reduce the computational expense, however, efforts to increase the computational efficiency of the optimization algorithm is still needed to make it possible for BBTO to be added to software toolboxes and commercial packages of topology optimization.

The rest of the paper is organized as follows. We will discuss different continuum formulations to represent the design domain in Section III, whereas efficient evolutionary algorithms that are recommended to solve the BBTO as a large-scale problem will be reviewed in Section IV. In Section V, we will present approaches to speed up computation using high-performance computing. In the last section, we will discuss some of example applications on topology optimization, discussing how such applications can benefit from BBTO.

III. CONTINUUM DOMAIN FORMULATIONS

The general, continuum topology optimization problem can be formulated as [6]:

$$\min_{\rho} F(u(\rho), \rho)$$

s.t. : $G_0(\rho) = \int_{\Omega} \rho(x) dV - V_0 \leq 0$ \hspace{1cm} (2)

$$G_l(u(\rho), \rho) \leq 0, \ l = 1, \ldots, L$$

$$\rho(x) = 0 \text{ or } 1, \ \forall x \in \Omega$$

where the density variable $\rho$ describes the material distribution depending on location $x$ in the design domain $\Omega$ and can take the value 1 for material and 0 for voids. The objective function $F(\rho, u(\rho))$ is to be minimized, where $u(\rho)$ is the state obtained from solving the governing equation and $G_0$ and $G_l$ are optimization constraints with the target volume $V_0$.

The function $F$ represents a fairly wide variety of objective functions especially it is not limited to aggregations over the structure which is often the case in classical mechanical problems. However note that the purpose of the formulation is to provide a basis for discussion, not to formally restrict the material discussed in this article. Many of the presented approaches could also serve for extended formulations (such as formulations with more than two materials, where $\rho$ can take more values than two, or multi-objective problems where $F$ returns a vector).

Typically, two major challenges have to be faced when dealing with topology optimization problems in this formulation (or extensions). The first challenge is the variability of possible solutions. For instance, in the discretized case the number of possible solutions can be quantified by a high but finite number as in (1). In general, material is freely distributed within the design domain $\Omega$ and able to form arbitrary shapes and connections. In fact, structural parts may be arbitrarily small, reaching the domain of anisotropic micro-structures i.e., there is a lack of closeness of the possible solutions [173]. In the face of this design freedom, any suitable BBTO optimization method needs a representation that is able to capture the appropriate solutions for the considered problem without oversimplifying or over-complicating it too much.

The second challenge is the computational expense for evaluations of the objective function. One evaluation typically requires at least one finite element analysis. Hence, in (2), $u(\rho)$ associates the design space with a mathematical physics model of the addressed problem. Commonly, the corresponding differential equations are solved by numerical approaches. Depending on the application domain, often solid mechanics, the problem is typically analyzed by a linear or non-linear finite element analysis. For instance, the famous 88-line topology optimization performs a linear elastic analysis to assess the compliance and sensitivities of the structure [174]. Many linear elastic problems of moderate mesh size can be analyzed on contemporary workstations within a moderate time of a few seconds up to a few minutes. Other topology optimization problems may require responses based on nonlinear analysis, such as compliant mechanisms [173] or crash problems [175]. Depending on the problem, also responses based on alternate physics solvers such as heat conduction, computational fluid dynamics or multi-physics simulations might be used (see e.g., [1] for an overview). However, the simulation time may increase significantly with increasing resolution of the analysis mesh. An example is the field of crashworthiness simulations, where a single evaluation can require hours or even days to run on several machines of a computational cluster. Therefore, an efficient optimization approach is of critical importance for real applications, for which performing as many as hundreds of evaluations might be a serious challenge for an engineer or designer within a company.

To address the two previously indicated challenges, we want to emphasize the importance of the parameterization of the problem in this section. While plenty of optimization algorithms are available in the literature to be used as black-box optimizers (see Section IV), at least as important as
the optimization scheme, is the meaning of the optimization variables with respect to the actual design. The right choice of parameterization will adjust the variability suitably and will limit the number of variables so that an efficient BBTO is possible.

Many efficient and successful optimization algorithms are originally inspired by the field of evolutionary computation. The biological inspiration is based on the concepts of “genotype” and “phenotype”. The phenotype of a solution is exposed to the actual evaluation and hence the selection operator. The genotype is the underlying vector of optimization variables that are “seen” by the optimization process and modified by its variational operators. Eiben and Smith define a representation as a “mapping from the phenotypes onto a set of genotypes”[176].

Common representations are binary encodings used in genetic algorithms [177]; for instance, when a continuous number is represented as a binary number. A more natural encoding is that of a binary encoding for decision problems such as the Travelling Salesperson Problem. Other natural encodings are real-valued design variables in evolution strategies [178], [179]; which is particularly useful for engineering optimization problems.

In the context of problem (2), formally, we can describe the representation, or encoding, \( R \) as:

\[
R : \rho(x) \rightarrow \theta \ \forall \ \rho(x) \in \mathcal{P}_R
\]  

where \( \theta \in \mathbb{R}^M \) is the vector of design variables and \( \mathcal{P}_R \) is the set of represented phenotypes. Importantly, the number \( M \) of design variables is the dimensionality of the search space spanned by the representation.

Explicit knowledge of the encoding is not necessary. Practically, it is only necessary to implement the inverse operation, i.e., decoding \( D \):

\[
D : \theta \rightarrow \rho(x)
\]

Then, a black-box optimizer can be applied to address the optimization problem (2) in the chosen representation resulting in the new formulation:

\[
\begin{align*}
\min_{\theta} \quad & F_D(\mathbf{u}(\theta), \theta) \\
\text{s.t.} \quad & G_0(\theta) = V(\theta) - V_0 \leq 0 \\
& G_l(u(\theta), \theta) \leq 0, \ l = 1, \ldots, L \\
& \theta_{lo} \leq \theta \leq \theta_{up}
\end{align*}
\]  

where \( \theta_{lo} \) and \( \theta_{up} \) are vectors of lower and upper bounds on the variables, \( V(\theta) \) is the volume covered by material, and the \( F_D \) indicates that the function contains the wrapping to (2).

