Microstructure Optimization with Constrained Design Objectives using Machine Learning-Based Feedback-Aware Data-Generation

Arindam Paul, Pinar Acar, Wei-keng Liao, Alok Choudhary, Veera Sundararaghavan, Ankit Agrawal

Abstract

Microstructure sensitive design has a critical impact on the performance of engineering materials. The safety and performance requirements of critical components, as well as the cost of material and machining of Titanium components, make dovetailing of the microstructure imperative. This paper addresses the optimization of several microstructure design problems for Titanium components under specific design constraints using a feedback-aware data-driven solution methodology. In this study, the microstructure is modeled with an orientation distribution function (ODF) that measures the volumes of different crystallographic orientations. Two algorithms are used to sample the entire microstructure space followed by machine learning-aided identification of a minimal subset of ODF dimensions which is subsequently explored by targeted sampling.

Conventional optimization methods lead to a unique microstructure rather than yielding a comprehensive space of optimal or near-optimal microstructures. Multiple solutions are crucial for the deployment of materials design for manufacturing as traditional manufacturing processes can only generate a limited set of microstructures. Our data sampling-based methodology not only outperforms or is on par with other optimization techniques in terms of the optimal property value, but also provides numerous near-optimal solutions, 3-4 orders of magnitude more than previous methods. Consequently, the proposed framework delivers a spectrum of optimal solutions in the microstructure space which can accelerate materials development and reduce manufacturing costs.

Nomenclature

σ	=	volume averaged stress (in MPa)
ε	=	volume averaged strain
α	=	constant
q	=	volume normalization vector
r	=	orientation
V	=	null space vector
S	=	compliance (in 1/GPa)
C	=	stiffness (in GPa)
C_{eff}	=	effective stiffness (in GPa)
A	=	orientation distribution function
χ	=	orientation dependent property
D	=	total number of dimensions of ODF vector
k	=	intended number of non-zero dimensions of ODF vector

1. Introduction

Exploring and harnessing the association between processing, structure, properties, and performance is a critical aspect of new materials exploration [1–8]. Variation in microstructure leads to a wide range of materials properties which in turn impacts the performance. The materials performance can be significantly improved by dovetailing the microstructure [9–12]. Titanium alloys are used for airframe panels, and optimizing the property is necessary for the safety and performance of the aircraft [13–16]. Furthermore, both the cost of the material and machining for Titanium panels are expensive [17, 18]. Titanium airframe panels are modeled as thin, rectangular, anisotropic plates. However, the panels are exposed to elevated temperatures in moving devices. Titanium is a lightweight, yet strong, uniquely versatile metal with properties such as high tensile strength to density ratio, high corrosion resistance and ability to withstand high temperatures without creeping. In addition, Titanium is a very ductile material that can be worked into many shapes. Titanium [19] has a very high melting point cap (3000

degrees Fahrenheit) which makes it able to bear high-heat environments. Combination of all these properties make Titanium alloys ideal for use in aircraft applications.

One of the major goals of materials design optimization is the trade-off of properties based on prioritizing one design goal over others [20, 21]. For microstructure optimization, it can involve enhancing properties in one direction while sacrificing other properties which are not as important for the design problem [22]. Techniques that allow tailoring of properties of polycrystalline alloys involves selection of preferred orientations of various crystals constituting the polycrystalline alloy. This paper addresses the problem by tailoring crystallite distribution for specific optimization design problems. The orientation distribution function (ODF) is used to quantify the microstructure [3, 23–25] which represents the volume fractions of crystals of different orientations of the microstructure.

In this paper, we aim to explore the microstructure optimization of multiple design problems for a Titanium panel. Two different mesh sizes to represent ODFs are explored in this work : 50 and 388. Three different properties: coefficient of expansion α, stiffness coefficient C_{11} and yield stress σ are optimized. We use two algorithms - allocation and partition to sample the entire microstructure space. There has been several works on application of machine-learning for accelerated materials discovery and design of new materials with select engineering properties [26–37]. In this work, we harness machine learning for microstructure search space reduction for identification of a minimal subset of ODF dimensions which is subsequently explored by targeted sampling. Our data sampling-based methodology not only outperforms or is on par with other optimization techniques in terms of the optimal property value, but also provides numerous near-optimal solutions up to 3-4 orders of magnitude more than previous methods.

2. Background and Related Works

The alloy microstructure is composed of multiple crystals with each crystal having its distinct orientation. The orientation distribution function (ODF) is one representation for depicting the volume fractions of the crystals for different orientations in the

microstructure. In this work, the microstructure of Titanium panels is modeled using ODFs [23–25, 38]. The ODF is a critical concept in texture analysis and anisotropy. The ODFs are represented by axis-angle parameterization of the crystal lattice rotation in the orientation space, as proposed by Rodrigues [39]. The Rodrigues' parameterization is generated by scaling the axis of rotation n as $r = ntan\frac{\theta}{2}$, where θ is the rotation angle. Orientation distributions can be described mathematically in any space appropriate to a continuous description of rotations [23, 38, 39].

The orientation space can be scaled down to a subset called the fundamental region. Each crystal orientation is depicted uniquely inside the fundamental region by a parameterization coordinate for the rotation r. The ODF, represented by $A(r)$, is the volume density of crystals of orientation r. Each microstructural orientation is associated with a nodal point ODF. The local scheme developed with the finite element discretization is advantageous since it can represent sharp textures, including extreme cases such as single crystals. The fundamental region is discretized into N independent nodes with N_{elem} finite elements and N_{int} integration points per element. A detailed explanation of the ODF discretization and volume averaged equations has been provided in [9, 40–42]. A single particular orientation or texture component is represented by each point in the orientation distribution. The orientation distribution information can be used to determine the presence of components, volume fractions and predict anisotropic properties of polycrystals.

3. Problem Statement

We aim to explore the microstructure optimization of multiple design problems for a Titanium panel. Two different mesh sizes to represent ODFs are investigated in this work : 50 and 388. Three separate properties: coefficient of thermal expansion α, stiffness coefficient C_{11} and yield stress σ are optimized. There are four different design problems explored, and both the upper and lower bounds are solved. Figure 1 illustrates a section of Titanium aircraft panel and the corresponding microstructure cross-section.

The ODF values are associated with an orientation of the microstructure. Using

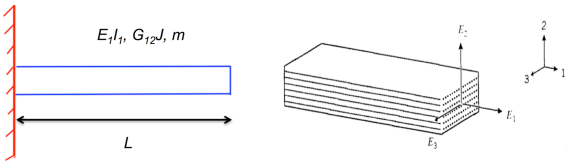


Figure 1: Geometric representation of Titanium panel. E and G indicate the Young's modulus and shear modulus values around the corresponding directions, J is the torsion constant, m is the unit mass, L is the length of the beam, and I_1 is the moment of inertia along axis 1.

the ODF approach is advantageous since the averaged material properties over a microstructural domain can be computed using the homogenization (averaging) equations which are linear with respect to the ODF values. This is true when the effects of crystal size and shape are ignored, and homogenous deformity is assumed in the volume element. Using the homogenization relation, the orientation-dependent averaged material property, $< \chi >$, can be computed using the material property values at different orientations, $\chi(r)$, and the ODF values, A.

$$< \chi >= \int_R \chi(r)A(r)dv,$$

where, the orientation is denoted by r. The ODF representation should satisfy the following volume normalization constraint in the microstructural domain.

$$\int_R A(r)dv = 1$$

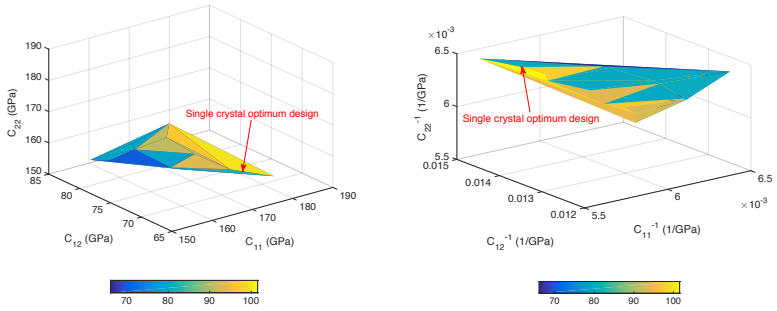
The optimization problems of interest aim to identify the best microstructure design to enhance the material properties. Since the ODF values quantify the microstructural texture, the goal is to identify the optimum ODF values for each problem. However, the ODF solution space is high-dimensional, and it leads to an optimization problem with numerous design variables. Here, one favorable approach would be generating a new solution space, which is called as property closure, which includes the complete range of properties obtainable from the space of the ODFs. In property closure approach, the material properties can be calculated with either upper or lower bound averaging assumption [40]. An example computation of property closure with upper and lower

bounds approaches is shown in Figure 2 for stiffness (C_{11}, C_{12} and C_{22}) and compliance (S_{11}, S_{12} and S_{22}) properties. The example computations for the averaged stiffness, $< C >$, and compliance, $< S >=< C^{-1} >$, are given next for the upper and lower bound approaches respectively.

$$< C >= \int_R CAdv$$

$$< S >=< C^{-1} >= \int_R C^{-1}A^{-1}dv$$

$$< S >= \int_R SAdv$$



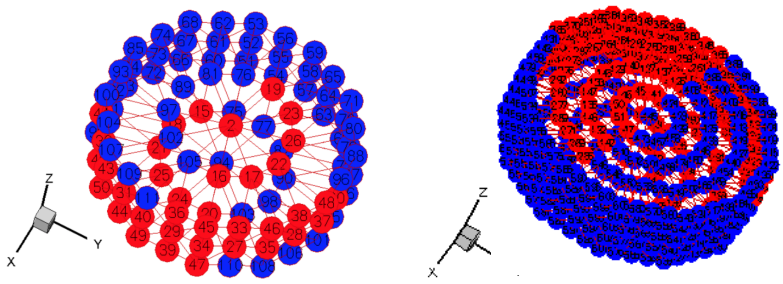
(a) Property closure for stiffness (C_{11},C_{12} and C_{22}) parameters

(b) Property closure for compliance (S_{11}, S_{12} and S_{22}) parameters

Figure 2: Property closures in C and S (C^{-1}) space for HCP Titanium. The color scale are represented in C space in both plots.

