Proceedings of the ASME 2020
International Design Engineering Technical Conferences
and Computers and Information in Engineering Conferences
IDETC/CIE2020
August 16-19, 2020, St. Louis, MO, USA

DETC2020-22681

METASET: AN AUTOMATED DATA SELECTION METHOD FOR SCALABLE DATA-DRIVEN DESIGN OF METAMATERIALS

Yu-Chin Chan, Faez Ahmed, Liwei Wang, Wei Chen*
Dept. of Mechanical Engineering
Northwestern University
Evanston, Illinois, USA

ABSTRACT

Data-driven design of mechanical metamaterials is an increasingly popular method to combat costly physical simulations and immense, often intractable, geometrical design spaces. Using a precomputed dataset of unit cells, a multiscale structure can be quickly filled via combinatorial search algorithms, and machine learning models can be trained to accelerate the process. However, the dependence on data induces a unique challenge: An imbalanced dataset containing more of certain shapes or physical properties than others can be detrimental to the efficacy of the approaches and any models built on those sets. In answer, we posit that a smaller yet diverse set of unit cells leads to scalable search and unbiased learning. To select such subsets, we propose METASET, a methodology that 1) uses similarity metrics and positive semi-definite kernels to jointly measure the closeness of unit cells in both shape and property space, and 2) incorporates Determinantal Point Processes for efficient subset selection. Moreover, METASET allows the trade-off between shape and property diversity so that subsets can be tuned for various applications. Through the design of 2D metamaterials with target displacement profiles, we demonstrate that smaller, diverse subsets can indeed improve the search process as well as structural performance. We also apply METASET to eliminate inherent overlaps in a dataset of 3D unit cells created with symmetry rules, distilling it down to the most unique families. Our diverse subsets are provided publicly for use by any designer.

*Address all correspondence to this author. Email: weichen@northwestern.edu

1 Introduction

Metamaterials are drawing increased attention for their ability to achieve a variety of non-intuitive properties that stem from their intentionally hierarchical structures [1]. While they traditionally consist of one unit cell that is repeated everywhere, multiple unit cells can also be assembled to create *aperiodic* mechanical metamaterials with, *e.g.*, spatially-varying or functionally-gradient properties [1, 2]. Over the past few years, conventional computational methods have been adapted to design these complex structures, including topology optimization (TO) of the microscale unit cells within a fixed macroscale structure [3, 4], and hierarchical and concurrent multiscale TO that design both the unit cell and macro-structure [5–7]. However, as the desire to attain even more intricate behaviors grows, so too does the complexity of the design process, which must account for the expensive physical simulations and, in aperiodic structures, the vast combinatorial design space and disconnected neighboring unit cells [1, 8].

Capitalizing on advances in computing power, data-driven metamaterials design can be a more efficient and therefore enticing solution to those challenges. Its success hinges on precomputed unit cell libraries or datasets, which can avoid costly on-the-fly physical simulations and multiscale TO in huge design spaces, as well as provide candidate unit cells that are better connected to their neighbors. Fig. 1 shows an overview of two common approaches in data-driven design: global optimization methods, and machine learning (ML) based methods. In the first case, combinatorial optimization algorithms can be

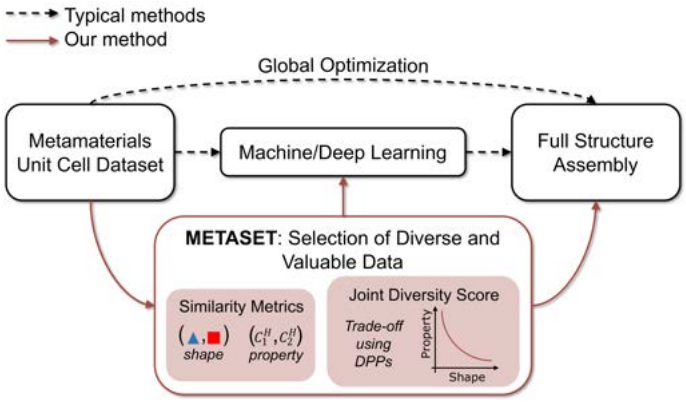


FIGURE 1: A high-level overview of data-driven metamaterials design, and how our proposed method, METASET, fits in. As an example, we show C^H, the homogenized elastic tensor, as the unit cell properties.

used to directly search for the set of unit cells that realize a target macroscale behavior while minimizing or constraining the boundary mismatch between neighboring cells [1, 8–10]. From another perspective, data-driven methods can use the dataset to train ML models that further accelerate design. For example, they have been used to rapidly predict homogenized physical properties as part of the optimization loop [11–14]. Additionally, deep generative models inspired by the computer vision field can learn embedded geometric descriptors that act as reduced dimensional design variables, and construct new designs, *e.g.*, optical 2D metamaterials [15, 16], almost instantaneously.

The efficacy of data-driven methods, however, relies highly on the size and coverage of the datasets. The search space of global optimization methods can quickly explode when the number of unit cells increases. Meanwhile, imbalanced datasets with skewed data distributions can reduce the chance of meeting certain property or compatibility requirements, and hobble the performance of ML models since they may not learn a less frequent property or shape as well [17]. Therefore, due to the importance of the data on downstream tasks, in this work we focus on the first step of data-driven design: dataset selection.

In existing literature, metamaterial datasets are often built using heuristics or the designer's intuition, with the assumption that the unit cells will offer sufficient coverage for the desired application. Many employ TO to inversely design unit cells that meet pre-specified target properties [1, 9, 11], and some expand the dataset by morphing the shapes [1, 11] or randomly flipping pixels or voxels [9]. Alternatively, Panetta *et al.* developed graph-based rules to create truss-like unit cells [18]. Although these are more feasible than enumerating over all possibilities, bias towards particular properties or shapes can be unintentionally introduced, deteriorating the performance of the design algorithm or the design itself.

Moreover, the point at which to stop generating new unit cells has thus far been heuristic with the same goal in mind: to cover a broad property space. The range of this space is sometimes restricted for specific applications [3], or strict symmetry and manufacturability constraints are implemented to limit the possible shapes [18]. More often, the property space is allowed to grow at will, *e.g.*, TO and shape perturbation are repeated until the change in the density of the property space is less than a given tolerance [9, 11]. While efficient, all of the works to date have only considered coverage in the property space alone, which can produce similar shapes or overlook those that might benefit the design with regards to boundary connectivity. In contrast, our work explores coverage in both property and shape space.

Improving imbalance arising from data with multiple classes has been extensively researched in computer science. The most relevant to our application are the data-preprocessing strategies such as undersampling to remove data from majority classes, oversampling to replicate data from minority classes, or combinations thereof [17]. However, the former can accidentally remove samples with important features, *i.e.*, decrease the diversity, and the latter can lead to model overfitting and increased training overhead [19]. Nor are they made to consider the diversity of data with features that have drastically different representations, like shape and property. The issue of downsampling a metamaterial database was addressed by Chen *et al.* [14], who compressed the size of their database by selecting the samples that are farthest from each other with respect to properties (not shape), allowing them to more efficiently fit a property prediction model. As far as we know, there is currently no method to assess or select a diverse set of unit cells that can simultaneously cover the shape and property spaces.

Despite the dearth in the metamaterials field, measuring and ranking items based on their quality as well as their contribution to the diversity of a whole set or subset is an ongoing research area. In computer science, for example, recommender systems rank diverse items such as online products to match users' preferences. These are based on the concept of diminishing marginal utility [20], wherein lower ranking items bestow less additional value onto the users. In design, too, researchers have developed methods to help designers sift through large sets of ideas by ranking them. In particular, to balance diversity against quality of designs, Ahmed *et al.* introduced the idea of clustering items into groups for subset selection [21], which employed submodular functions that follow the property of diminishing marginal utility. Ahmed *et al.* [22] also showed the application of Determinantal Point Processes [23], which model the likelihood of selecting a subset of diverse items as the determinant of a kernel matrix, to the diverse ranking task. The latter, in particular, are elegant probabilistic models that capture the trade-off between competing ideas like quality and diversity. While the goal of maximizing the determinant is similar to the optimality criterion used in generating D-optimal designs [24] in design of experiments, Determinantal Point Processes are not restricted to linear

kernels and have advantages in that calculating marginals, computing certain conditional probabilities, and sampling can all be done in polynomial time. This paper shows that it can also be used for coverage in multiple spaces defined over the shapes and properties of unit cells.