As expressed in (3) only a subset of the possible phenotypes may be represented. Therefore, a representation needs to include close-to optimum phenotypes for the given problem. This implies that choosing the representation requires some previous knowledge of favorable solutions. The representation strongly influences the fitness landscape of the problem: the choice can introduce or avoid discontinuities or smoothness and it is possible to support or restrain the capabilities of the optimizer, which may heavily affect the progress of the optimization. Representations themselves can be complex, for instance, by involving developmental steps with a small number of parameters that encode more versatile phenotypic variation (see for instance [180]). Hence, depending on the representation, the optimizer has to deal with different search spaces. A well-chosen representation facilitates beneficial changes obtained from the search operators. This is related to the topic of evolvability in the field of evolutionary computation, which captures the ability of a representation to efficiently improve solutions [181], [182], [183].

Based on this argumentation, we conclude that the representation is an essential part of the topology optimization problem and that the representation has a decisive effect on the realizability of a black-box optimization. In the following subsections, we describe known representations from the literature with assets and limitations based on three definitions of representations [184]: grid, geometric and indirect representations. Fig. 4 shows an overview of these three fundamental categories.

In this overview, as in the majority of the related literature, we focus on distributions of a single material type and voids according to (2) but extensions for several materials are imaginable by additional variables for most of the presented approaches. Similarly, most approaches only consider two-dimensional formulations but seem to be extensible to three-dimensional problems.

A. Grid Representation

We classify a representation as a “Grid Representation” if the genotype encodes properties of fixed locations in a grid that discretizes the design space [184], [63]. The grid representations include the bit-array representation which is illustrated in Fig. 2 (a). The material distribution \( \rho(x) \) is discretized in a vector of design variables \( \rho \) with elements \( \rho_i, \ i = 1, \ldots, N \). According to our proposed notation, a majority of publications are described by \( \theta_i = \rho_i \), with \( M = N \). Since binary representations are discussed briefly already in Sec. II-A, and quite critically in literature [185], we continue the discussion with less known geometric and indirect representations.

B. Geometric Representation

We classify a representation as “Geometric Representation” if the genotype encodes properties of a set of fixed or movable shape primitives that define the geometry of the structure within the design domain [184], [63]. These primitives, or components, can be fixed or moved around freely in the design domain, or restricted by enforcing some connectivity, for instance by relating them as nodes and edges of a graph. The number of primitives can be defined by the user or can be adaptive.

Each primitive \( k \) is defined by a number of parameters \( \theta^{(k)} \) such as coordinates, radius, length or shape parameters defining a geometric domain \( \phi_k(\theta^{(k)}) \). In the most simple cases, the area or volume defined by the domain is filled with a default material or describes a void, but of course, further parameters may determine the material properties. Assuming
each primitive simply defines a geometric space filled with a default material, this can be expressed by

\[ \rho(x) = \begin{cases} 
1 & x \in \phi_k \\
0 & x \in \Omega \setminus \phi_k
\end{cases} \tag{6} \]

Note, however, there are approaches, for which this relation might be inversely defined or depend on an optimization variable for the material type. In the considered case (6), however, the complete space occupied by material is given by

\[ \Phi(\theta) = \bigcup_{j=1}^{K} \phi_k(\theta^{(k)}) \tag{7} \]

In the following, we briefly review the existing formulations.

1) Voronoi and H-representation: One of the first approaches developed as an alternative to the grid representation is the Voronoi cell parameterization proposed by Schoenauer [27], [186], [187], [188]. The genotype encodes the coordinates and the material properties of a number of Voronoi cells. Concretely, the parameters encode the Voronoi site (i.e., the center) of the Voronoi cell and an additional binary variable indicating material or void. In this way, the design domain is subdivided in a number of polyhedral-shaped subdomains. Fig. 2(b) illustrates this representation. The idea was picked up later again in [26], [28] and in more recent work on the design of electromechanical devices [189], [190]. Another early geometric representation is the holes representation (or H-representation), where the material is removed in the area of a superposition of rectangular shape primitives. This was investigated by Schoenauer et al. [186], [187], [188] and, relying on their references, even as early as 1993 [191], [192]. The primitives are defined by four real numbers, coordinates of the center and length in both dimensions. A similar approach has been explored by Saxena et al. with the material mask overlay strategy [73], [47], in which the shape primitives are circular masks, defined by coordinates, radius and a binary variable for material. In more recent publications, this variable is defaulted to the removal of material [49], [193], [194], [195], effectively resulting in a method closely resembling the holes representation with circles instead of rectangles. The circular material mask method is shown in Fig. 2(g).

2) Level Set Method: Geometric representations define the boundary between material and void implicitly by the superposition of the shape primitives. If this boundary is defined mathematically by a superposition of basis functions, we enter the domain of level set methods for topology optimization. Level set methods define the boundary between material and void as the contour of a level set function. Often level set functions are composed of a large number of basis functions in order to also represent fine shape variations and are therefore typically optimized with gradient-based methods [196], [197], [198]. The basis functions, e.g. radial basis functions, can be considered as shape primitives, or a complete level set function as a parameterized shape. Hence, we can consider them geometric representations as defined in this section.

Initially, the computational cost of genetic algorithms was found to be too large when using a topology description function [199]. However, recently, methods using explicit level set functions appeared whose basis functions are suitable for BBTO. For instance, a structural boundary may be represented by a Kriging model, interpolating values of a level set function at knot points [58], [81], [59], [79]. Another approach is the Movable Morphable Components (MMC) framework, where beam shaped components define the level set boundary, hence this relates to level set methods based on geometric basis functions [200], [201], [202], [203]. The MMC approach lead to an evolutionary level set method applied to crashworthiness topology optimization [41], [103], [43], [44], [113]. Yet another representation with more design freedom is based on closed B-splines (CBS) [204] instead of predefined beams.
3) Graph Representation: For some of the mentioned geometric approaches, the shape primitives move around freely in the design domain by means of varying their coordinates. However, for many problems, a concept of connectivity between the components can be useful. For instance, a geometric representation can be augmented by a graph that defines connections so that the genotype encodes locations of the nodes and properties of edges such as form or thickness. Edges may represent beams or spline curves, which are connected in the nodes.