In the present work, we will utilize both upper and lower bound averaging techniques to identify the optimum microstructure solutions. The material of interest is polycrystalline α-Titanium as shown in Figure 3 (a), red color shows independent orientations, blue color shows dependent orientations resulting from the crystallographic symmetries. We will model this hexagonal close-packed (HCP) structure using 111 ODF values defined in the Rodrigues fundamental region as shown in Figure 3 (a). However, we will only use 50 independent ODF values for modeling purpose since the remaining

ODF values can be determined using the crystallographic symmetries. In Fig. 3 (b), a finer finite element mesh, that can improve the numerical resolution of microstructural texture representation, having 388 independent ODF values is illustrated.



(a) ODF representation indicating the location of 50 independent nodes in the orientation space in red

(b) ODF representation indicating the location of 388 independent nodes in the orientation space in red

Figure 3: Finite element discretization of the orientation space of HCP Titanium.

In this work we solve for the best microstructure design that maximizes desired properties which are coefficient of thermal expansion α_x, stiffness coefficient C_{11} and yield stress σ_y and satisfies the design constraints. The material properties of the objective function are computed using the upper bound averaging approach. For design constraints both upper and lower bound averaging approaches are utilized.

Four design problems are presented in this work, and each of them is solved using both upper and lower bound approach. Upper bound sub-problems for design problems 1 and 2 are solved in mesh sizes of both 50 and 388, while the lower bound sub-problems are solved in 50 dimensions. Both the upper and lower bound design sub-problems 3 and 4 are solved in mesh size of 388. The finer mesh with the 388 ODF values is expected to provide a more accurate representation as the Rodrigues domain is discretized with more variables. The design constraints of the optimization problems reflect certain stiffness needs of engineering designs.

Problem 1:

max α_x

Upper Bound: (mesh dimension 50 and 388)

$$\textit{subject to } 161 \leq C_{11} \leq 165 \text{ GPa} \quad (1a)$$
$$\textit{subject to } 75 \leq C_{12} \leq 78 \text{ GPa} \quad (1b)$$

Lower Bound: (mesh dimension 50)

$$\textit{subject to } 0 \leq C_{11} \leq 125 \text{ GPa} \quad (2a)$$
$$\textit{subject to } 90 \leq C_{12} \leq 95 \text{ GPa} \quad (2b)$$

Problem 2:

max C_{11}

Upper Bound: (mesh dimension 50 and 388)

$$\textit{subject to } 75 \leq C_{12} \leq 78 \text{ GPa} \quad (3)$$

Lower Bound: (mesh dimension 50)

$$\textit{subject to } 90 \leq C_{12} \leq 95 \text{ GPa} \quad (4)$$

Problem 3: (mesh dimension 388)

max σ_y

Both Bounds:

$$\textit{subject to } \leq S_{11} \leq 0.15 \text{ 1/GPa} \quad (5a)$$

$$\textit{subject to } \leq S_{22} \leq 0.1 \text{ 1/GPa} \quad (5b)$$

Problem 4: (mesh dimension 388)

$$\textbf{max } \sigma_y$$

Upper Bound:

$$\textit{subject to } 120 \leq C_{11} \leq 130 \text{ GPa} \tag{6a}$$

$$\textit{subject to } 90 \leq C_{12} \leq 95 \text{ GPa} \tag{6b}$$

$$\textit{subject to } 0 \leq S_{11} \leq 0.15 \text{ 1/GPa} \tag{6c}$$

$$\textit{subject to } 0 \leq S_{22} \leq 0.1 \text{ 1/GPa} \tag{6d}$$

Lower Bound:

$$\textit{subject to } 0 \leq C_{11} \leq 125 \text{ GPa} \tag{7a}$$

$$\textit{subject to } 0 \leq C_{12} \leq 75 \text{ GPa} \tag{7b}$$

$$\textit{subject to } 0 \leq S_{11} \leq 0.15 \text{ 1/GPa} \tag{7c}$$

$$\textit{subject to } 0 \leq S_{22} \leq 0.1 \text{ 1/GPa} \tag{7d}$$

It is important to note that the set of constraints are representative examples, and actual constraints may differ from them in the real design. However, it was ensured that the design constraints resembled real-world problems. Apart from the specific set of design constraints for the problem, they should also obey the following generic constraints.

$$\text{A} \geq 0$$

$$\int Adv = 1$$

4. Method

The proposed methodology is divided into two phases. In the first phase, a data repository is created using two sampling algorithms. In the second phase, we evaluate which combinations of ODF dimensions lead to optimal solutions by machine learning. The following flow-diagram illustrates the overall methodology.

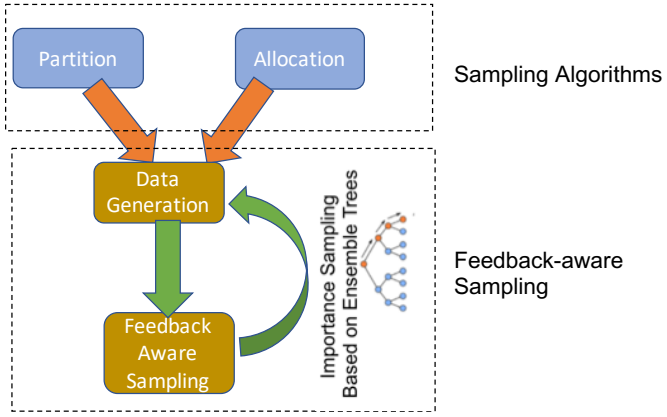


Figure 4: Flow diagram of our methodology. The green arrows depict the data generation process, and the orange arrow signifies the feedback-aware sampling.

4.1. Data Generation

Two data generation algorithms are explored for dataset creation, namely partition and allocation [22]. In the first step, we run our data-generation algorithms to generate around 5 million valid solutions for each set of constraint.

Partition. In this method, a unit length is partitioned into k small segments, where $k > 1$ and can vary between 2 to D. D is the total number of dimensions for the ODF and k is the intended number of non-zero dimensions. For HCP Titanium structure, D is 50 for coarse mesh and 388 for finer mesh. We consider the unit length 1 divided into k random intervals or making k-1 random cuts between the interval [0,1]. For k=2, there is one random cut possible but that cut can be anywhere between [0,1] and we would have an ODF with 2 non-zero dimensions. Similarly for k=3, there be 2 random cuts and so on till k=D. We run the sampling for each value of k, and then for each k, we run it several times so as to generate multiple ODFs for the given number of cuts/dimensions.

Allocation. This randomly generates k values at a time, where k can vary between 1 to D, where D is the number of dimensions for the ODF vector. In this algorithm, k is the intended number of non-zero dimensions for the ODF, and $D - k$ dimensions are set to 0 for the given density function (df) vector. The sum of the product of the volume

fraction (vf) and df across each dimension must add up to 1. Therefore, we continue selecting a value until k values are selected, or the remainder is sufficiently small. k=1 is the trivial case in which the product of the vf and df equal to 1.

The idea behind all the data generation algorithms is based on the heuristic that in a valid microstructure obeying all the constraints, only a few dimensions in the ODF vector are non-zero. However, these methods are complementary to each other and try to sample the entire feature space. While the allocation method tries to find a minimal subset of ODF dimensions that would be non-zero generating a polycrystal solution, the partition method seeks to widen the search across all the dimensions - 50 for Problems 1 and 2, and, 388 for Problems 3 and 4.

4.2. Feedback Aware Sampling

In the first phase of our methodology, we had generated a dataset using two sampling algorithms. In the second phase, we attempt to investigate which combination of non-zero ODF dimensions lead to optimal or near-optimal solutions. For this purpose, we select the top 10 % and bottom 10 % based on the desired design objective and label them as 'High' and 'Low', and perform random forest-based [43] machine learning models on this data subset, where the ODFs become the feature vector. For instance, in design problem 1, as the objective was maximizing the coefficient of expansion α_x, ODF vectors yielding the highest 10 % and bottom 10 % of α_x are labeled as 'High' and 'Low'. Random Forests are ensemble learning methods that construct multiple decision trees [44] to predict the output, and the prediction is decided by a vote across the ensemble of decision trees.