Our contributions: We propose METASET, an automated methodology that simultaneously considers the diversity of shape and property to select subsets of unit cells from an existing dataset. By doing so, we can achieve scalable data-driven design of metamaterials by using smaller yet diverse subsets. In addition, we can eliminate bias in imbalanced datasets through diversification to improve any ML model trained on these sets. As a part of METASET, we also introduce using similarity metrics to efficiently assess the diversity of the shapes and properties of 2D and 3D metamaterials. We propose that a weighted sum of Determinantal Point Process (DPP) kernels based on the shape and property similarities can measure and allow the maximization of the joint diversity with respect to both shape and property spaces. For the first time in data-driven metamaterials design — to our knowledge — using 2D design as example, we reveal that diverse subsets can expedite and even enhance the design performance and compatibility of aperiodic mechanical metamaterials compared to using the full dataset. Finally, applying METASET to the generation of 3D unit cells, we identify diverse families of isosurface unit cells and discover that these extend beyond the ones commonly considered in the design of functionally-graded structures [2, 25]. The diverse subsets are available publicly.[1]

Organization of the paper: Our methods are detailed in Sec. 2. As case studies and validation, we apply METASET to diversify large datasets of 2D and 3D unit cells. We use diverse 2D subsets to design metamaterials with non-intuitive target displacement profiles (Sec. 3.1). In the 3D case, we use METASET to provide unique and diverse families of isosurface unit cells (Sec. 3.2).

2 METASET: Assessing and Optimizing Diversity

The inner workings of METASET consist of three main steps: 1) Defining similarity metrics for metamaterials that quantify the difference between pairs of 2D or 3D shapes as well as their properties (Sec. 2.1); 2) Using a DPP-based submodular objective function to measure the joint coverage of a set of unit cells in shape and property spaces using pairwise similarity kernel matrices (Sec. 2.2); 3) Maximizing the joint diversity with an efficient greedy algorithm while allowing trade-off in the two spaces to be tuned to suit the desired application (Sec. 2.3). In this section, we describe these components and summarize the METASET methodology with Algorithm 1.

[1] https://github.com/lychan110/metaset

2.1 Similarity Metrics for Metamaterials

A diverse metamaterials dataset should ideally contain unit cells that are sufficiently different, *i.e.*, not similar, such that they cover the shape and property spaces. To measure the diversity of a set, then, we first need to quantify the similarity between the properties and shapes of unit cells.

Property Similarity Since mechanical properties are generally scalar values that can be formatted as a vector, *e.g.*, by flattening the elastic tensor, we can use any similarity metric between vectors. In this work, we use the Euclidean distance. We note that the properties do not need to be the tensor components; rather, they can be other values of interest such as elastic or shear moduli, or Poisson's ratios. Neither do they need to be limited to scalar mechanical properties. For instance, acoustic bandgap that are dynamic properties could be considered as long as the similarity between two of them can be quantified.

Shape Similarity Shape similarity metrics are key in many computer vision and graphics applications, *e.g.*, object retrieval from databases and facial recognition. In these methods, the shapes are usually first represented by structural descriptors extracted from individual shapes [26], or by embedded features learned via data-driven methods such as clustering or ML [27]. The distances between features are then measured in either Euclidean [26] or Riemmanian space [28, 29]. Since Riemannian metrics are based on geodesic distances, they are suitable if one needs invariance to deformation, *i.e.*, if one considers a shape to be the same after bending.

For metamaterials, however, we must rule out deformation and rotation invariant metrics since any transformation of a unit cell impacts its properties. Additionally, to ensure that METASET is compatible with small or imbalanced datasets, we avoid using embedded features. We also seek metrics that are efficient — particularly in 3D — but still detailed enough to discriminate fine details, as well as able to form positive semi-definite similarity matrices for the later step involving DPPs. Hence, in the next sections, we introduce Euclidean metrics based on structural features, namely, division-point-based descriptors for 2D and Hausdorff distance for 3D.

2D Shape Similarity Metric For 2D unit cells, which are usually binary images resulting from TO, there are multiple methods to compute similarities between them. We propose using a descriptor-based approach by first extracting a division-point-based descriptor [30] to reduce the dimension of the image into a vector that extracts salient features at different levels of granularity. This has been applied to the field of optical character recognition [31, 32]. The binary image of a unit cell is first recursively divided into sub-regions that contain an equal number of solid pixels. The coordinates of all division points, *i.e.*, points

at the intersection of two division lines between each sub-region, are then obtained as descriptors of the unit cell. This process is repeated until the desired level of detail is captured, constructing a k-d tree of the distribution of solid materials. In the later 2D case study (Sec. 3.1), we perform the division seven times for each unit cell, which in our case extracts a sufficient amount of detail without needing a very high-dimensional descriptor. This results in the coordinates of 62 division points that constitute a 124-dimensional vector as the shape descriptor.

Using the above method, we can represent each 2D unit cell with a vector, then use the Euclidean norm to find the distance between any pair. However, the input for a DPP is a positive semi-definite similarity matrix, L, so we transform the distance to a similarity metric by selecting from a wide variety of transformations. For the 2D case, we use a radial basis function (RBF) kernel with unit bandwidth, *i.e.*, $L_{i,j} = \exp(-0.5\, d(i,j)^2)$, where $d(i,j)$ is the distance between i-th and j-th unit cells. In practice, the choice of an appropriate transformation is equivalent to choosing the right distance metric between items. Our empirical study showed that different standard choices (like the ones we used) mainly affect the distribution of similarity values but do not significantly affect the final outcome or the key findings of our work.

3D Shape Similarity Metric As for 3D unit cells, mesh formats such as STL are commonly used so that researchers can manufacture the metamaterials through additive manufacturing. However, since performing analysis on 3D shapes is undoubtedly more computationally intense, we suggest representing each unit cell as points on the surface of the original mesh, *i.e.*, point clouds, which are more efficient for extracting and processing 3D features [33]. This extra conversion can take little computation with well-established sampling methods, *e.g.*, randomly sampling the surface of a mesh with the probability of choosing a point weighted by the area of the triangular faces.

We then use a distance metric commonly utilized to measure the distance between sets of points, the Hausdorff distance. In essence, it computes the difference between two clouds as the maximum of the nearest neighbor distances of each point. This is expressed as [34]:

$$h(A,B) = \max_{a \in A} \left[\min_{b \in B} \|\cdot\| \right], \tag{1}$$

where a is a point (x,y,z) within cloud A and b is a point in the second cloud B. The notation $\|\cdot\|$ indicates that any distance can be used; for metamaterials, we use the Euclidean norm, $\|a - b\|_2$. In our implementation, we computed the nearest neighbor norms using a GPU-enabled code by Fan *et al.* [35]. To obtain a symmetric distance, we can take the maximum as follows:

$$d_{\mathrm{H}}(A,B) = d_{\mathrm{H}}(B,A) = \max\left[h(A,B), h(B,A)\right], \tag{2}$$

Finally, we convert the distances into a DPP similarity kernel, L, using the following transformation: $L_{ij} = \frac{1}{1+d(i,j)}$.

2.2 Determinantal Point Processes for Joint Diversity in Two Spaces

Given a similarity kernel matrix L, we can now measure the diversity of a dataset using Determinantal Point Processes (DPPs), which are models of the likelihood of choosing a diverse set of items. They have been used for set selection in ML, *e.g.*, diverse pose detection and information retrieval [23, 36], and recently in ranking design ideas based on diversity and quality [22]. Viewed as joint distributions over the binary variables that indicate item selection, DPPs capture negative correlations. This means that, intuitively, the determinant of L is related to the volume that the set covers in a continuous space. In other words, the larger the determinant, the more diverse the set.