A representation based on Bézier Curves was proposed by Tai and Chee [31], [205]. They connect loads and supports by Bézier curves and the end and control points form the genotype. The structure is obtained by assigning material to all elements passed by the curve, where surrounding elements are added based on a thickness property of the curve. In later publications, the variables are explicitly encoded in a graph and optimized with graph specific crossovers [33], [34], [206], [207]. The graph representation with Bézier curves is shown in Fig. 5.

Although the authors of the mentioned papers focused on a graph representation using Bézier curves, they performed similar investigations also for beams instead of splines [208]. Sauter et al. proposed a similar representation using complex-shaped beams connected by a graph [209], [210]. Their approach considers straight, variable-thickness, and curved variable-thickness beams with specialized operators such as splitting and merging of beams. Another, hybrid approach, combining a grid and a geometric representation is proposed by Balamurugan et al. [68], which uses a bit array optimization that yields a skeleton starting point for a graph optimization, where the nodes are rectangles of material connected by edges defined by the skeleton. In [84], [143], the structure is defined by piecewise linear segments with different length and orientations. Recently, a constructive solid geometry is proposed by Ahmed et al. [211], [212], [42], [45]. In their work, nodes of the graph define start and end points of rectangular bar primitives and are optimized. The bars i.e. edges of the graph are obtained by Delaunay triangulation and combined by using solid modeling operators such as the constructive solid modeling union operator that yields a material domain for overlapping bars. Fig. 6 shows the principle of this approach.

C. Indirect Representation

A representation classifies as an “Indirect Representation” if the genotype encodes variable properties of a generative model, which implicitly defines material locations or a geometry [184], [63]. Boundaries between geometric representations and indirect representations can be slightly blurred. For example, the geometric primitives in the Voronoi representation and level set methods are also implicitly defined. However, in contrast to a descriptive model, the genotype $\theta$ of an indirect representation encodes properties of a more abstract process to obtain the material distribution in the structure. These parameters are, for instance, parameters of rules or a development process, possibly inspired by onto- or morphogenesis in nature. Usually, the mapping $D(\theta) \rightarrow \rho(x)$ can be very non-linear, possibly requiring a complex, dynamic simulation of a development process. Indirect representations known in the literature are presented in this section.

1) Lindenmayer system: Hornby used a generative Lindenmayer system (L-system) representation to evolve a family of three-dimensional table-like structures [213], [214], [215]. The L-system is a grammatical rewriting system, modeling the cellular division process of organisms, e.g., branched topology in plants. Although Hornby’s work does not explicitly address a topology optimization task and no finite element analysis for evaluating the performance is used, the task of finding a table-like structure resembles a topology optimization in the sense that the distribution of material in a voxel-based design domain is optimized. Explicitly, a more recent approach uses the L-system as an indirect representation for topology optimization [216], [37]. The L-system model features successive cell division processes. The parameters that are encoded in the genotype are an axiom, production rules and additional physical and geometrical properties such as the thickness of edges. An illustration of the topology growing process is shown in Fig. 7. The topology is obtained by performing the development, starting from the axiom and successive application of the rules that lead to splits of design domain regions, where each split introduces a new edge. Besides minimum compliance, this representation has been applied to the optimization of wing designs [217], [218].

2) Gene-regulatory Network: Another biologically-inspired indirect representation was proposed by Steiner et al. [219], [220], [221], [222]. The representation encodes a gene-

![Fig. 5. Example of a Bézier curves representation connected by a graph [33].](image1)

![Fig. 6. Example of the solid geometry modeling representation [212].](image2)

![Fig. 7. The starting point and the first three developmental steps of the cellular division process of the L-system representation [37].](image3)
regulatory network that determines the behavior of a cellular growth process, where cells grow inside a design domain. The genotype is composed of parameters that form regulatory and structural units. The state of the cell and its environment influence the regulatory unit that can trigger the activation of the structural units. In the basic variant, the model includes the physical interaction of cells, cellular division, and cell specialization into material and void cells [220].

An enhanced model [221] includes a chemical Gauss-shaped gradient that is used for chemotaxis, i.e., an additional force that is acting on the cells and can also be read by the regulatory units for gene activation. The gradient is pre-defined and does not change during the developmental process and can be considered as a means of providing global directional and location information. Additionally, forces on cell adherence and repulsion are introduced. Fig. 8 shows an example of the growing process into a topology.

3) **State-based Representation**: Another method for an indirect representation of the structure is to represent the design space based on the structural state [62], [63]. Every evaluation using a finite element analysis not only yields the objective function value but also detailed state information on, for instance, nodal displacements, stresses and energies that are locally related to every element. Based on this state information, groups of similar elements are obtained by clustering in the space of selected local state features. For a small number of element clusters, a low dimensional representation is obtained, where elements are close in feature space, but not necessarily close, or even connected, in the design domain. This representation is illustrated in Fig. 9. The result of optimizing this representation is used to iteratively change the structure so that the corresponding state will change and the representation is adapted when necessary. Hence, the optimization is turned into a sequence of optimizations with a lower dimensional representation. This approach forms a generic topology optimization approach, where a gradient-like optimization is performed based on a model that substitutes elements’ sensitivities based on the feature input [63]. Alternative approaches to train the model are by finite differences [157] or by training of a neural network model [61], [223].

4) **Compositional Pattern Producing Network**: An abstraction model of a development process is proposed in the form of Compositional Pattern Producing Networks (CPPN) [224].

The CPPN avoids to explicitly model a development from a small starting point to its final form. A CPPN model is similar to an artificial neural network model. Applied for topology optimization it processes coordinate information on the finite elements in the design domain as inputs. These inputs flow through nodes with regular mathematical activation functions. The network output then determines the state of the material in the considered elements. The CPPN model, i.e., neurons, functions and their parameters are optimized with a neuro-evolution by augmenting topologies algorithm [225]. Its usage as representation for optimizing material in a design domain, i.e., topology optimization, has been studied in two independent works [226], [76]. The concept is shown in Fig. 10.