The motivation behind this step is to evaluate ODF dimensions which are important for generating optimal solutions. This step extracts the features that are most important for generating 'High' values. However, as the target is to generate a polycrystalline solution, we proceed to the second iteration of sampling. However, during this step, instead of sampling across all ODF dimensions, we select only those dimensions that are advantageous in providing near-optimal solutions.

Table 1: Number of solutions within 0.01%, 0.02%, 0.05% and 1% of the optimal solutions for the fourth set of constraints

		ML-Guided Sampling			
Bound	**Mesh Size**	within 0.01%	within 0.02%	within 0.05%	within 1%
Upper	388	140	280	759	$1.255\text{x}10^3$
Lower	388	0	$6.223\text{x}10^3$	$1.078\text{x}10^5$	$1.084\text{x}10^5$

Table 2: Comparison of coefficient of expansion α_x, and stiffness parameters (C_{11} and C_{12}) between traditional optimization approaches and ML-Guided Sampling for design problem 1 (Equations 1, 2)

.

Bound	**Mesh Size**	**Linear Programming and Genetic Algorithm**			**ML-Guided Sampling**		
		α_x(in 1/K)	C_{11}(in GPa)	C_{12}(in GPa)	α_x(in 1/K)	C_{11}(in GPa)	C_{12}(in GPa)
Upper	50	$8.5506\text{x}10^{-6}$	161.0000	75.0000	$8.4903\text{x}10^{-6}$	161.0631	75.0450
Upper	388	$8.8560\text{x}10^{-6}$	161.0000	75.0000	$8.8392\text{x}10^{-6}$	161.0519	75.0486
Lower	50	$9.3682\text{x}10^{-6}$	126.6925	90.0000	$9.3790\text{x}10^{-6}$	129.9803	91.6693

5. Results

In this section, we evaluate the data-driven approach proposed in this paper for generating optimal and near-optimal solutions. The proposed method is comparable to solutions produced by prior state-of-the-art techniques and delivers numerous optimal or near-optimal solutions with distinct microstructure designs. The near-optimal solutions for this problem correspond to different microstructure configurations having same or similar values for yield stress. Furthermore, in our study, several different objectives are solved, and the proposed approach is successful for both coarse (50 dimensions) and fine (388 dimensions) meshes for HCP Titanium. Table 1 presents the total number of near-optimal solutions, or in other words, solutions that are proximal to the optimal solutions.

Acar et al. in their previous works [40, 41] used a genetic algorithm based scheme to solve the upper bound problem. In [42], the upper bound approach was transformed to a lower bound approach by converting the problem from stiffness domain to compliance (reciprocal of stiffness) domain and thereby transforming a non-linear problem into a

Table 3: Comparison of stiffness parameters (C_{11} and C_{12}) between traditional optimization approaches and ML-Guided Sampling for design problem 2 (Equations 3, 4)

.

Bound	**Mesh Size**	**Linear Programming and Genetic Algorithm**		**ML-Guided Sampling**	
		C_{11}(in GPa)	C_{12}(in GPa)	C_{11}(in GPa)	C_{12}(in GPa)
Upper	50	167.8562	75.0000	167.8538	75.0013
Upper	388	170.2609	75.0000	169.8015	75.0049
Lower	50	144.2199	95.0000	144.1442	94.9546

Table 4: Comparison of yield stress (σ_y) and compliance parameters (S_{11} and S_{12}) between traditional optimization approaches and ML-Guided Sampling for design problem 3 (Equation 5)

.

Bound	**Mesh Size**	**Linear Programming and Genetic Algorithm**			**ML-Guided Sampling**		
		σ_y(in MPa)	S_{22}(in 1/GPa)	S_{12}(in 1/GPa)	σ_y(in MPa)	S_{22}(in 1/GPa)	S_{12}(in 1/GPa)
Upper	388	423.9396	0.0071	0.0098	423.9396	0.0071	0.0097
Lower	388	423.8462	0.0150	0.1073	422.8327	0.0200	0.0999

Table 5: Comparison of yield stress σ_y, stiffness parameters (C_{11}, C_{12}), and compliance parameters (S_{11} and S_{22}) between traditional optimization approaches and ML-Guided Sampling for design problem 4 (Equations 6, 7)

.

Bound	**Approach**	**Mesh Size**	**Linear Programming and Genetic Algorithm**				
			σ_y(in MPa)	C_{11}(in GPa)	C_{12}(in GPa)	S_{11}(in 1/GPa)	S_{22}(in 1/GPa)
Upper	LP	388	421.8096	175.0000	69.6976	0.0075	0.0095
Upper	ML	388	421.8094	174.9997	69.6976	0.0074	0.0094
Lower	GA	388	423.6050	124.8043	78.3030	0.01612	5.8017×10^{-8}
Lower	ML	388	422.8341	119.8148	80.7035	0.0200	0.0999

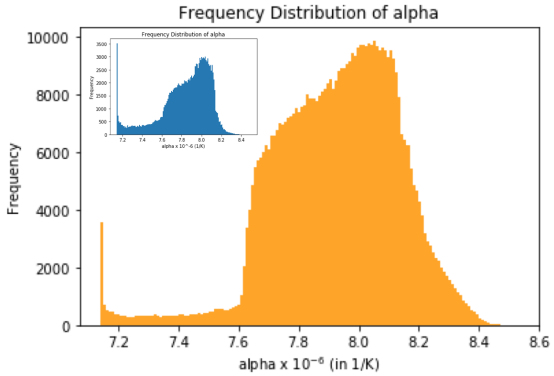
linear problem that is LP-solvable. In [22], a data-driven approach for arriving at a near-optimal solution was expounded for upper and lower bound problems for optimization of the yield stress of cantilevered Galfenol beam under vibrational constraints. The proposed work improves on the previous methodology by identifying a minimal subset of ODF dimensions using machine learning.

For the upper and lower bound approaches, our solutions are compared against the genetic algorithm based scheme and LP-based methods respectively. The proposed data sampling approach based on the sampling algorithms surpassed the yield stress obtained from genetic algorithm based solver for the upper bound approach. Additionally, the results for the lower bound are comparable to the optimal values achieved by the LP method. It is important to note that only the LP solution (used for the lower restricted approach by Acar et al. [42]) yields the theoretical maximum value in contrast to the genetic algorithm solver scheme used by them for the upper bound approach [40].

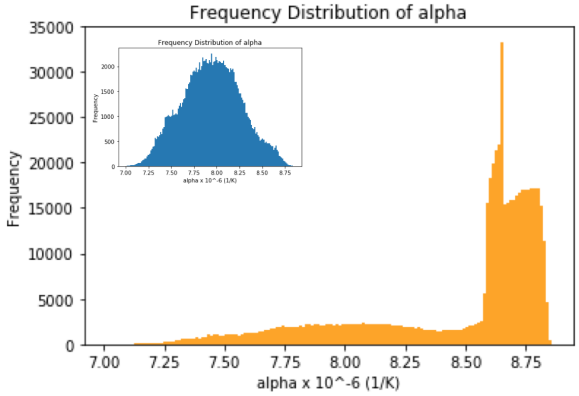
Figures 5,6 represent the frequency distribution for the feedback-driven data-generation of coefficient of expansion and C_{11} for upper (mesh sizes 50 and 388) and lower bounds (mesh size 50) for first set of constraints (Equations 1, 2) for ML-Guided sampling. A comparison of the frequency distribution of the ML-agnostic overall sampling process with the ML-guided sampling indicates the efficacy of the ML-guided approach to effectively extract the non-zero ODF dimensions suitable for more optimal solution. Without an ML-guided approach, it would become increasingly intractable to identify which combinations of ODF dimensions should be non-zero for generating optimal solutions.

Figures 7, 8 illustrate finite element discretized sensitivity ODF cross-sections (mean and standard deviation) and frequency distribution of the maximal desired values across ODF dimensions for design problem 1 and 2. The frequency plots of the ODF dimensions for the top 1% values indicate that there are certain dimensions that are more likely to be non-negative for producing near-optimal solutions as compared to most other dimensions. Thus, ML-guided sampling helps to isolate those dimensions and in particular, isolate those combinations of dimensions that are non-zero for generating near-optimal solutions.