To model our data, we construct DPPs through L-ensembles [37], using a positive semi-definite matrix, L, to define a DPP. Hence, given the full unit cells dataset of size N, which we denote as ground set G, DPPs allow us to find the probability of selecting any possible subset of unit cells. For any subset M, it assigns the following probability:

$$\mathbb{P}(M) = \frac{det(L_M)}{det(L+I)}, \tag{3}$$

where $L_M \equiv [L_{ij}]_{ij \in M}$ is the submatrix of L with entries indexed by elements of the subset M, and I is a $N \times N$ identity matrix. The probability of a set containing two items increases as the similarity between them decreases. Therefore, the most diverse subset of any size has the maximum likelihood $\mathbb{P}(M)$, *i.e.*, the largest determinant.

Unlike submodular clustering approaches, DPPs only require the similarity kernel matrix, L, as an input and do not explicitly need the data to be clustered or to define a function to model diversity. This also makes them more flexible, since we only need to provide a valid similarity kernel, rather than an underlying Euclidean space or clusters.

For METASET, we calculate two different similarity values — one in shape space and another in property space — between any two unit cells. Hence, for all the unit cells combined, we have one kernel matrix corresponding to each of the two spaces. In order to measure the joint coverage in both spaces, we take a weighted sum of the two matrices, thus allowing the trade-off between diversifying in shape or property space:

$$L = (1-w) \cdot L_P + w \cdot L_S, \tag{4}$$

where L, L_P and L_S are the joint, property and shape similarity kernels, respectively, and w is a weight parameter than can be varied between 0 and 1. It is possible to combine two kernel matrices in many ways. However, we choose this formulation due to two reasons. First, the weighted sum of two positive semi-definite matrices is also positive semi-definite, which is a prerequisite for a DPP kernel. Second, it allows us to control the amount of diversity in both spaces using a single knob, w. We conducted multiple experiments on simulated data with easy-to-

verify coverage metrics and found that this approach is effective in capturing diversity in both spaces. In the interest of space, we have not included these experiments and directly report results on using joint kernels for more complex 2D and 3D metamaterial unit cells.

2.3 Algorithm for Optimizing Diversity

To optimize the diversity of a subset M, we maximize the log determinant of its joint similarity kernel, *i.e.*, $f = \log[det(L_M)]$. Note that the log determinant of a positive semi-definite matrix is monotonically non-decreasing and submodular. In general, finding the set of items that maximizes a submodular diversity function is NP-Hard. When solving such problems, a well-known limit due to Feige [38] is that any polynomial-time algorithm can only approximate the solution up to $1 - \frac{1}{e} \approx 67\%$ of the optimal.

However, this is where choosing a submodular function f as the objective comes in handy. It turns out that greedily maximizing this function is guaranteed to achieve the optimality bound. We use this property to substantially accelerate diversity optimization using a scalable greedy algorithm [39], which has theoretical approximation guarantees and is widely used in practice. At each step, the algorithm picks an element, *i.e.*, a unit cell, that provides the maximum gain in the objective function (lines 5-8 in Algorithm 1). This makes greedy maximization of diversity the best possible polynomial-time approximation to an otherwise NP-Hard problem.

Algorithm 1: METASET algorithm. After calculating the similarity kernels, a polynomial-time greedy maximization of the gain on the weighted combination between diversity in shape and property space is performed. The output is a subset of all unit cells such that the joint diversity is maximized.

Data: Ground set G of size N of all unit cells
Result: Subset M of size N_M
1 Calculate shape and property similarity kernels, L_S and L_P;
2 Calculate joint similarity kernel L;
3 Find subset M;
4 $M \leftarrow \emptyset$;
5 **while** $|M| \neq N_M$ **do**
6 Pick an item G_i that maximizes $\delta f(M \cup i)$;
7 $M = M \cup \{G_i\}$;
8 $G = G - G_i$;
9 **return** M;
10 Use M as input to downstream task such as data-driven design or machine learning;

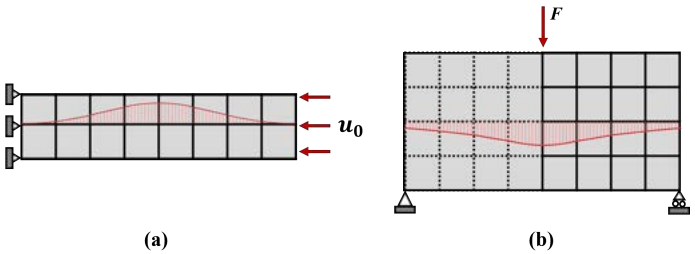


FIGURE 2: Problem settings of the 2D design examples, (a) the cantilever, and (b) the MBB beam, both of which should achieve the target displacement profiles shown in red.

3 METASET in Data-Driven Metamaterials Design

Now that we have set the stage for METASET, we demonstrate through 2D and 3D mechanical metamaterials case studies the advantages of selecting diverse subsets of metamaterial unit cells based on their shape and properties. The most salient of these are the abilities to accelerate search algorithms while enhancing the final design performance (Sec. 3.1), and discover unique unit cell families in order to build an economical and diverse dataset for design (Sec. 3.2).

3.1 Application to 2D Aperiodic Mechanical Metamaterials

Inserting METASET prior to the assembly stage in the general data-driven design flow (Fig. 1) can augment the performance and results of global search algorithms. In this section, we show this improvement through 2D data-driven design examples featuring mechanical metamaterials that are given target displacement profiles and constraints on the connectivity of neighboring unit cells. First, to emphasize that METASET can be added with little extra cost to any existing data-driven framework, we utilize a 2D dataset of unit cells from our previous work in [11, 12] (which we will briefly describe in Sec. 3.1.1). Next, we use METASET to select several subsets that illustrate the effect of size and diversity on the search and final designs (Sec. 3.1.2). Finally, we employ genetic algorithm to assemble the full structures to meet our design requirements (Sec. 3.1.3).

The first design example is a cantilever discretized into 2×4 unit cells, as shown in Fig. 2a. Here the design objective is to optimize the combination of unit cells so that the horizontal centerline of the cantilever achieves a bridge-like shape (red curve) when a prescribed displacement, $\mathbf{u}_0$, is applied to its right end. In the second example, we design a classical example from the TO field, the MBB beam, such that its horizontal centerline conforms to the red curve when loaded with a vertical force F (Fig. 2b). Due to the structural symmetry, we only need to design the right half of the beam with 4×4 unit cells, outlined by the solid black lines. The full structure can then be obtained by symmetry, *i.e.*, reflecting over the vertical centerline.

3.1.1 2D Unit Cells via Topology Optimization and Perturbation

In [11, 12], we previously proposed using a combination of TO and stochastic shape perturbation to generate a large dataset of 2D unit cells. To initialize the dataset, we ran TO for each uniformly sampled target properties, the components of homogenized elastic tensors, and then iteratively perturbed the shape of the unit cells with the most extreme or uncommon properties. By doing so, we created a dataset of 88,000 unit cells that covered a relatively large property space within reasonable computational cost. Note that we did not build this dataset with geometry in mind, leading to many similar shapes. Also, even though we aimed to fill the less populated regions of the property space by perturbing unit cells in those locations, there is a higher concentration of final unit cells with lower property values (the lower left corners in Fig. 4). For details, please see [11].

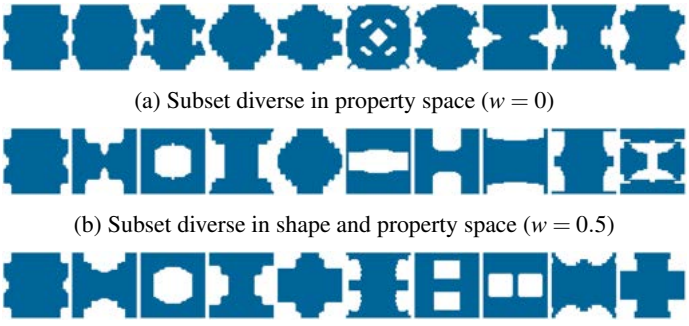
Before applying METASET, we preprocess the data by randomly sampling unit cells from the original dataset of size 88,000. We select 17,380 unit cells that have a volume fraction greater than 0.70. This fraction was chosen so that the resultant unit cells are less likely to have very thin features, which makes them more feasible for manufacturing. Additionally, when computing shape diversity, if unit cells occupy very different volume fractions, a diverse subset is more likely to be dominated by flimsy low volume structures, which have the least probability of overlap with other unit cells (high volume fraction unit cells are more probable to have some overlap with each other). However, as we will show in Sec. 3.1.3, this preprocessing does not impede the chances of designing well-connected structures that met the targets quite well.