5) **Variational Autoencoder**: With the recent advent of neural network models with many layers and millions of tunable parameters, these have become state of the art for many machine learning problems for which plenty of data is available, for instance in the domains of object recognition, image segmentation, and natural language processing. Based on these deep neural networks a new type of unsupervised dimensionality reduction method has become more popular, so-called Autoencoders [227]. The target of the training process is a reconstruction of the input at the output layer. The network architecture is composed of an encoder, a latent layer (a kind of bottleneck for the information) and a decoder. Hence to achieve a precise reconstruction, the low-dimensional latent layer needs to be an abstract representation that captures most information on the content. As an extension, *variational* autoencoder trains to reproduce the data by capturing the probability distribution of the input data in the latent space. In the context of topology optimization, this representation has been explored by using large datasets of topology optimization results to train the autoencoder for reconstruction of optimized structures, see Fig. 11 for an illustration. Naturally, the reconstruction quality depends on the size of the latent representation, which can be chosen freely. Yu et al. [228] use the obtained model to predict the optimal structures for given boundary conditions. Guo et al. [124] use the representation to perform optimization in latent space for a different objective function which was used to create the training data. The

![Fig. 8. A cellular growth model based on motile polarized cells and voxelization [221]. The cells iteratively move, divide and specialize into a structure.](image)

![Fig. 9. State-based representation approach [62]: The structure is divided into clusters based on the structural state information obtained from the simulation.](image)
work in the topology domain has been restricted to two-dimensional cases but in the domain of object classification autoencoders are already applied to voxel spaces [229], which may provide insights on directions for future applications to three dimensions.

Multi-objective optimization: Problems dealing with two or more (often conflicting) objectives can be addressed by so-called multi-objective evolutionary algorithms (MOEAs) instead of the typical weighting schemes in GTO. Research on MOEAs has flourished in the last 15 years giving rise to a wide variety of approaches. Currently, the most popular MOEAs for few objectives (two or three) are NSGA-II [238] and SMS-EMOA [239]. For more than three objectives (the so-called many-objective problems), there are many proposals [240], but some frequently used MOEAs include NSGA-III [241] and MOEA/D [242]. A few papers solved multi-objective topology optimization problems. For instance, NSGA-II was used to minimize volume and compliance in [42], [84], and to solve a many-objective multi-component problem considering manufacturability in [80], [81]. Besides multi-physics and multi-discipline problems, typical structural problems involve several load cases, which can be considered as different objectives.

Expensive objectives: In real-world problems, it is very common to deal with expensive objective functions, whose evaluation may take minutes, hours or even days. Such problems are normally handled using surrogate methods, e.g., [243]. For instance, Efficient Global Optimization (EGO) [244] with a Kriging model is used in [125] for topology optimization of fluid problems. Large approximation errors may have a negative impact on search performance, leading to false optima. However, the uncertainty of the surrogate model can be beneficial to the evolutionary algorithm by smoothing noisy fitness landscapes, and thus accelerating the convergence and avoiding trapping in local optima [245], [246]. The selection of the surrogate model and the infill criterion is crucial for the success of the surrogate-based optimization process [247]. We refer the reader to [248], [249] for detailed discussions on opportunities for surrogate-assisted evolutionary computation, as well as [250] for parallel-based algorithms. Hybrid algorithms with integrated local searches and/or the inclusion of gradient or approximate gradient information can be an alternative for problems with computationally expensive evaluations. Such hybrid algorithms would be worth a separate review, however some examples of hybrid algorithms that use both gradients and global search methods are [62], [69], [251], [252]. A review on comparing different derivative-free algorithms, including increasing problems sizes can be found in [253].

IV. EFFICIENT EVOLUTIONARY ALGORITHMS

Many types of evolutionary algorithms have been proposed over the years with the aim of addressing challenging problems such as the following:

Constrained optimization: Constraints are very common in real-world applications, and a number of constraint-handling techniques specially designed for evolutionary algorithms have been proposed. Mezura and Coello [172] categorized the constrained-handling approaches into feasibility rules, e.g., [230], stochastic ranking, e.g., [231], ε-constrained method [232], Penalty functions, [233], [234], special operators, e.g., [235], multi-objective concepts, e.g., [236], and ensemble of constraint-handling techniques, e.g., [237].

Classical topology optimization formulations as in (5), at least have to balance the material versus a structural performance metric, where one of them typically comes as a constraint. Hence any algorithm applied to a topology optimization has to be able to handle constraints. Besides volume, classical constraints result are geometric, e.g. resulting from manufacturing constraints. Depending on the applied representation, structural feasibility (e.g. connectivity) can be a main constraint. It can be handled implicitly by using repair mechanisms or explicitly by using constraint-handling techniques [25].
parallel algorithms can be found in [255].

**Niching:** The use of niching [257], which allows clustering together solutions that share some similarities is a useful tool when dealing with highly nonlinear problems and deceptive landscapes. Niching mechanisms are implemented in BBTO methods and showed tangible success as in [21], [158], [79]. From the application perspective, niching improves the capability of population-based algorithms to find more than just one solution. A set of conceptually different solutions to choose from can support engineers in their creative process to select the best structure according to additional criteria not obvious to the optimization as are typically present in real-world applications.

The selection of the optimization algorithm is crucial to achieving satisfactory results. Although general algorithms can be devised based on the design-domain formulation, customization of these algorithms can be more effective to the problem at hand. For instance, local search algorithms have the advantage of fast convergence. However, they may not be effective in solving problems with noisy multi-modal landscapes. On the other hand, a memetic evolutionary algorithm with a niching mechanism can solve a wide range of problems, but at a higher computational cost. Advantages of robustness, global-seeking capability, and wide applicability are common characteristics of global search algorithms.