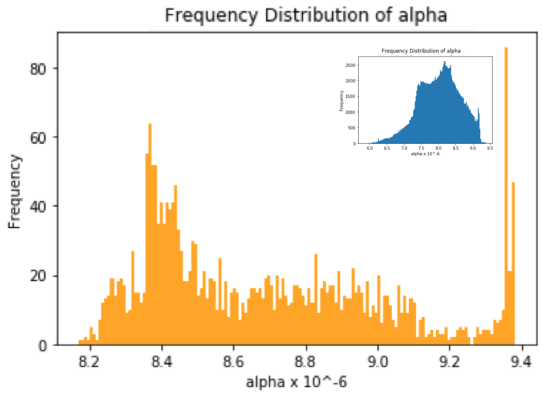
The frequency distribution and sensitivity plots for design problems 3 and 4 are



(a) upper bound (mesh size 50)

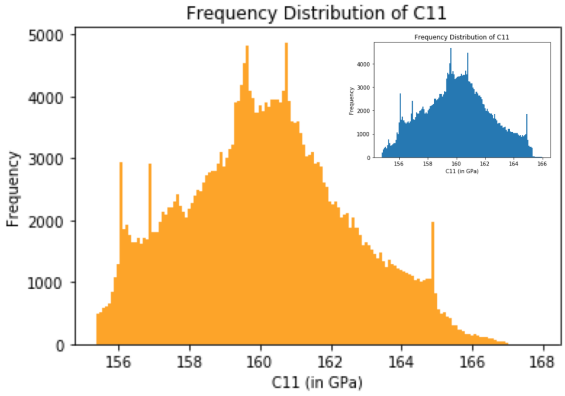


(b) upper bound (mesh size 388)

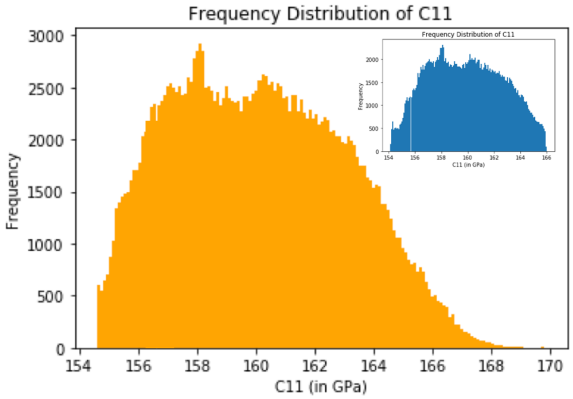


(c) lower bound

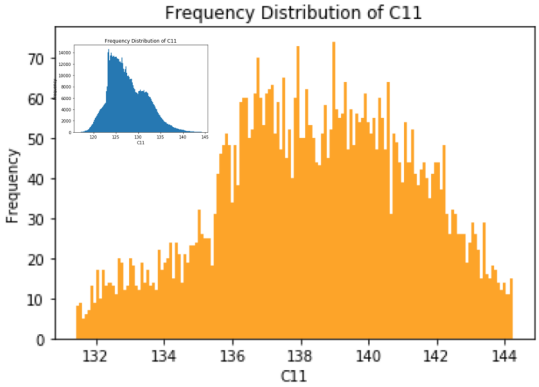
Figure 5: Frequency distribution of coefficient of expansion for upper (mesh sizes 50 and 388) and lower bounds (mesh size 50) for first set of constraints (Equations 1, 2) for ML-Guided sampling. The overall frequency distribution of entire sampling process is presented inset.



(a) upper bound (mesh size 50)

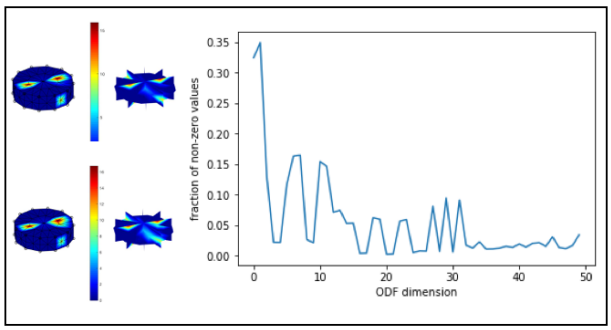


(b) upper bound (mesh size 388)

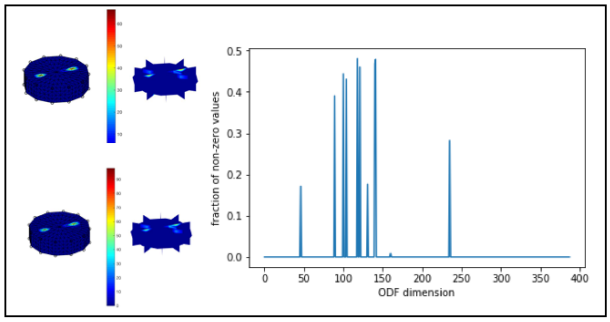


(c) lower bound

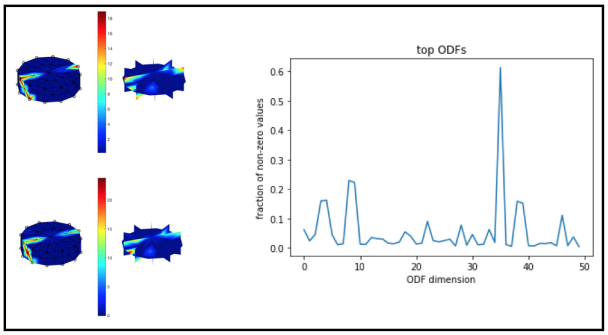
Figure 6: Frequency distribution of C_{11} for upper (mesh sizes 50 and 388) and lower bounds (mesh size 50) for second set of constraints (Equations 3, 4) for ML-Guided sampling. The overall frequency distribution of entire sampling process is presented inset.



(a) upper bound (mesh size 388)

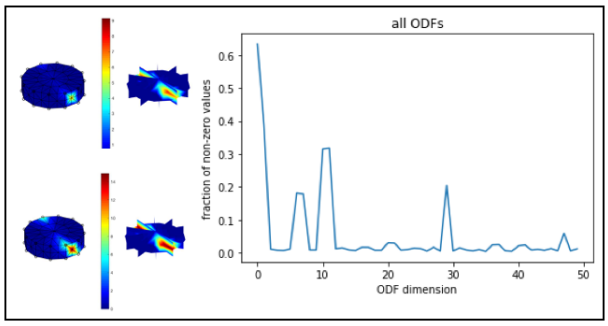


(b) upper bound (mesh size 50)

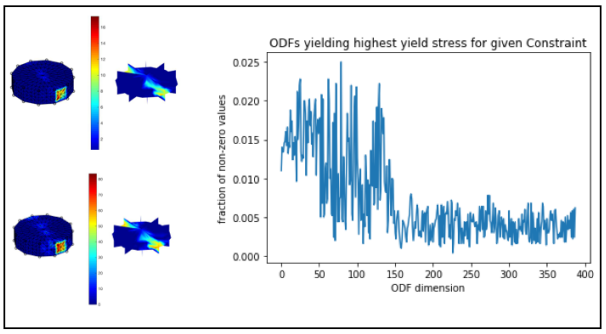


(c) lower bound

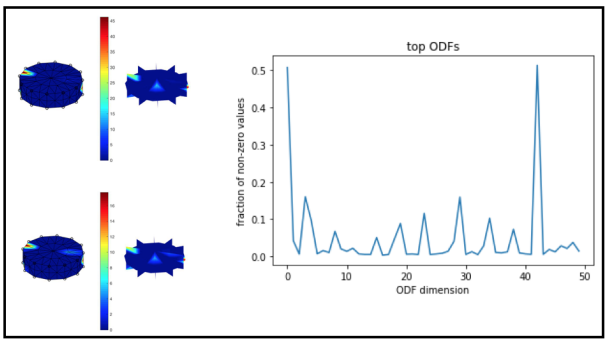
Figure 7: Finite element discretized sensitivity ODF cross-sections (mean and standard deviation) and frequency distribution(inset) of the highest 1% yield stress values across ODF dimensions for design problem 1.



(a) upper bound (mesh size 388)



(b) upper bound (mesh size 50)



(c) lower bound

Figure 8: Finite element discretized sensitivity ODF cross-sections (mean and standard deviation) and frequency distribution(inset) of the highest 1% yield stress values across ODF dimensions for design problem 2.

presented in the Appendix. Examples of finite element microstructure (FEM) cross-sections of near-optimal ODF solutions for all four objective problems are presented in the Appendix. The potential of our methodology to produce many optimal solutions for the upper bound sub-problem in the neighborhood of the LP solution for design problem 1 and 2 for both mesh sizes demonstrate that our method can be advantageous for any mesh size. Nonetheless, more finer the mesh or more the number of dimensions in an ODF, an ML-guided approach becomes more imperative to generate many near-optimal solutions.

6. Conclusion and Future Work

The selection of materials and geometry to optimize desired properties has been a cardinal problem in materials science. The proposed strategy expounds the potential of data-driven approaches for solving a constrained microstructure design objective for both upper and lower bound problems. It outperforms the maximum solutions obtained using Genetic algorithms and is close to the theoretical maximum solution obtained using LP. The proposed targeted sampling approach first explores the entire sample space and then selectively generates solutions that optimize the given design objective. It generates numerous near-optimal solutions, 3 to 4 orders of magnitude higher than prior methods. Past methodologies including LP techniques lead to a unique or handful of optimal solutions. One of the challenges of inverse materials problems is establishing production feasibility of proposed microstructure design. Many cost-aware manufacturing processes can generate specific microstructures and thus, discovering hundreds of thousands of optimal microstructures can help identify more cost-effective candidates for design, thereby accelerating the design to deployment step.