3.1.2 METASET Results

For the dataset of 17,380 2D unit cells, which we now refer to as the full or ground set G, we calculate the property and shape similarity matrices, L_P and L_S, respectively, as described in Sec. 2.1. Taking their weighted sum forms the joint DPP kernel matrix L (Eq. 4), whose determinant scores the diversity in terms of both shape and property (Sec. 2.2). To explore this, we rank several subsets using the greedy algorithm from Sec. 2.3 by varying their sizes, N_M, and kernel weights, w. From the results, we can make two observations:

1. By increasing w, we shift from ranking a subset based on diversity in the property space alone, to a mixture of both spaces, and to the shape space only. In essence, trade-off between shape and property diversity can be easily controlled.
2. By observing $det(L_M)$ of the subsets as more items, *i.e.* unit cells, are added, we find that the gains in shape and property diversity saturate at approximately $N_M = 20$. Thus, a very small number of unit cells are sufficient to cover both spaces.

Ten example unit cells from the subsets with $w \in \{0, 0.5, 1\}$ are shown in Fig. 3, where the subset optimized for only shape diversity (Fig. 3c) displays the most variety of topologies compared to the subset diverse in only properties (Fig. 3a). Meanwhile, the balanced subset contains a mixture of unit cells akin to both extreme sets (Fig. 3b). This may be counter-intuitive since similar shapes should have similar mechanical properties. However, note that upon close inspection, the property diverse unit cells exhibit tiny features that lead to low elastic property values. Such small details in the shape may lead to a larger change according to the physical simulations and the property similarity metric, *i.e.*, the Euclidean norm.



(a) Subset diverse in property space ($w = 0$)

(b) Subset diverse in shape and property space ($w = 0.5$)

(c) Subset diverse in shape space ($w = 1$)

FIGURE 3: Examples of 2D unit cells from the diverse subsets used in the cantilever and MBB design problems.

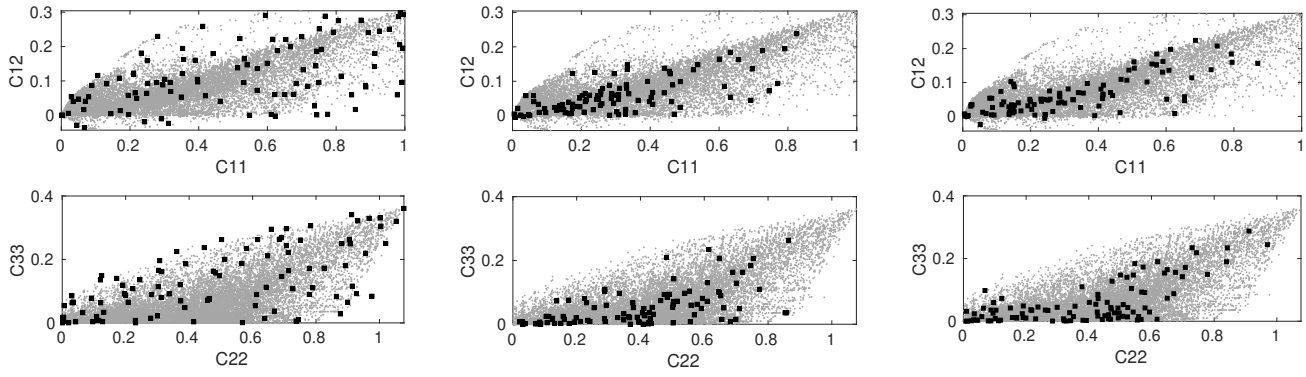
Comparing the properties of the unit cells in diverse subsets to the ground and randomly sampled sets (Fig. 4), we can confirm that the property diverse subsets cover all regions of the original property space, even the sparsely populated areas. As expected, the shape diverse subset does not do as well, and the random subset contains tight clusters in certain areas.

The result that only 20 unit cells is needed to cover the shape and property spaces is quite interesting since a main tenet of data-driven design thus far is that "more is better" — larger datasets provide more candidates from which we can choose compatible unit cells. So, to explore the impact of the subset size on the data-driven approach, we selected the top 20 as well as top 100 ranking unit cells from each subset to move on to the next step: full structure assembly.

3.1.3 Design Results: A Study on the Effect of Size and Diversity

Using subsets of unit cells with varying sizes and levels of diversity for metamaterials design using global optimization, we can elucidate 1) the effect of subset size on the search algorithm's efficiency, and 2) the impact of diversity on the final design performance as well as the compatibility of neighboring unit cells. We choose the following nine datasets:

- P_{20}: Property diverse subset of size 20
- SP_{20}: Shape and property diverse subset of size 20
- S_{20}: Shape diverse subset of size 20 diverse
- R_{20}: Random subset of size 20
- P_{100}: Property diverse subset of size 100



(a) Property DPP (b) Random Sample (c) Shape DPP

FIGURE 4: The property space of the 2D unit cell subsets optimized for property and shape diversity, and a randomly sampled set, plotted against the full dataset. We observe that property DPPs cover the space well, hence it is more likely to have unit cells near any target property combination.

- SP_{100}: Shape and property diverse subset of size 100
- S_{100}: Shape diverse subset of size 100
- R_{100}: Random subset of size 100
- G: Full dataset of size 17,380

To design the cantilever and MBB beam described earlier (Fig. 2), we pass each of the datasets to a global optimization method, which in these examples is a a single objective genetic algorithm to better illustrate the effect of size and diversity on the final results. Although this approach is simple, we chose it to restrict our design to the discrete choice of unit cells in our subsets, whereas most gradient-based algorithms for data-driven metamaterials design map continuous design variables to the nearest existing, or interpolated, unit cell in dense databases [1, 9].

Specifically, the genetic algorithm is used to choose the combination of unit cells from each given dataset that minimizes the mean squared error (MSE) between the achieved and target displacement profiles. Also, since detached neighbours are not desirable, we add a compatibility constraint by requiring that the number of disconnected unit cells, N_{dc}, in the full structure be equal to zero. The optimization problem is formulated as:

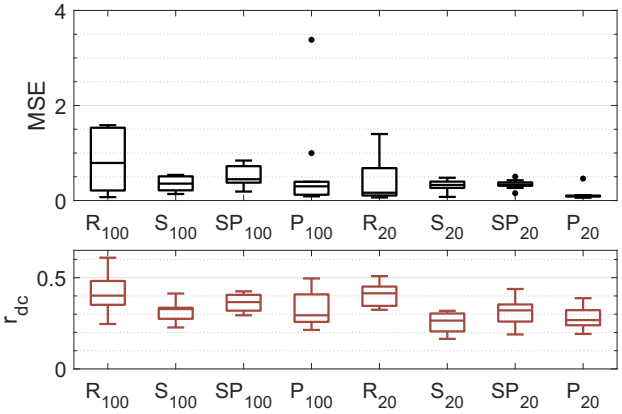
$$\begin{aligned} \underset{\boldsymbol{l}}{\text{minimize}} \quad & \frac{1}{n}\|\boldsymbol{u}(\boldsymbol{l}) - \boldsymbol{u}_t\|_2^2 \\ \text{subject to} \quad & \boldsymbol{K}(\boldsymbol{l})\boldsymbol{U} = \boldsymbol{F}, \\ & N_{dc}(\boldsymbol{l}) = 0, \\ & l_i \in \{1,2,\cdots,N_M\}, \quad i = 1,2,\ldots,N_f, \end{aligned} \tag{5}$$

where $\boldsymbol{u}$ is the displacement of n nodes located on the centerline of the structure, $\boldsymbol{u}_t$ is the discretized target displacements, $\boldsymbol{K}$ is the global stiffness matrix, and $\boldsymbol{U}$ and $\boldsymbol{F}$ are global displacement and loading vectors, respectively. The number of unit cells in the given dataset is N_M while the number in the full structure is N_f, and $\boldsymbol{l} = [l_1, l_2, \ldots, l_{N_f}]^T$ is a vector of the indices of the chosen unit cells.