### V. Speeding up Computation Using GPU and Multiprocessing

Thanks to the advent of high-performance computing hardware and software, engineers and designers can simulate mechanical structures and physical systems governed by sophisticated mathematical models. These simulation algorithms can be further coupled with mathematical optimization frameworks to explore the optimized topology, geometry, and material layout of physical designs. The bottleneck of this pipeline lies in the process of evaluating the physical/mechanical performance (e.g., structural compliance, target deformation, etc.) given a set of system parameters, which typically requires solving a set of partial differential equations with thousands to millions of unknowns. This evaluation happens in every iteration of the optimization loop. Developing new simulation techniques, in particular, those fast and predictive approaches, play an essential role in reducing the computational cost associated with these optimization procedures. In this section, we will review a few branches of techniques to boost the efficiency of large-scale numerical simulation of complex mechanical systems, with focuses on hardware and data structures.

#### A. Hardware Acceleration

The most expensive step in topology optimization is the evaluation of the system’s objective for the current material distribution, which usually requires solving a large-scale sparse system. The rapid development of parallel computing hardware, in particular, General-Purpose Graphics Processing Unit (GPGPU), opens up new possibilities for accelerating these solvers. A vast literature [258], [259], [260], [261], [262] has been devoted to the development of GPU-based algorithms to solve large-scale systems that are sparse, linear and symmetric positive definite.

A large variety of algorithms have been developed in the effort to extend a traditional CPU-based iterative solver, such as a preconditioned conjugate gradient solver, to its GPU implementation for direct efficiency gain. These iterative solvers are particularly suited to be implemented on a modern GPUs for parallelization. The high memory bandwidth, high peak computational throughput, and the single instruction multiple data (SIMD) architecture of the GPU hardware allow the fast implementation of several critical operations in such solver, including the matrix-free implementation of the sparse matrix-vector multiplication [263], the red-black Gauss-Seidel smoothing [264], and the geometric multi-level preconditioning [265]. The combination of these parallelization techniques in one solver enables orders of magnitude acceleration compared to its streamlined CPU counterpart, opening up possibilities to a broad range of new applications, such as real-time simulation [265], evolutionary computing [266], meshfree simulation [267], unstructured mesh simulation [268], and super-scale topology optimization with up to a billion of degrees of freedom [262]. In addition, regarding the physical equations, GPU algorithms have proved the efficiency in accelerating a diversity of applications, such as thermal conductivity [258], elastodynamics [269], and so on. In particular, the linear elastic GPU solver based on finite element discretization [259] [262] has been drawing the most attention in the topology optimization community.

Most GPU-accelerated numerical solvers run on a single GPU with relatively small data size—up to millions of unknowns with double precision—due to the advantage of GPU’s high internal memory bandwidth and peak computation throughput. The communication between CPU and GPU is minimized, in order to maximize the latency of the entire system. This restriction makes scalability become one of the primary challenges for a GPU solver. For example, for a three-dimensional elastic FEM system, an NVidia GeForce GPU with 6G memory can hold the hierarchy of multigrid matrices (stored in CRS format and double precision) with up to two million elements. In contrast, a CPU-based parallel solver (e.g., the FEM solver in PETSc [270]) can scale the problem up to 83 million voxels on a cluster where memory is not the bottleneck. To reduce the memory usage on GPU, a matrix-free representation for the sparse matrix turns out to be efficient (e.g. see [271], [272]). Instead of explicitly storing the sparse matrix, a matrix-free approach reads very few parameters from memory and assembles the local matrix block on-the-fly when a linear operator such as matrix-vector multiplication is performed. Such technique (e.g., see [262]) enables high-resolution topology optimization results at the level of 14 million active FEM elements on a single Nvidia GPU with 8G memory.

For large-scale simulation problems, the development of numerical algorithms that fit GPU clusters (multiple GPUs connected by PCIe bus) or heterogeneous platforms (CPU multiprocessors enhanced by multiple GPUs) have been drawing attention in many areas. Some pioneering explorations in this direction include [273] and [274]. The key philosophy to
build these hardware architecture-aware numerical algorithms is to keep the data on GPU for local computation for as long as possible, in order to minimize the data transfer cost over the PCIe bus. Different strategies have been taken on different parallelization level. One strategy is to assign processors different tasks for coarse-grained parallelism. For example, in [273], multiple GPUs are assigned different computational tasks for uncertainty propagation to achieve task-level parallelism. Heterogeneous computing platforms have also demonstrated their efficacy in accelerating large-scale simulations. For example, a multigrid solver based on the Schur-Complement scheme [274] has increased the resolution to one billion active elements, on a heterogeneous CPU-GPU workstation, taking advantage of both processors.

B. Adaptivity and Sparsity

Density [12], [275] and interface [196], [197] are the two most commonly used material representations for topology optimization. Among the various data structures that have been developed in the literature, uniform grids play an essential role in tracking the material evolution for their multiple advantages in computation. These advantages include cache-coherent memory access, regular subdivisions for parallelization, simple data layout, and, in particular, the existence of efficient numerical PDE solvers. A multigrid FEM solver discretized on a uniform grid has been established as one of the standard physical equation solvers for topology optimization (e.g. [276]).

1) Adaptivity: The main weakness of a uniform grid lies in its lack of adaptivity. To improve this, researchers inverted hierarchical data structures maintaining grid cells with different resolutions, e.g., the octree grid [278] and the AMR grid [279]. These data structures can adaptively refine a single grid cell to allocate computational resources around regions of interest. High-performance solvers [280] are combined with an octree grid to solve large-scale problems with adaptivity. Instead of adding new cells, researchers also deform grid cells, e.g., by translating [281] or bending [282] grid lines, to get finer discretizations around important regions. This strategy preserves most of the computational advantages of a uniform grid. Multiple overlapping grids with different resolutions [283] have also been used to improve the adaptivity of a uniform grid structure.

2) Sparsity: The target structures for topology optimization usually occupy a small portion of the design space, exhibiting sparse and adaptive features. A discretization taking advantage of this sparse nature has the potential in boosting the efficiency for both simulation and optimization. A wealth of sparse data structures have been invented, e.g., OpenVDB [284] and SPGrid [285], to reduce the computational cost of by maintaining active elements selectively. The essential idea is to establish an efficient mapping from a grid cell index in the real, sparse space to an index in the virtual, compact storage, enabling the allocation of computational resources only to discretizations occupied by or near the real structures. Examples of these mappings include a standard hash table (e.g., local level set [286]), an octree (e.g., adaptive distance field [287], OpenVDB [284]), or via a Virtual Memory Page Table and the Translation Lookaside Buffer (TLB) (e.g., SPGrid [285]). These data structures enable large-scale optimization. For example, Liu et al [277] invented a new narrowband grid structure to perform topology optimization for super-resolution structures (1 billion voxels) on a regular workstation, see Fig. 12.