The analysis of constrained microstructure optimization problems depicts that certain combinations of ODF dimensions are non-zero more often in the ODF vector of the near-optimal solutions. Further, a subset of these combinations usually generate ODFs that preferentially produces more near-optimal solutions compared to other combinations. The success of this and similar works reveal the potential of data-driven methods for property predictions of different materials and different design constraints, and for

both upper and lower bound problems. Leveraging data-driven techniques can play an essential role in the expedition of precise design of materials with process constraints. This study demonstrates the power of carefully designed sampling approaches to identify numerous near-optimal solutions for a constrained non-linear optimization problem and can prompt the development of alternative sampling schemes that can reach optimal solutions faster and deliver numerous near-optimal solutions. The sampling schemes are generalizable and agnostic of the problem domain and can be used in other scientific domains as well.

Acknowledgments

This work is supported primarily by the AFOSR MURI award FA9550-12-1-0458. Partial support is also acknowledged from the following grants: NIST award 70NANB14H012; NSF award CCF-1409601; DOE awards DE-SC0007456, DE-SC0014330; and Northwestern Data Science Initiative.

Data Availability

The raw data required to reproduce these findings can be generated by implementing the data-generation algorithms for the given design problems and performing random-forest regression as outlined in the manuscript.

References

[1] G. B. Olson, Computational design of hierarchically structured materials, Science 277 (5330) (1997) 1237–1242.

[2] A. Agrawal, A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science, APL Materials 4 (5) (2016) 053208.

[3] V. Sundararaghavan, N. Zabaras, Linear analysis of texture–property relationships using process-based representations of rodrigues space, Acta Materialia 55 (5) (2007) 1573–1587.

[4] A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, S. R. Kalidindi, Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters, Integrating Materials and Manufacturing Innovation 3 (1) (2014) 1–19.

[5] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, npj Computational Materials 2 (2016) 16028.

[6] P. Deshpande, B. Gautham, A. Cecen, S. Kalidindi, A. Agrawal, A. Choudhary, Application of statistical and machine learning techniques for correlating properties to composition and manufacturing processes of steels, in: Proceedings of the 2nd World Congress on Integrated Computational Materials Engineering (ICME), Springer, 2013, pp. 155–160.

[7] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Physical Review B 89 (9) (2014) 094104.

[8] R. Cang, Y. Xu, S. Chen, Y. Liu, Y. Jiao, M. Y. Ren, Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design, Journal of Mechanical Design 139 (7) (2017) 071404.

[9] R. Liu, A. Kumar, Z. Chen, A. Agrawal, V. Sundararaghavan, A. Choudhary, A predictive machine learning approach for microstructure optimization and materials design, Scientific reports 5.

[10] H. Xu, R. Liu, A. Choudhary, W. Chen, A machine learning-based design representation method for designing heterogeneous microstructures, Journal of Mechanical Design 137 (5) (2015) 051403.

[11] R. Liu, A. Agrawal, W.-k. Liao, A. Choudhary, M. De Graef, Materials discovery:

understanding polycrystals from large-scale electron patterns, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, 2016, pp. 2261–2269.

[12] R. Liu, Y. C. Yabansu, Z. Yang, A. N. Choudhary, S. R. Kalidindi, A. Agrawal, Context aware machine learning approaches for modeling elastic localization in three-dimensional composite microstructures, Integrating Materials and Manufacturing Innovation (2017) 1–12.

[13] W. D. Brewer, R. K. Bird, T. A. Wallace, Titanium alloys and processing for high speed aircraft, Materials Science and Engineering: A 243 (1-2) (1998) 299–304.

[14] V. N. Moiseyev, Titanium alloys: Russian aircraft and aerospace applications, CRC press, 2005.

[15] A. Bratukhin, B. Kolachev, V. Sadkov, et al., Technology of production of titanium aircraft structures, Mashinostroenie, Moscow.

[16] R. Boyer, Titanium for aerospace: rationale and applications, Advanced Performance Materials 2 (4) (1995) 349–368.

[17] A. Machado, J. Wallbank, Machining of titanium and its alloys–a review.

[18] E. Ezugwu, Z. Wang, Titanium alloys and their machinability a review, Journal of Materials Processing Technology 68 (1997) 262–274.

[19] M. J. Donachie, Titanium: a technical guide, ASM international, 2000.

[20] R. Grandhi, S. Modukuru, J. Malas, Integrated strength and manufacturing process design using a shape optimization approach, Journal of Mechanical Design 115 (1) (1993) 125–131.

[21] H. Xu, Y. Li, C. Brinson, W. Chen, A descriptor-based design methodology for developing heterogeneous microstructural materials system, Journal of Mechanical Design 136 (5) (2014) 051007.

[22] A. Paul, P. Acar, R. Liu, W.-k. Liao, A. Choudhary, V. Sundararaghavan, A. Agrawal, Data sampling schemes for microstructure design with vibrational tuning constraints, AIAA Journal 56 (3) (2018) 1239–1250.

[23] A. Heinz, P. Neumann, Representation of orientation and disorientation data for cubic, hexagonal, tetragonal and orthorhombic crystals, Acta Crystallographica Section A: Foundations of Crystallography 47 (6) (1991) 780–789.

[24] V. Randle, O. Engler, Introduction to texture analysis: macrotexture, microtexture and orientation mapping, CRC press, 2000.

[25] U. F. Kocks, C. N. Tomé, H.-R. Wenk, Texture and anisotropy: preferred orientations in polycrystals and their effect on materials properties, Cambridge university press, 2000.

[26] Z. D. Pozun, K. Hansen, D. Sheppard, M. Rupp, K.-R. Müller, G. Henkelman, Optimizing transition states via kernel-based machine learning, The Journal of chemical physics 136 (17) (2012) 174101.

[27] J. Jung, J. I. Yoon, H. K. Park, J. Y. Kim, H. S. Kim, An efficient machine learning approach to establish structure-property linkages, Computational Materials Science 156 (2019) 17–25.

[28] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, A. Agrawal, Elemnet: Deep learning the chemistry of materials from only elemental composition, Scientific reports 8 (1) (2018) 17593.

[29] A. Rahnama, S. Clark, S. Sridhar, Machine learning for predicting occurrence of interphase precipitation in hsla steels, Computational Materials Science 154 (2018) 169–177.

[30] A. Paul, D. Jha, R. Al-Bahrani, W. keng Liao, A. Choudhary, A. Agrawal, Chemixnet: Mixed dnn architectures for predicting chemical properties using multiple molecular representations, in: NeurIPS Workshop on Machine Learning for Molecules and Materials, 2018.
URL https://arxiv.org/abs/1811.08283

[31] A. G. Kusne, T. Gao, A. Mehta, L. Ke, M. C. Nguyen, K.-M. Ho, V. Antropov, C.-Z. Wang, M. J. Kramer, C. Long, et al., On-the-fly machine-learning for high-

throughput experiments: search for rare-earth-free permanent magnets, Scientific reports 4.

[32] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, T. K. Woo, Rapid and accurate machine learning recognition of high performing metal organic frameworks for co2 capture, The journal of physical chemistry letters 5 (17) (2014) 3056–3060.

[33] K. Tsutsui, H. Terasaki, T. Maemura, K. Hayashi, K. Moriguchi, S. Morito, Microstructural diagram for steel based on crystallography with machine learning, Computational Materials Science 159 (2019) 403–411.

[34] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, npj Computational Materials 3 (1) (2017) 54.

[35] M. Mozaffar, A. Paul, R. Al-Bahrani, S. Wolff, A. Choudhary, A. Agrawal, K. Ehmann, J. Cao, Data-driven prediction of the high-dimensional thermal history in directed energy deposition processes via recurrent neural networks, Manufacturing Letters 18 (2018) 35 – 39.

[36] G. Pilania, J. E. Gubernatis, T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, Computational Materials Science 129 (2017) 156–163.

[37] D. Jha, S. Singh, R. Al-Bahrani, W.-k. Liao, A. Choudhary, M. De Graef, A. Agrawal, Extracting grain orientations from ebsd patterns of polycrystalline materials using convolutional neural networks, Microscopy and Microanalysis 24 (5) (2018) 497–502.

[38] H.-J. Bunge, Texture analysis in materials science: mathematical methods, Elsevier, 2013.

[39] A. Kumar, P. Dawson, Computational modeling of fcc deformation textures over rodrigues' space, Acta Materialia 48 (10) (2000) 2719–2736.

[40] P. Acar, V. Sundararaghavan, Utilization of a linear solver for multiscale design and optimization of microstructures in an airframe panel buckling problem, in: 57th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2016, p. 0156.