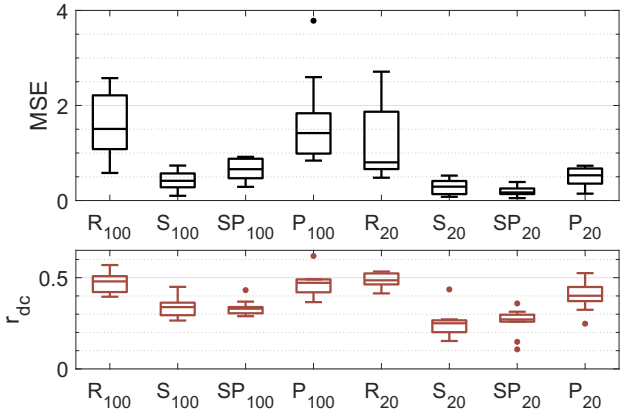
Due to the stochasticity of genetic algorithms, we run the optimization ten times for each dataset and report the MSE of the final topologies in Fig. 5. In addition, we show a measure of the connectivity of the final structure: the mean ratio of disconnected pixels on the boundaries of touching unit cells, r_{dc}. Similar to N_{dc} in the constraint (Eq. 5), a fully compatible structure should have r_{dc} as zero.

When given the full dataset, G, the search algorithm is overwhelmed and not able to find any designs with satisfactory MSE for either example (Table 1). In two runs for the cantilever and one for the MBB beam, it also fails to meet the compatibility constraint. This can be attributed to a vast search space since the number of possible unit cell combinations grows exponentially as the size of the dataset increases. A larger set may also contain more redundant shapes or properties that contribute little to diversity, exacerbating the search challenge and possibility of local optima.

Conversely, every run using the 20- and 100-item subsets satisfy the constraint. In fact, our results (Fig. 5) show that smaller, diverse subsets can perform better than all other sets under the same search algorithm and termination criteria. In the cantilever example, the smallest property diverse subset P_{20} achieves the lowest mean MSE, followed by S_{20} and SP_{20}. For the MBB beam, it is the SP_{20} and S_{20} sets. Another interesting result we highlight is that the best connected, *i.e.*, lowest r_{dc}, structures result from the diverse subsets that consider shape. For cantilever, this is the S_{20} set, although P_{20} is close, and for MBB it is S_{20} and SP_{20}. The slightly different results between the two examples illustrate that the effectiveness of property or shape diversity is problem-dependent. Still, small yet diverse subsets can outperform random or full datasets regardless of the problem.



(a) Cantilever results using subsets. P_{20}, S_{20} and SP_{20} achieve the best MSE overall while S_{20} and P_{20} perform best for r_{dc}.



(b) MBB beam results using subsets. The best overall MSE is obtained by SP_{20} and S_{20}, and the best r_{dc} by S_{20} and SP_{20}.

FIGURE 5: The final objective values (MSE) and ratios of disconnectivity (r_{dc}) of 10 runs per subset. Lower values are better. Overall, the shape diverse subsets achieve the best connectivity, and small yet diverse subsets attain lower MSE.

TABLE 1: Mean of the final results for the 2D examples, with the lowest values in bold.

	G	R_{100}	S_{100}	SP_{100}	P_{100}	R_{20}	S_{20}	SP_{20}	P_{20}
Bridge MSE	58.2626	1.2385	0.3530	0.5040	0.6215	0.4264	0.3104	0.3399	**0.1266**
Bridge r_{dc}	0.4943	0.4105	0.3151	0.3618	0.3287	0.4121	**0.2566**	0.3126	0.2786
MBB MSE	1.3E+18	1.5341	0.4278	0.6454	1.6648	1.2395	0.2865	**0.2017**	0.4926
MBB r_{dc}	0.5184	0.4770	0.3406	0.3347	0.4653	0.4836	**0.2488**	0.2578	0.3996

In Fig. 6, we show the final topologies of the runs that achieve the minimum MSE for some of the datasets. As expected from the low performance and compatibility using the full G, the designs in Fig. 6c contain disconnected and oddly matched unit cells. Following from the worse r_{dc} for property diverse sets, the topologies for P_{20} are visually less compatible (Fig. 6a) than those of shape diverse sets (Fig. 6b), with neighbors connected by small features. This can be associated with the observation in Sec. 3.1.2 that METASET tends to include unit cells with small features as it maximizes property diversity, leading to subsets with less compatible unit cells. These final designs further enforce our findings that shape diversity can improve boundary compatibility.

Although our constrained genetic algorithm provides satisfactory designs, we must point out that this global method was implemented to showcase the impact of subset size and diversity. While more elegant optimization techniques would be better suited for practical applications, we nevertheless believe that the insights gained from this study — that selecting diverse subsets can accelerate and benefit metamaterial design — can be generalized to other data-driven methods. This is an exciting direction for future works.

3.2 Discovery of Diverse 3D Unit Cell Families

In the 3D study, we demonstrate another benefit of METASET: eliminating inherent bias in datasets by optimizing the diversity of a subset, which can subsequently be leveraged for data-driven design or ML of, *e.g.*, property prediction or generative models (Fig. 1). We first introduce a new method based on periodic functions to generate 3D families of unit cells with the same underlying structure but varying densities, which although fast creates a great number of overlapping shapes. Our goal in applying METASET to this 3D data is to sift through the overlaps to discover diverse sets of unique isosurface families.

3.2.1 3D Unit Cell Generation using Families of Level-Set Functions

Triply periodic isosurface unit cells, whose symmetries follow those of crystal structures [40], are often used in 3D mechanical metamaterials design due to excellent surface area-to-performance ratios and manufacturability [2]. In addition, their representation as level-set functions allows the density of the unit cells to be easily manipulated for functionally-graded structures [2, 25] and tailorable acoustic bandgaps [41]. A level-set function $f(x,y,z) = t$ is an implicit representation of geometry where the t-isocontour, *i.e.*, the points where $f = t$,

(a) Using property diverse subset of size 20 (P_{20}), the unit cells are connected, but by small features.

(b) Using shape diverse subset of size 20 (S_{20}), we observe superior connectivity between neighboring unit cells.

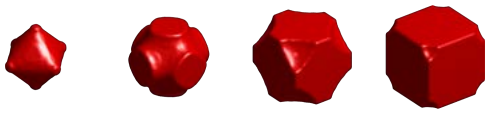
(c) Using the full dataset G of size 17,380, we find the final topologies are critically disconnected.

FIGURE 6: Final topologies with the lowest MSE out of 10 runs using 20-item diverse sets (a,b), which are more well-connected than the one using the full dataset (c). Left: cantilever example; right: classical MBB beam example showing the full structure after symmetry.

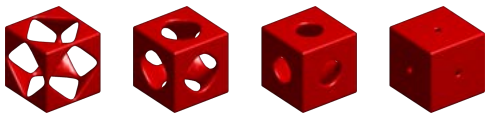
describes the surface of the structure, while the locations where $f < t$ are solid material, and void where $f > t$. Thus, by varying the isovalue t, an entire family of isosurface unit cells can be extracted from one level-set function.