One of main challenge remaining in the creation of efficient data structures for large-scale topology optimization applications lies in the automated co-optimization of both the discretization and the material. It is essential to allocate the computational resources in an automatic way so that the local discretizations can accommodate the modeling of smooth, complex, and thin features emerged from the background. The ultimate goal is to enable the interleaving evolutions of both structure and discretization to obtain designs with better performance.

VI. APPLICATIONS

Industrial engineering design optimization problems often involve black-box simulations. More precisely, simulations of structures or systems, for which the users do not have a precise mathematical model for determining analytical sensitivities or in general a limited expertise on the complex modeling behind the simulation. Where in academia, most often analytically traceable models are considered, in industrial development processes users are often confronted with simulation models, where only partial domain knowledge is available. These models are treated often more like experimental set-ups and validated by comparison to data from real-world experiments.

From an optimization perspective, these models are black boxes for which a general optimization approach independent of gradients is the possibly only option, as long as specialized optimizers are not available.

Interesting domains are for instance virtual development processes, in automotive or aerospace industries. Examples of real-world applications are simultaneous topology, shape and sizing optimisation, multidisciplinary optimization of aircraft morphing wings as well as automotive components and structures. Black-box simulation may be problems that involve simulations which are challenging due to anisotropic materials (e.g. handling of composites), multiphysics simulations (e.g. thermo-mechanical or fluid-structure couplings), and simulations of manufacturing processes (e.g. metal-sheet forming or stamping processes) or optimization constraints resulting thereof. The black-box context is especially relevant when the simulations consider transient processes that have to be addressed by explicit finite element solvers. For the mentioned types of simulations, GTO methods are not always readily available in general, even less in the industrial process, due to the non-linearities, or complexity of the simulated experiments, also with respect to boundary conditions or a multitude of involved components that deviate from the assumptions of simple academic white-box models.

A. Vehicle crashworthiness design

One interesting domain for topology optimization is the industrial vehicle development that involves the optimization
of many different design stages and objectives [288], [289]. Especially crashworthiness topology optimization is a possible application for the methods proposed in this article and is currently a very active field of research. Crashworthiness simulations are an important step in the automotive development process to ensure passive safety of the vehicle while reducing the number of physical prototypes for testing. Topology optimization of these models is hindered by the significant non-linearities and transient nature of the models, for which the derivation of analytical sensitivity information cannot be easily obtained. Due to the physical and numerical simulation noise and the high computational cost of the simulation, finite-differencing approaches are often difficult to realize or infeasible, as well. Under some assumptions, sensitivity information can be obtained for instance in a ground structure approach [290], [291]. This offers a chance for future hybrid approaches as well that may be able to deal efficiently with complex crash mechanics such as material failure or very non-linear objective functions such as accelerations.

Currently, specialized heuristic methods are applied to crashworthiness topology optimization, and among them, approaches that apply evolutionary optimization. Optimization of crashworthiness has been addressed by the state-based representation optimized with evolution strategies [62], [63], the evolutionary level set method [103], [44] and a Kriging-based optimization for a domain parameterized via a thickness clustering [292]. The challenging objective functions that are tackled include for instance control of energy absorption, intrusion or acceleration characteristics. Besides the variety for crash problems, these methods could also be an alternative to existing heuristics for objective functions found in impact mechanics problems, which are dominated by even more extreme dynamics and non-linearities, see for instance [293].

B. Design under manufacturing constraints

1) General-purpose manufacturing constraint: Early works of manufacturing constraints for topology optimization were not originated from manufacturability concerns, rather from the numerical instability and mesh-dependency considerations. It is well-known that topology optimization in a continuum domain tends to generate numerous small holes, a.k.a. checkerboards [294], [295]. For GTO, regularization schemes have been developed. There are two main categories, filtering methods (e.g., [296], [297]) and constraint methods (e.g., [295], [298]). Recently, the PDE-based filtering method [299], [300] gained popularity due to its ease of implementation and computational efficiency. For BBTO, early works [18], [21] with explicit bitmap geometric representations suffered from the same checkerboard issues. However, recent works [58], [59] based on level-set methods have naturally eliminated checkerboard features due to their implicit geometric representations. Topology optimization with regularization schemes not only avoided numerical instabilities but generated results with much simpler geometries. The reduced shape complexity usually means better general manufacturability. In BBTO methods, manufacturing constraints can be incorporated just as any other constraints, and the geometric complexity can be controlled by optimization parameters.

2) Process-specific manufacturing constraint: Manufacturing constraints modeled for specific manufacturing processes are referred to as process-specific manufacturing constraints. Such constraints are usually imposed on part geometries and are more detailed than the general-purpose manufacturing constraints. There are three major advantages for BBTO to incorporate process-specific manufacturing constraints.

First, for the BBTO, intermediate topologies often have clear part boundaries. Instead, for density-based GTO, intermediate topologies are “blurry”. Since it is straightforward to evaluate the manufacturability of topologies with clear part boundaries, process-specific manufacturing constraints can be seamlessly integrated into BBTO. For GTO, however, in order to evaluate manufacturability of “blurry” intermediate topologies, modeling simplification and numerical approximation are often required, which still remains a challenging task. For example, in order to generate simply-connected topologies for additive manufacturing, the virtual temperature method [301] has been developed to avoid the formation of enclosed cavities by constraining the maximum temperature of an additional thermal conductivity analysis. This can intuitively be achieved, in a BBTO framework, by adding a penalty term to the objective function if the intermediate topology is checked as multiply-connected.