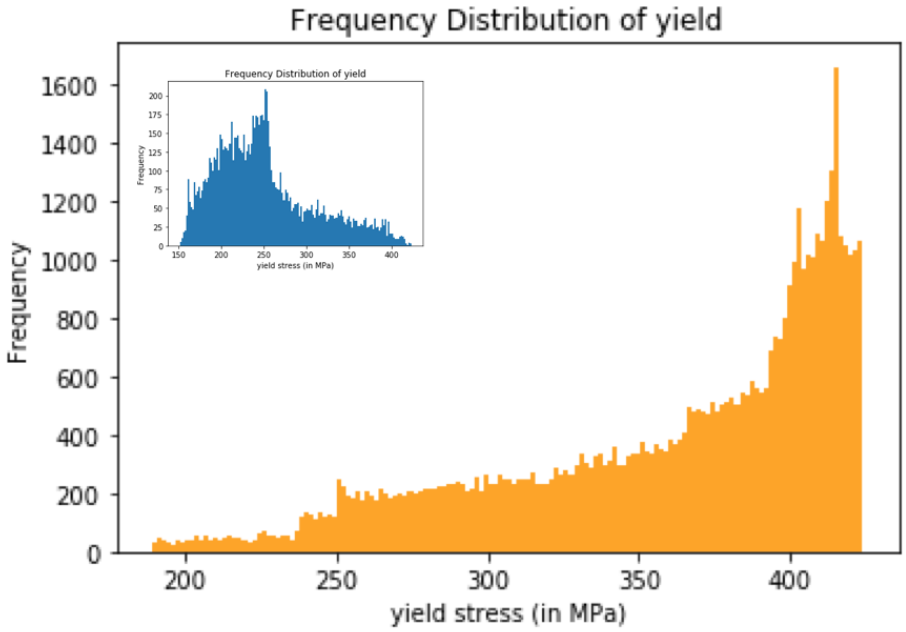
[41] P. Acar, V. Sundararaghavan, Utilization of a linear solver for multiscale design and optimization of microstructures, AIAA Journal 54 (5) (2016) 1751–1759.

[42] P. Acar, V. Sundararaghavan, Linear solution scheme for microstructure design with process constraints, AIAA Journal 54 (12) (2016) 4022–4031.

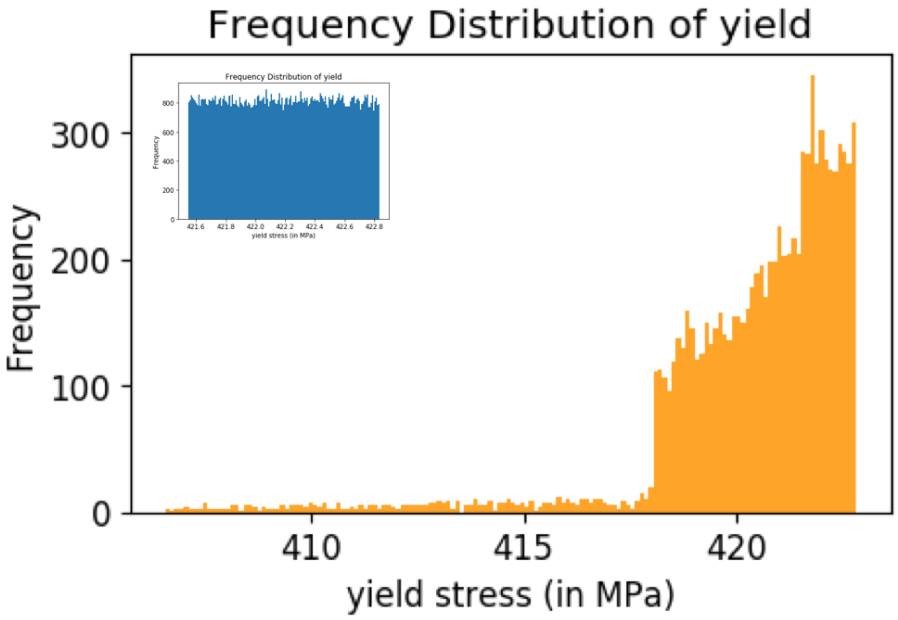
[43] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, R news 2 (3) (2002) 18–22.

[44] T. M. Oshiro, P. S. Perez, J. A. Baranauskas, How many trees in a random forest?, in: International Workshop on Machine Learning and Data Mining in Pattern Recognition, Springer, 2012, pp. 154–168.

Appendix

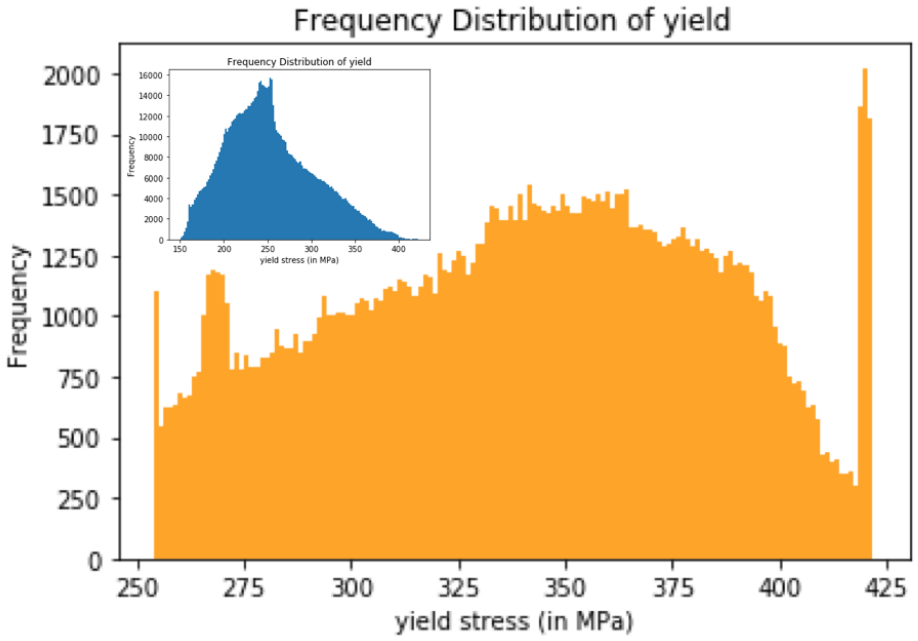


(a) upper bound

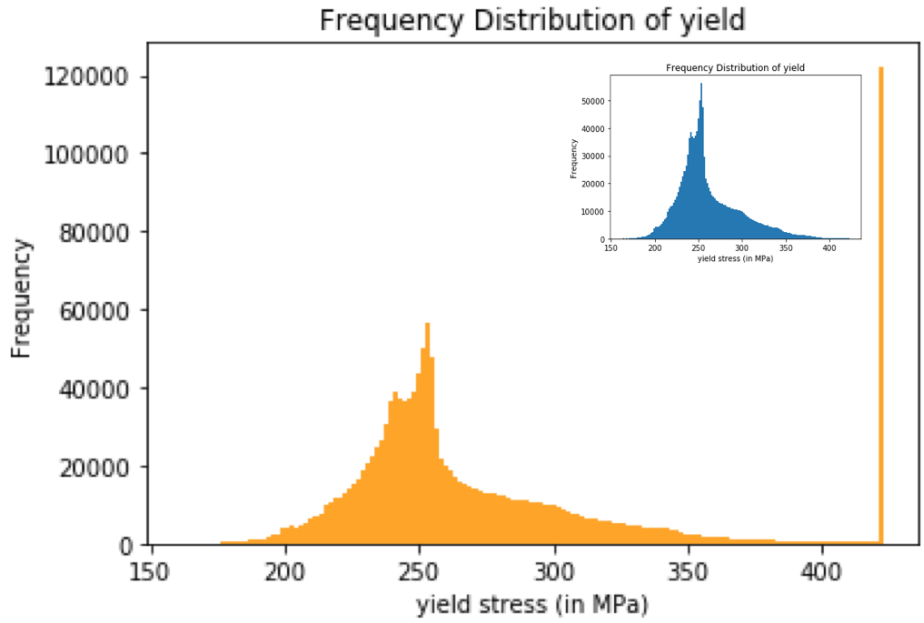


(b) lower bound

Figure 9: Frequency distribution of yield stress for upper and lower bounds for third set of constraints (Equation 5) for ML-Guided sampling. The overall frequency distribution of entire sampling process is presented inset.