The most prevalent type of isosurfaces used in metamaterials design is a special subset known as Triply Periodic Minimal Surfaces (TPMS). However, only a few TPMS families have been used since their functions are complex to derive [40]. For example, Maskery *et al.* use six families in their design work [2], while Li *et al.* use four [25]. Moreover, it has not been investigated whether these few families cover the gamut of shapes and properties needed for design applications. Suppose a researcher wishes to design a new functionally-graded 3D metamaterial by tuning the densities of isosurface functions, but does not know beforehand which families would best suit their application. Due to the computational expense of design in 3D, they may desire to select a smaller set of families that can then be used in their optimization method. In this section, we present METASET as a procedure to choose those families such that the resultant subset has large coverage over different properties and shapes. In doing so, we also demonstrate that METASET removes bias in datasets by maximizing diversity.

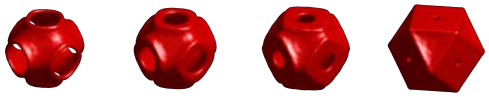
Creation of New Isosurface Families Before selecting diverse families, we must first generate an initial pool to choose from. Thus, to build a large 3D dataset, we propose a new method to create isosurface families based on the level-set functions of crystallographic structure factors, which describe how particles are arranged in a crystal unit cell [42]. In contrast to most unit cell generation methods, our approach here does not set targets in the property space or use TO, and different from TPMS functions, a larger variety of shapes can be found without complex derivations.



(a) Family $A_{229,(001)}(X,Y,Z) \leq t$



(b) Family $A_{229,(001)}(X,Y,Z) \geq t$



(c) Family $A^2_{229,(001)}(X,Y,Z) \leq t^2$

FIGURE 7: Examples of unit cells from isosurface families generated by the structure factor for space group No. 229 and $(hkl) = (001)$. The effect of increasing t to create a family is shown from left to right.

In crystallography, structures that are invariant under the same symmetry operations belong to the same space group, of which there are a total of 230 for 3D structures. For the purposes of our work, we will focus on the 36 cubic groups, No. 195 through 230, to obtain our level-set functions. Experimentally, the space group of a crystal can be determined through, *e.g.*, X-ray techniques, by scattering radiation off a lattice plane denoted by (hkl), and then observing the diffraction pattern. These symmetric patterns have been analytically modeled as *structure factors*, which are periodic functions of the form:

$$f_{group,(hkl)}(X,Y,Z) = A + iB, \tag{6}$$

where $A = \cos\left(hX + kY + lZ\right)$, $B = \sin\left(hX + kY + lZ\right)$, $X = 2\pi x$, $Y = 2\pi y$, and $Z = 2\pi z$. The equations of these structure factors are listed in [42] for all space groups and their allowable (hkl).

We can split each structure factor into six isosurface families by separating A and B in Eq. 6 (inspired by [40]), and converting them into level-set functions as follows:

$$\begin{aligned} A_{group,(hkl)}(X,Y,Z) &\leq t, \\ A_{group,(hkl)}(X,Y,Z) &\geq t, \\ A^2_{group,(hkl)}(X,Y,Z) &\leq t^2, \end{aligned} \tag{7}$$

and similarly for $B_{group,(hkl)}$. These, respectively, correspond to

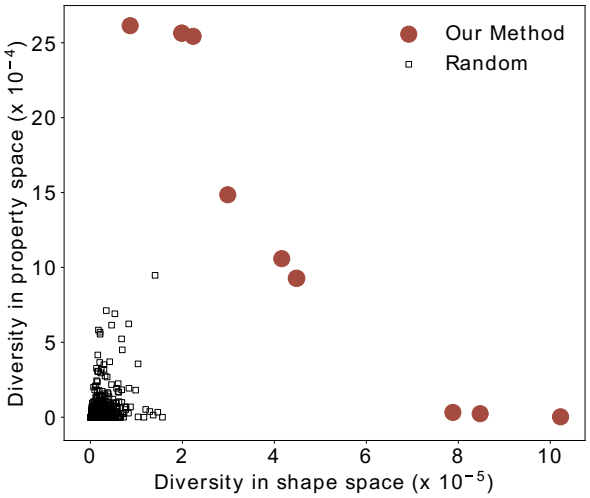


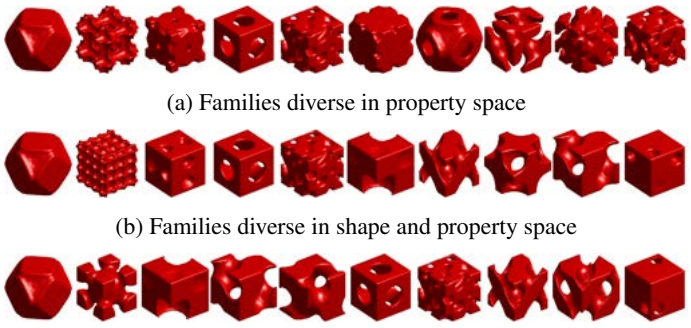
FIGURE 8: Trade-off between diversity in property vs. shape space. The minimum diversity in shape space for optimized sets has a diversity score greater than 98.9% of random samples.

setting as solid material the function values that are less than t (Fig. 7a), greater than t (Fig. 7b), and in between $-t$ and t (leading to a "thin-walled" structure; Fig. 7c).

Thus, instead of using the limited TPMS functions, we can use the structure factors of all 36 cubic space groups and their corresponding (hkl) to generate a greater number of isosurface families for data-driven design. In this way, we quickly created 483 families without performing property-driven optimization. Although efficient, this method also causes an imbalance in geometry, since several structure factors differ only by a coefficient and lead to overlapping families. For example, space groups No. 195 and 196 are related as $A_{195,(hkl)} = 4 \cdot A_{196,(hkl)}$, and therefore generate the same structures. Next, we demonstrate the prowess of METASET in systematically removing such overlaps when selecting diverse subsets.

3.2.2 METASET Results As the families are comprised of a range of densities and therefore shapes and properties, we need to capture the similarities of individual unit cells while assessing the similarities between families. Hence, we generate 100 samples from each family covering all densities between 0.01 and 0.99, giving 48,300 unit cells total. Each unit cell is represented as a 4096-dimensional point cloud by first converting its level-set field into a triangle mesh [43], and then sampling uniformly on the triangular faces [44]. We also remove any small disconnected features during post-processing, and find the homogenized elastic tensors of each unit cell using a code modified from [45].

To quantify the similarity between two families, we assume each family is a collection of points, where each point corresponds to a unit cell. This reduces the problem of finding similarity between two families to one between two point sets, which



(a) Families diverse in property space

(b) Families diverse in shape and property space

(c) Families diverse in shape space

FIGURE 9: Examples of subsets of 3D isosurface families selected by METASET.

we calculate using the Hausdorff distance (Sec. 2.1). In property space, the similarity between two families is related to the maximum of the pairwise Euclidean distances between the effective elastic tensor components of each unit cell. For shape similarity between families C and D, we take the Hausdorff distance twice: first using Eq. 2 to calculate $d_{\mathrm{H}}(c,d)$ between individual unit cells $c \in C$ and $d \in D$, and then substituting this into Eq. 1 to obtain the *inter-familial* distance, $h(C,D)$. Intuitively, this means that the shape similarity between two families is the maximum of the similarities between closest-in-shape pairs of unit cells. Therefore, rather than simply averaging the features of each family, the inter-familial similarities also consider the diversity of unit cell members within each family.

Our goal is to find a small subset of 10 families out of 483 that are diverse in both shape and property space. Using Eq. 4, we vary the weight w between 0 and 1, and run the greedy algorithm (Sec. 2.3) for each weight to find the optimal subsets. Next, we calculate the diversity of each subset in both property space (determinant of L_P) and shape space (determinant of L_S). As a baseline, we also randomly sample 1000 sets of families and measure their diversity in each space as well. Fig. 8 shows these diversity scores. Despite drawing several random subsets (which are representative of the distribution of the similarity values between pairs of unit cells), 98.9% of them still fall short of the optimized subset with the lowest shape diversity score. This is compelling evidence that 1) the original randomly generated dataset was severely imbalanced, and 2) METASET is able to combat such bias and select more diverse subsets. Additionally, the optimized scores (our method) in Fig. 8 illustrate the trade-off between the shape and property similarity metrics: the diversity in property space drops as we select sets that are more diverse in shapes, and vice versa. This can be leveraged to tune the dataset to specific applications, *e.g.*, finding optimal designs or ML.