Second, since sensitivity information is required for GTO, only differentiable constraints can be incorporated. It is true that many non-differentiable constraints can be modified and
approximated as differentiable. However, such numerical treatments usually lead to less accurate modeling and convergence issues. In contrast, both differentiable and non-differentiable manufacturing constraints can be integrated with BBTO because sensitivity analysis is not required. For example, the resin filling time consideration for liquid composite molding process has been incorporated into BBTO [147]. Such detailed manufacturing constraint will be difficult to implement using GTO due to the challenge of obtaining its analytical sensitivity.

Lastly, it is more challenging to solve GTO with multiple manufacturing constraints. Since each manufacturing constraint model comes with approximation and simplification, the accumulation effect of multiple constraints will make the problem even more difficult to converge, if it is possible at all. For BBTO, it is relatively straightforward to formulate such problems as multi-objective and generate Pareto frontiers to present the optimized results. Most existing manufacturability-driven GTO had only one manufacturing constraint, e.g., casting [302], additive manufacturing [303]. Multiple manufacturing constraints have been successfully applied to BBTO [80], [81].

Continuous relaxation of BBTO problems has become an interesting research trend. For example, the multi-component topology optimization considering manufacturing and assembly cost constraints have previously been modeled as black-box problems and solved by genetic algorithms [80], [81]. A continuously relaxed gradient-based formulation for the multi-component topology optimization has recently been proposed [304], [305], [306], which enabled efficient GTO and showed promising computational efficiency improvement. However, problems with many nonlinear constraints such as maximum stress constraints, and process-specific manufacturing constraints are challenging for GTO where the resultant multimodal landscapes are required to explore for comprehensive studies. These multiple local optima cannot be easily explored by GTO methods even with a multi-start strategy as the number of possible starting point can be exponentially large (same argument as BBTO regarding the number of decision variables). On the other hand, a GA with a niching method can explore their landscapes quite easily and efficiently.

C. Multiphysics applications

Exploring the trade-offs between physical properties is essential for multiphysics topology optimization problems. In particular, it is crucial to design algorithms that are capable of exploring the trade-offs of the various possible combinations of properties, such as to find metamaterials with both optimal mechanical and acoustic properties and to design soft materials with specified electromagnetic bandwidths for soft antenna designs. The numerical measurements of these different physical properties are computationally expensive, particularly when multiple PDE systems need to be solved given one specific design. These computational costs make it challenging to incorporate the simulation models into the standard gradient-based optimizers. For multi-physics optimization where each physics is evaluated by a closed computer-aided engineering package that can only be evaluated in sequence, BBTO has less difficulty in integration and coordination, although combined solver packages could solve some of these issues with GTO methods.

Gradient-free approaches, in particular, the algorithms relying on large-scale data generation and data mining, and involving human intelligence in the design loop, will potentially play a central role in designing the next-generation topology optimization algorithms for multiphysics applications. Instead of working on the optimization of materials or structures directly to achieve multiple physical performances, these algorithms rely on generating samples in a performance space spanned by the different physical properties. It requires running large-scale simulations on parallel computing platforms to obtain these physical properties in the precomputation stage. These features can be correlated, e.g., Young’s modulus and Poisson’s ratio, or independent, e.g., material stiffness and electrical conductivity. The trade-offs between these properties can be explored straightforwardly on the boundary of the low-dimensional manifold in the performance space. A typical example of this type of computational pipeline can be seen in [307], where multiple physical properties are explored as independent axes of the space in the data generation phase and combined to explore the trade-offs in the optimization stage. Another direction to explore these multiphysics trade-offs is to develop interactive visualization tools [308] for large-scale precomputed data. These tools provide an efficient way to incorporate human designers into the optimization loop for making decisions between multiple design objectives.

D. Design-dependent physics

Within the multiphysics applications framework, a noteworthy group of problems can be denominated as design-dependent. In such problems, the structure is interacting with one or more distinct physical fields through a well-defined interface, when in the presence of acoustic waves, fluid flow, etc., or a volume, when associated with electromagnetics, temperature variation, etc. The problem is said to be design-dependent only when the loads on the structure can change their location, direction, and magnitude during optimization due to modifications in the multiphysics interfaces or volumes.

Design-dependent problems represent a challenge for topology optimization. For example, in thermoelastic design, thermal stresses are dependent on the structural design. The current methods lead to ill-behaved optimization problems and might produce solutions that are opposed to design practices related to thermal stresses [1], [309]. Thermoelastic problems, although design-dependent, can be treated as a single physics problem when the temperature field or variation is known. Other design-dependent problems present inherent interaction between two or more physics, e.g., fluid-structure interaction (FSI). In these problems, fluid pressure and/or viscous loads are dependent on the structural boundary location and the optimization method must explicitly track the fluid-structure interfaces in order to account for accurate equilibrium. Although optimization algorithms with discrete design variables and explicitly defined boundaries, e.g., in hydrostatic loaded
structures [310], acoustic-structure interaction [311], [312] and FSI with viscous fluid flow [313], [314] have used to design such systems, the potential of BBTO is yet to be explored.

VII. CONCLUSION

In this paper, we presented topology optimization as a black-box optimization problem. The problem is mostly studied in the literature based on the traditional binary-coded formulation for minimum compliance. However, the subject of problem formulation is rich in innovation and different formulations have been proposed to diminish the design space for more efficient and effective optimization. Having said that, the design representation should be formulated considering topological attainability and how the topological details affect optimality of the achievable solutions.

The BBTO can be considered as an application of large-scale global optimization. However, the problem is mostly studied using standard evolutionary algorithms that may not perform efficiently with high dimensionality. To advance this topic of research, it is crucial to develop global search algorithms that are capable to deal efficiently with high-dimensional complex problems. We provided some examples of algorithms and techniques that can be utilized in BBTO methods.

Furthermore, we presented various techniques to speed up computation at software and hardware levels. The increasing trend in using multi-core processors, in personal and portable computers, makes the issue of computational cost less significant than before. BBTO can be considered as a complement to topology optimization, filling the gap in the applicability of GTO for new or non-conventional problems. Thus, we reviewed some example applications including crashworthiness design, design for manufacturability, multiphysics, and design dependent physics. Finally, we can conclude that although recent advancements succeeded to overcome major challenges of BBTO, the existing literature is still long far from fully utilizing all advances in evolutionary computation and high-performance computing. More empirical studies using the presented methods are needed to increase the community’s knowledge on their best usage.