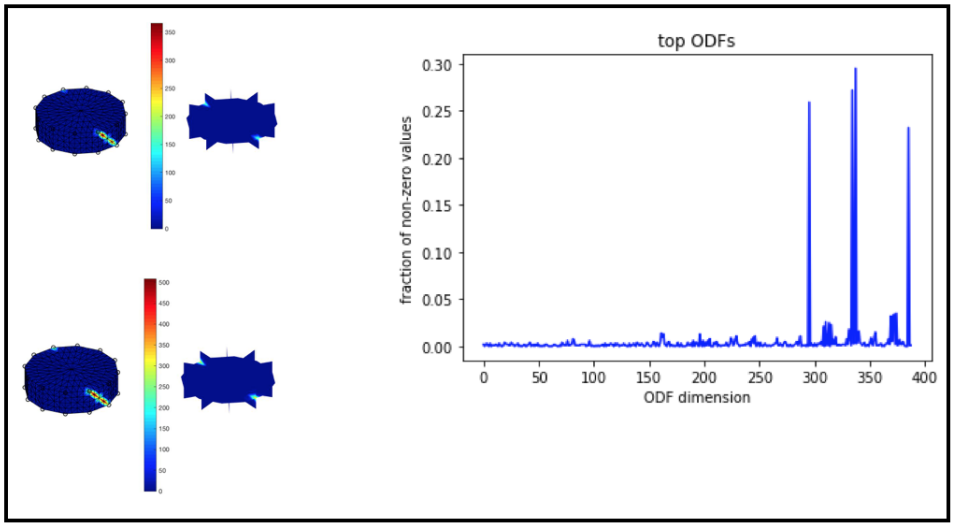


(a) upper bound

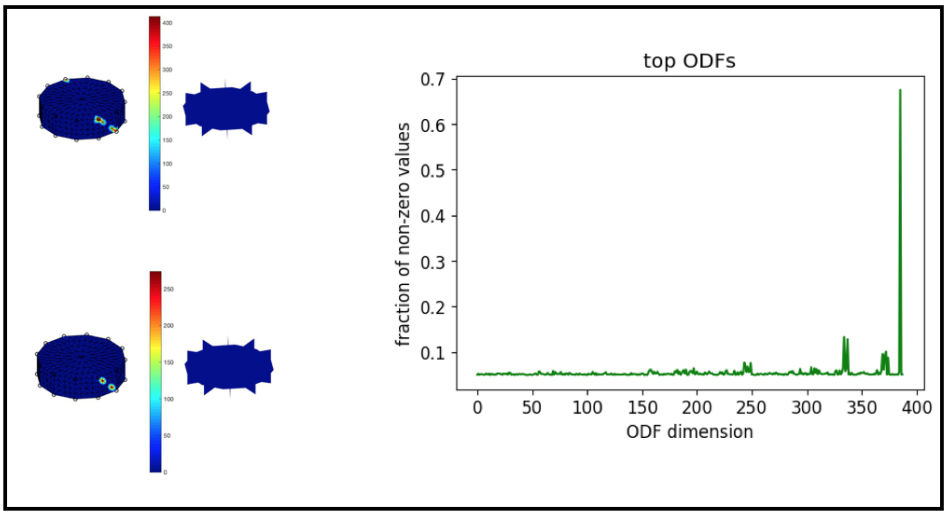


(b) lower bound

Figure 10: Frequency distribution of yield stress for upper and lower bounds for fourth set of constraints (Equations 6, 7) for ML-Guided sampling. The overall frequency distribution of entire sampling process is presented inset.

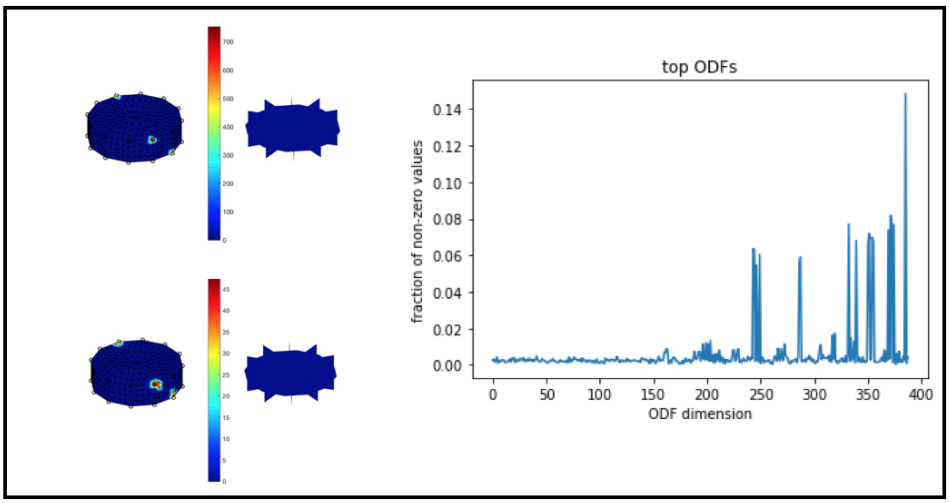


(a) upper bound

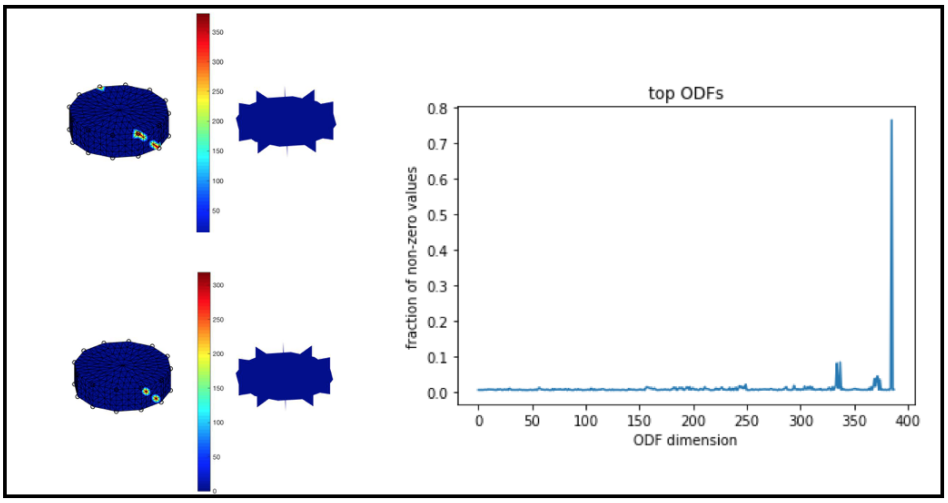


(b) lower bound

Figure 11: Finite element discretized sensitivity ODF cross-sections (mean and standard deviation) and frequency distribution(inset) of the highest 1% yield stress values across ODF dimensions for design Problem 3.

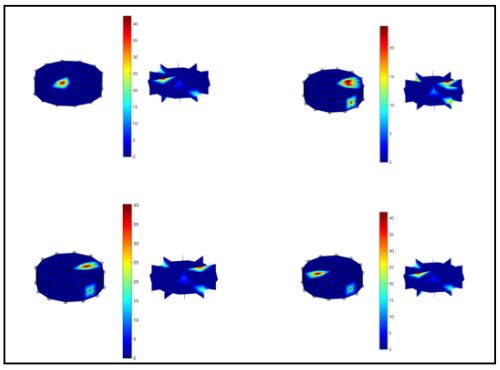


(a) upper bound

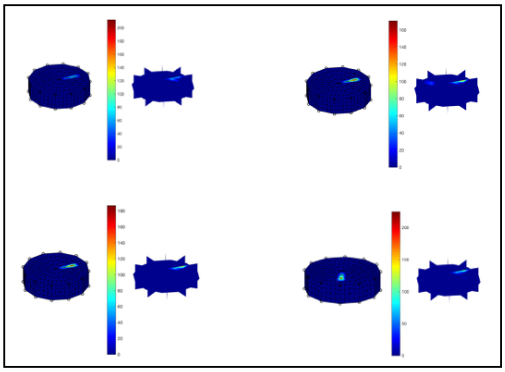


(b) lower bound

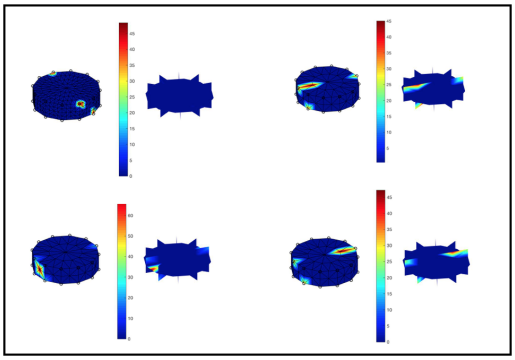
Figure 12: Finite element discretized sensitivity ODF cross-sections (mean and standard deviation) and frequency distribution(inset) of the highest 1% yield stress values across ODF dimensions for design Problem 4.