We point out that the trade-off might raise this question: How can a set of families which are quite diverse in property space have low diversity in shapes, even though similar shapes

are expected to possess similar properties? On careful inspection of the left side of Fig. 8, one can notice that the sets of families with higher diversity in property space and seemingly "low" diversity in shape space actually have larger shape diversity scores than the majority of the random sets. This shows that the highest diversity in property space is achieved by a set of families which are also very diverse in shape. An interesting avenue of future research is whether the Hausdorff distance can affect this trade-off front. Using Hausdorff, we observed that even if two families have near complete overlap, with just one outlier, they can still have a large distance between them. While this property is mathematically desirable, and widely used in measuring similarity between point clouds, further research is needed to understand if this metric is the most appropriate choice for design applications.

Some example subsets of diverse isosurface families are shown in Fig. 9, where the 50^{th} (median) sample from each family are pictured. Intriguingly, we note that in the shape diverse set (Fig. 9c), families generated from the same space group and (hkl), but different level-set forms (Eq. 6) appear. For example, the second and third to last items in Fig. 9c have the equations $A_{203,(111)} \geq t$ and $A_{203,(111)} \leq t$. One could think of these as completely different shapes with almost no overlaps—which is further validation of the shape diversity chosen by METASET.

Moreover, the property-only and shape-only diverse sets (Figs. 9c and 9a, respectively) share very few of the same families, while the set that is equally shape and property diverse (Fig. 9b) contains a mixture of the former two. Common TPMS used for metamaterials design, such as the Primitive, Gyroid and Diamond (see [2, 25]) are also included among our diversified families. In future works, we will explore the use of these families for data-driven metamaterials design and ML. We provide the data of the METASET results publicly so that the diverse families can be employed by any designer in their work as well.

4 Discussion

Although we illustrated the benefits of METASET in Sec. 3, there are nevertheless some topics worthy of examining in the future. From our design of 2D aperiodic structures, we saw that whether one subset leads to better design performance is problem dependent. Shape diverse subsets may increase the chance to find compatible neighboring unit cells, while property diverse sets might enhance problems that require a wider range of target properties at the cost of connectivity. This dependence extends to ML tasks in the data-driven design framework (Fig. 1) as well. To train property prediction models, one may need a property diverse dataset, while for a deep generative model that learns geometric features, a shape diverse set might be more appropriate. Along these lines, it would be interesting to further validate the improved performance of diverse datasets for design and ML tasks using our subsets of diverse 3D unit cell families in a future work.

In the 2D examples, we also observed that smaller subsets led to designs with performance closer to the targets; in fact, we found using METASET that only 20 unit cells were enough to form a diverse subset. In most cases, the benefits of reducing the search space, model training time, or storage space of the dataset could outweigh any loss of data. However, certain applications such as ML may need large datasets. A key benefit of using a METASET, even for large subset sizes, is that it reduces bias by rank ordering all items in the dataset. The items with the highest redundancy in shape or property (like duplicates) are pushed towards the end of the rank-ordered list, so that ML algorithms trained on any subset will be less biased. While it is not difficult to increase the size of the set, determining how much data is enough is more challenging since this too is contingent on the application. The effectiveness of subset size and property versus shape diversity on specific tasks in metamaterials design is an important question for future studies.

Finally, we remark that the capability of METASET depends on the choice of similarity metrics as well as the definition of the joint similarity kernel, both of which are avenues of further research. In this work, the 2D descriptors and Hausdorff distance worked well in measuring shape similarity, but there are a wealth of other choices that may provide different results. Extending METASET to more complex properties, like dynamic ones, may necessitate a metric other than the Euclidean norm. For the joint DPP kernel, we chose a simple weighted sum to join the shape and property matrices, but other methods to combine kernels while preserving submodularity, are also possible. However, swapping these to best suit the application is easily done since the input of the DPPs-based greedy algorithm in METASET is a positive semi-definite similarity kernel that can be obtained from any appropriate metric or definition.

5 Conclusion

In this paper, we propose a methodology, METASET, that incorporates diversity of both shape and property space into data selection to improve the downstream tasks in data-driven design. It is efficient and flexible, allowing the emphasis on either shape or property to be easily traded by measuring and maximizing the joint diversity of subsets through a weighted DPP similarity kernel. To calculate this kernel matrix, we introduced similarity metrics that cater specifically to 2D and 3D metamaterials. Although in this work we focused on the design of mechanical metamaterials, we note that METASET can be transferred to other metamaterial domains, or indeed any other design problems that need to balance design space against some performance or quality space.

By way of our 2D aperiodic metamaterial design examples, we demonstrated that small yet diverse subsets of unit cells can boost the scalability of the global search algorithm, and lead to

designs that better achieve target properties with greater boundary compatibility. This revelation shakes a common belief in the field of data-driven mechanical metamaterials design that a larger and denser dataset is required to design well-connected structures while still meeting the target performance. To our knowledge, this is the first time that such a result has been studied and presented. In the 2D examples, we also found that different problems may require a different weight of shape and property diversity based on the design goals. The ease at which a subset's size as well as the weight of shape and property diversity can be explored is yet another advantage of METASET.

In our 3D case study, we not only proposed a new method to generate triply periodic isosurface families using crystallographic structure factors, but also verified that METASET can effectively reduce imbalance in metamaterial datasets. Similar to well-known TPMS unit cells, each of our 3D families are represented as level-set functions whose density parameter can be easily manipulated to design functionally-gradient metamaterials, or to tune an individual unit cell for greater compatibility with its neighbors. Different from established works, however, our dataset of families are optimized for shape and property diversity using METASET rather than arbitrarily chosen.

The methods proposed in this paper to achieve diversity for shape and property space is broadly applicable to domains outside data-driven metamaterials design too. In design ideation, our method can be used to select ideas to show to a designer that are functionally different from each other while achieving different performance goals. It can also be integrated with existing multi-objective optimization algorithms as a niching method. To contribute to the growth and capability of data-driven metamaterials design methods and other fields, we have shared diversified subsets of 2D and 3D unit cells, as well as the corresponding equations of isosurface families at `https://github.com/lychan110/metaset`. These unit cells can be directly plugged by any metamaterials designer in their application.

Acknowledgements

We are grateful for support from the NSF CSSI program (Grant No. OAC 1835782). Yu-Chin Chan thanks the NSF Graduate Research Fellowship (Grant No. DGE-1842165). Liwei Wang would like to acknowledge support from Zhiyuan Honors Program for Graduate Students of Shanghai Jiao Tong University for his predoctoral visiting study at Northwestern University.

REFERENCES

[1] Schumacher, C., Bickel, B., Rys, J., Marschner, S., Daraio, C., and Gross, M., 2015. "Microstructures to control elasticity in 3d printing". *ACM Transactions on Graphics (TOG),* ***34***(4), pp. 1–13.

[2] Maskery, I., Aremu, A., Parry, L., Wildman, R., Tuck, C., and Ashcroft, I., 2018. "Effective design and simulation of surface-based lattice structures featuring volume fraction and cell type grading". *Materials & Design,* ***155***, pp. 220–232.

[3] Choi, H., Baek, A. M. C., and Kim, N., 2019. "Design of non-periodic lattice structures by allocating pre-optimized building blocks". In Volume 1: 39th Computers and Information in Engineering Conference, American Society of Mechanical Engineers.

[4] Vogiatzis, P., Ma, M., Chen, S., and Gu, X. D., 2018. "Computational design and additive manufacturing of periodic conformal metasurfaces by synthesizing topology optimization with conformal mapping". *Computer Methods in Applied Mechanics and Engineering,* ***328***, pp. 477–497.

[5] Deng, J., and Chen, W., 2017. "Concurrent topology optimization of multiscale structures with multiple porous materials under random field loading uncertainty". *Structural and Multidisciplinary Optimization,* ***56***(1), pp. 1–19.

[6] Du, Z., Zhou, X.-Y., Picelli, R., and Kim, H. A., 2018. "Connecting microstructures for multiscale topology optimization with connectivity index constraints". *Journal of Mechanical Design,* ***140***(11).