ACKNOWLEDGMENT

Renato Picelli acknowledges the financial support by the São Paulo Research Foundation (FAPESP), grant number [2018/05797-8]. Carlos A. Coello Coello gratefully acknowledges support from CONACyT project no. 2016-01-1920.

REFERENCES


[304] Y. Zhou et al., “Gradient-based multi-component topology optimization for stamped sheet metal assemblies (MTO-III) as a scientist in the Complex System Optimization and Analysis group where he researched combinations of topology optimization with evolutionary algorithms and machine learning. He completed his industrial doctorate in cooperation with the Technische Universität Darmstadt in 2017. From 2017 to 2018 as a senior researcher with the Technische Universität Darmstadt in 2017. From 2017 to 2018 as a senior researcher with
Renato Picelli is a Research Fellow at the Polytechnic School of the University of São Paulo. He received his PhD on "Evolutionary Topology Optimization of Fluid-structure Interaction Problems" from the University of Campinas, Brazil. His PhD work is acknowledged as one of the three existent methodologies for gradient-based topology optimization of design-dependent fluid- and acoustic-structure interaction problems. He was a visiting researcher at Delft University of Technology (TU Delft) in 2014, Royal Melbourne Institute of Technology (RMIT) in 2015, University of California San Diego (UCSD), in 2016 and a post-doctoral researcher at Cardiff University, in the UK, for three years.

Bo Zhu is an assistant professor in the Computer Science Department at Dartmouth College. He received his Ph.D. in Computer Science from Stanford University. He also worked as a postdoc researcher at MIT CSAIL. Bo’s research interests encompass computer graphics, computational physics, and computational fabrication. In particular, he focuses on building computational approaches to automate the process of exploring complex physical systems.

Yuqing Zhou received the B.S. degree in mechanical engineering from Northeastern University, Shenyang, China, in 2012, and the M.S. and Ph.D. degrees in mechanical engineering from University of Michigan, Ann Arbor, MI, USA, in 2014 and 2018, respectively. He is currently a Research Scientist with the Toyota Research Institute of North America. His current research interest is topology optimization.

William Vicente is an assistant professor in structural dynamics and computational modeling at the School of Agricultural Engineering at University of Campinas. His current research interests include apply the finite element method and the structural topology optimization procedures to obtain innovative and efficient designs of implements/structures and materials that can be applied in the agricultural engineering.

Francesco Iorio received his M.Sc. in Computer Science (Information Technology) from the University of Liverpool, UK in 2008. He is currently working on his Ph.D. in computer science from the University of Toronto, Canada. In 2004 he joined Alias—Wavefront where he worked on scalable, high-performance 3D computer graphics for advanced visualization. In 2007 he joined the IBM High Performance Computing Group in Dublin, where he was the lead solution architect for all Cell Broadband Engine processor (Cell/B.E.) related projects in the fields of scientific computing, finance and high-throughput data analysis. In 2009 he joined Autodesk Research in Toronto, Canada, where he is Director of Computational Science Research and Distinguished Scientist. His research focuses on mathematical modeling, machine learning, creative AI and general design and exploitation of parallel algorithms on a variety of high-performance computing platforms. He has authored and co-authored more than 20 patents and publications.

Markus Olhofer is Chief Scientist at the Honda Research Institute Europe since 2010 and responsible for the Complex Systems Optimisation and Analysis Group of the institute. Markus Olhofer graduated in Electrical Engineering with a thesis on face detection and image recognition with a Diplom Ingenieur degree in 1997. After that he worked as PhD student at the Institute for Neural Computation (Institut für Neuroinformatik), Ruhr-Universität Bochum and received his PhD in 2000 from that university. He joined the Future Technology Research Division at Honda R&D Europe (Deutschland) GmbH in 1998 and since 2001 he works at the Honda Research Institute Europe GmbH.

Wojciech Matusik is an Associate Professor of Electrical Engineering and Computer Science at the Computer Science and Artificial Intelligence Laboratory at MIT, where he leads the Computational Fabrication Group. Before coming to MIT, he worked at Mitsubishi Electric Research Laboratories, Adobe Systems, and Disney Research Zurich. He studied computer graphics at MIT and received his PhD in 2003. He also received a BS in EECS from the University of California at Berkeley in 1997 and MS in EECS from MIT in 2001. His research interests are in direct digital manufacturing and computer graphics. In 2004, he was named one of the worlds top 100 young innovators by MITs Technology Review Magazine. In 2009, he received the Significant New Researcher Award from ACM Siggraph. In 2012, Matusik received the DARPA Young Faculty Award and he was named a Sloan Research Fellow.

Carlos Artemio Coello Coello (M’98–SM’04–F’11) received Ph.D. degree in computer science from Tulane University, New Orleans, LA, USA, in 1996. He is a Professor (CINVESTAV-3F Researcher) with the Department of Computer Science of CINVESTAV-IPN, Mexico City, México. He has authored and co-authored over 450 technical papers and book chapters. He has also co-authored the book “Evolutionary Algorithms for Solving Multi-Objective Problems” (Second Edition, Springer, 2007). His publications currently report over 48,900 citations in Google Scholar (h-index is 80). His research interests include evolutionary multi-objective optimization and constraint-handling techniques for evolutionary algorithms. He is a recipient of the 2007 National Research Award from the Mexican Academy of Sciences in the area of Exact Sciences, the 2013 IEEE Kiyo Tomiyasu Award and the 2012 National Medal of Science and Arts in the area of Physical, Mathematical and Natural Sciences. He is a member of the Association for Computing Machinery and the Mexican Academy of Science.

Kazuhiro Saitou received the B.Eng. degree in mechanical engineering from the University of Tokyo, Tokyo, Japan, in 1990, and the M.S. and Ph.D. degrees in mechanical engineering from Massachusetts Institute of Technology, Cambridge, MA, USA, in 1992 and 1996, respectively. He is currently a Professor in the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA. His research interest is algorithmic and optimal synthesis of products and systems, computational design for manufacture and assembly, and manufacturability-driven, multi-component structural topology optimization.