(a) upper (mesh size 50)

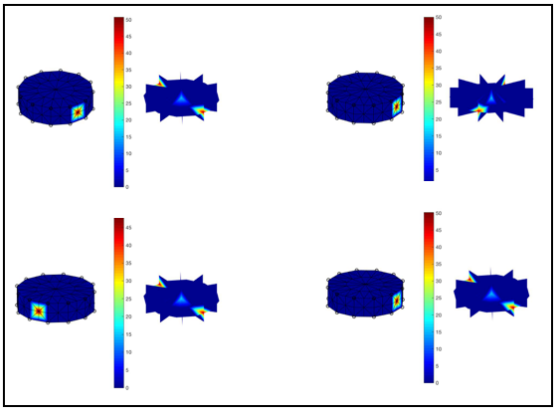


(b) upper (mesh size 388)

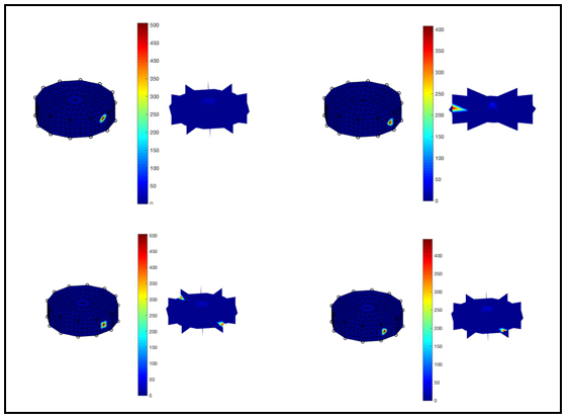


(c) lower

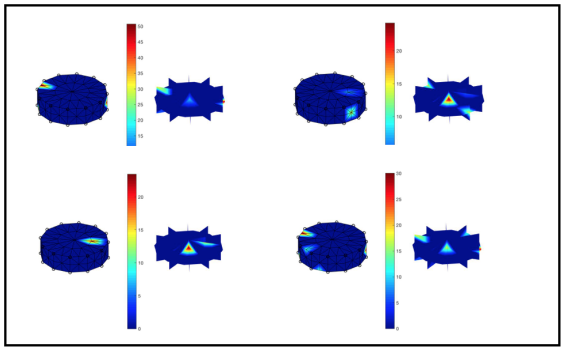
Figure 13: Finite element microstructure cross-sections for examples of near-optimal ODF solutions for design problem 1 (Equations 1, 2)



(a) upper (mesh size 50)

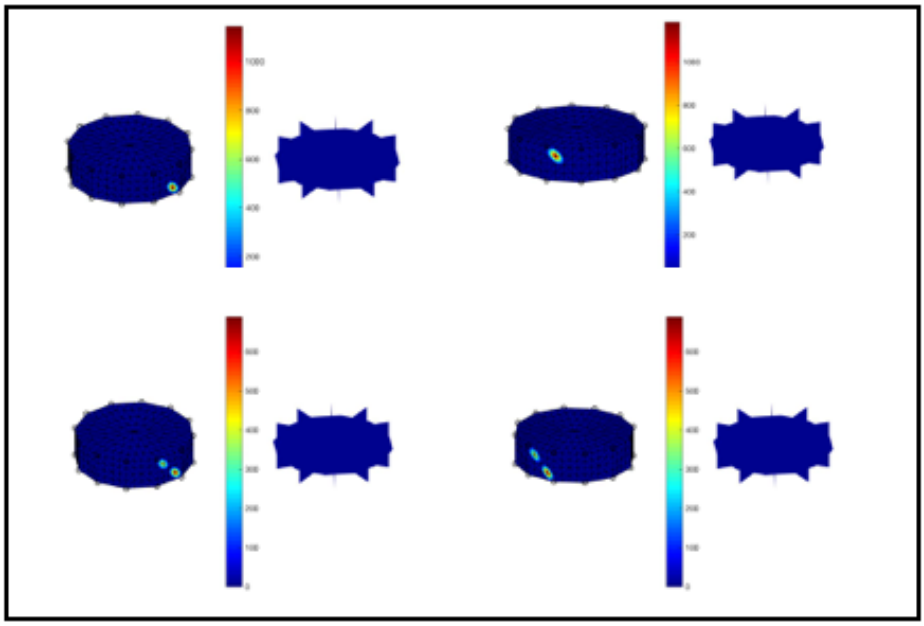


(b) upper (mesh size 388)

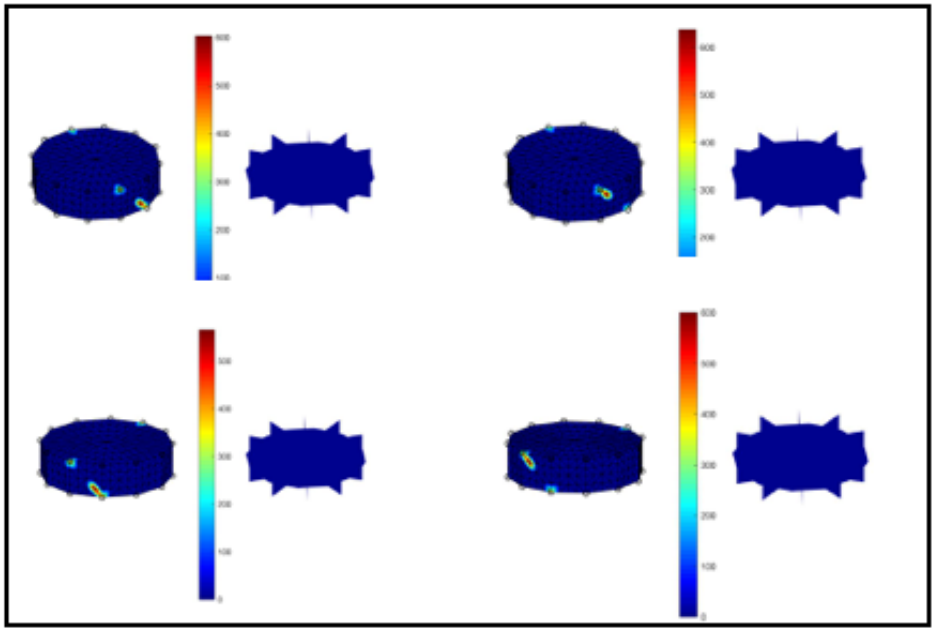


(c) lower

Figure 14: Finite element microstructure cross-sections for examples of near-optimal ODF solutions for design problem 2 (Equations 3, 4)

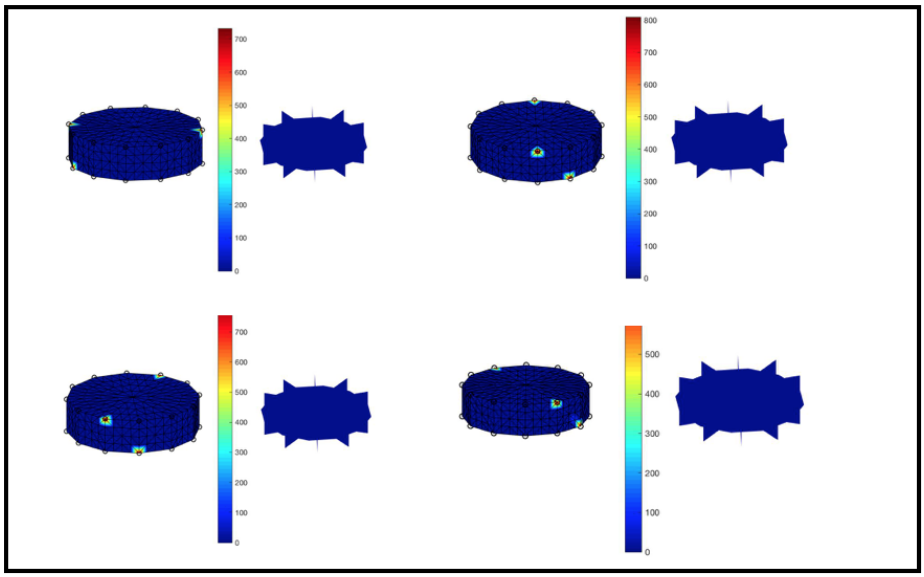


(a) upper

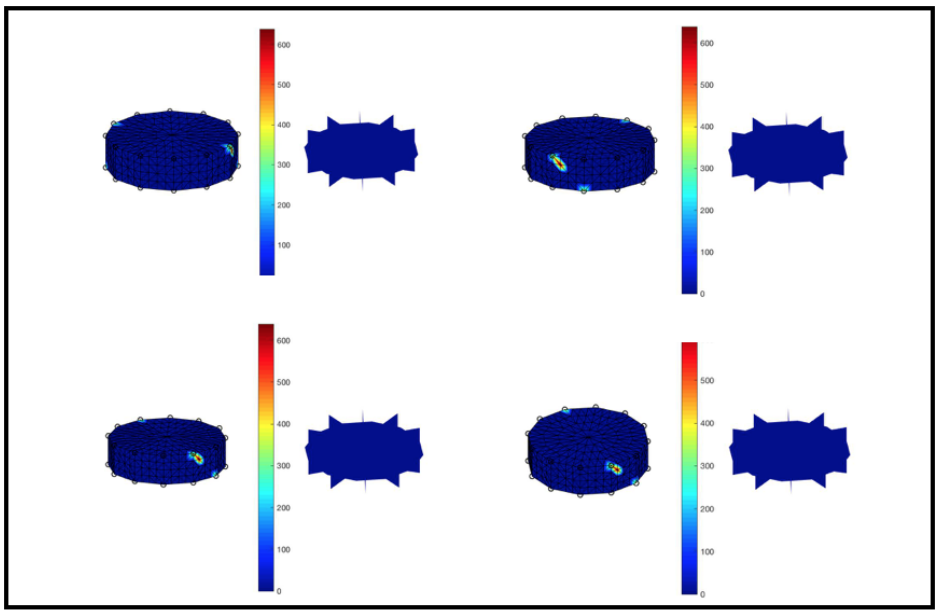


(b) lower

Figure 15: Finite element microstructure cross-sections for examples of near-optimal ODF solutions for design problem 3 (Equation 5)



(a) upper



(b) lower

Figure 16: Finite element microstructure cross-sections for examples of near-optimal ODF solutions for design problem 4