[7] Liu, H., Zong, H., Tian, Y., Ma, Q., and Wang, M. Y., 2019. "A novel subdomain level set method for structural topology optimization and its application in graded cellular structure design". *Structural and Multidisciplinary Optimization,* ***60***(6), pp. 2221–2247.

[8] Coulais, C., Teomy, E., de Reus, K., Shokef, Y., and van Hecke, M., 2016. "Combinatorial design of textured mechanical metamaterials". *Nature,* ***535***(7613), pp. 529–532.

[9] Zhu, B., Skouras, M., Chen, D., and Matusik, W., 2017. "Two-scale topology optimization with microstructures". *ACM Transactions on Graphics,* ***36***(4), p. 1.

[10] Callanan, J., Ogunbodede, O., Dhameliya, M., Wang, J., and Rai, R., 2018. "Hierarchical combinatorial design and optimization of quasi-periodic metamaterial structures". In Volume 2B: 44th Design Automation Conference, American Society of Mechanical Engineers.

[11] Wang, L., Chan, Y.-C., Liu, Z., Zhu, P., and Chen, W., 2020. "Data-driven metamaterial design with laplace-beltrami spectrum as "shape-DNA"". *Structural and Multidisciplinary Optimization*.

[12] Bostanabad, R., Chan, Y.-C., Wang, L., Zhu, P., and Chen, W., 2019. "Globally approximate gaussian processes for big data with application to data-driven metamaterials design". *Journal of Mechanical Design,* ***141***(11).

[13] White, D. A., Arrighi, W. J., Kudo, J., and Watts, S. E., 2019. "Multiscale topology optimization using neural network surrogate models". *Computer Methods in Applied Mechanics and Engineering,* ***346***, pp. 1118–1135.

[14] Chen, D., Levin, D. I. W., Sueda, S., and Matusik, W.,

2015. "Data-driven finite elements for geometry and material design". *ACM Transactions on Graphics (TOG),* ***34***(4), pp. 1–10.

[15] Ma, W., Cheng, F., Xu, Y., Wen, Q., and Liu, Y., 2019. "Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy". *Advanced Materials,* ***31***(35), p. 1901111.

[16] Liu, Z., Zhu, D., Rodrigues, S. P., Lee, K.-T., and Cai, W., 2018. "Generative model for the inverse design of metasurfaces". *Nano Letters,* ***18***(10), pp. 6570–6576.

[17] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G., 2017. "Learning from class-imbalanced data: Review of methods and applications". *Expert Systems with Applications,* ***73***, pp. 220–239.

[18] Panetta, J., Zhou, Q., Malomo, L., Pietroni, N., Cignoni, P., and Zorin, D., 2015. "Elastic textures for additive fabrication". *ACM Transactions on Graphics (TOG),* ***34***(4), pp. 1–12.

[19] Branco, P., Torgo, L., and Ribeiro, R. P., 2016. "A survey of predictive modeling on imbalanced domains". *ACM Computing Surveys,* ***49***(2), pp. 1–50.

[20] Coombs, C. H., and Avrunin, G. S., 1977. "Single-peaked functions and the theory of preference.". *Psychological Review,* ***84***(2), pp. 216–230.

[21] Ahmed, F., Fuge, M., and Gorbunov, L. D., 2016. "Discovering diverse, high quality design ideas from a large corpus". In ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.

[22] Ahmed, F., and Fuge, M., 2017. "Ranking ideas for diversity and quality". *Journal of Mechanical Design,* ***140***(1).

[23] Kulesza, A., Taskar, B., et al., 2012. "Determinantal point processes for machine learning". *Foundations and Trends® in Machine Learning,* ***5***(2–3), pp. 123–286.

[24] de Aguiar, P. F., Bourguignon, B., Khots, M., Massart, D., and Phan-Than-Luu, R., 1995. "D-optimal designs". *Chemometrics and intelligent laboratory systems,* ***30***(2), pp. 199–210.

[25] Li, D., Dai, N., Tang, Y., Dong, G., and Zhao, Y. F., 2019. "Design and optimization of graded cellular structures with triply periodic level surface-based topological shapes". *Journal of Mechanical Design,* ***141***(7).

[26] Bustos, B., Keim, D. A., Saupe, D., Schreck, T., and Vranić, D. V., 2005. "Feature-based similarity search in 3d object databases". *ACM Computing Surveys (CSUR),* ***37***(4), pp. 345–387.

[27] Rostami, R., Bashiri, F. S., Rostami, B., and Yu, Z., 2018. "A survey on data-driven 3d shape descriptors". *Computer Graphics Forum,* ***38***(1), pp. 356–393.

[28] Sharon, E., and Mumford, D., 2006. "2d-shape analysis using conformal mapping". *International Journal of Computer Vision,* ***70***(1), pp. 55–75.

[29] Su, Z., Wang, Y., Shi, R., Zeng, W., Sun, J., Luo, F., and Gu, X., 2015. "Optimal mass transport for shape matching and comparison". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* ***37***(11), pp. 2246–2259.

[30] Vamvakas, G., Gatos, B., and Perantonis, S. J., 2010. "Handwritten character recognition through two-stage foreground sub-sampling". *Pattern Recognition,* ***43***(8), pp. 2807–2816.

[31] Das, N., Reddy, J. M., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M., and Basu, D. K., 2012. "A statistical–topological feature combination for recognition of handwritten numerals". *Applied Soft Computing,* ***12***(8), pp. 2486–2495.

[32] Sarkhel, R., Das, N., Das, A., Kundu, M., and Nasipuri, M., 2017. "A multi-scale deep quad tree based feature extraction method for the recognition of isolated handwritten characters of popular indic scripts". *Pattern Recognition,* ***71***, pp. 78–93.

[33] Kobbelt, L., and Botsch, M., 2004. "A survey of point-based techniques in computer graphics". *Computers & Graphics,* ***28***(6), pp. 801–814.

[34] Huttenlocher, D., Klanderman, G., and Rucklidge, W., 1993. "Comparing images using the hausdorff distance". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* ***15***(9), pp. 850–863.

[35] Fan, H., Su, H., and Guibas, L. J., 2017. "A point set generation network for 3d object reconstruction from a single image". In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[36] Kulesza, A., and Taskar, B., 2011. "k-dpps: Fixed-size determinantal point processes". In Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 1193–1200.

[37] Borodin, A., 2009. "Determinantal point processes". *arXiv preprint arXiv:0911.1153*.

[38] Feige, U., Mirrokni, V. S., and Vondrak, J., 2011. "Maximizing non-monotone submodular functions". *SIAM Journal on Computing,* ***40***(4), pp. 1133–1153.

[39] Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L., 1978. "An analysis of approximations for maximizing submodular set functions—i". *Mathematical Programming,* ***14***(1), pp. 265–294.

[40] Wohlgemuth, M., Yufa, N., Hoffman, J., and Thomas, E. L., 2001. "Triply periodic bicontinuous cubic microdomain morphologies by symmetries". *Macromolecules,* ***34***(17), pp. 6083–6089.

[41] Abueidda, D. W., Jasiuk, I., and Sobh, N. A., 2018. "Acoustic band gaps and elastic stiffness of PMMA cellular solids based on triply periodic minimal surfaces". *Materials & Design,* ***145***, pp. 20–27.

[42] Shmueli, U., ed., 2010. *International Tables for Crystallography: Volume B, Reciprocal space*. International Union of Crystallography.

[43] Vogiatzis, P., Chen, S., and Zhou, C., 2017. "An open source framework for integrated additive manufacturing and level-set-based topology optimization". *Journal of Computing and Information Science in Engineering,* ***17***(4).

[44] Kleiman, Y., 2017. Github repository, sample_mesh. `https://github.com/hexygen/sample_mesh`.

[45] Dong, G., Tang, Y., and Zhao, Y. F., 2018. "A 149 line homogenization code for three-dimensional cellular materials written in matlab". *Journal of Engineering Materials and Technology,* ***141***(1).