



# Institut für Numerische Simulation

Rheinische Friedrich-Wilhelms-Universität Bonn

Wegelerstraße 6 • 53115 Bonn • Germany  
phone +49 228 73-3427 • fax +49 228 73-7527  
[www.ins.uni-bonn.de](http://www.ins.uni-bonn.de)

James Barker, Gregor Bollerhey and Jan Hamaekers

## A Multilevel Approach to the Evolutionary Generation of Polycrystalline Structures

INS Preprint No. 1525

November 2015



# A Multilevel Approach to the Evolutionary Generation of Polycrystalline Structures

James Barker<sup>a,b,\*</sup>, Gregor Bollerhey<sup>a</sup>, Jan Hamaekers<sup>a</sup>

<sup>a</sup>*Department of Virtual Material Design, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53754 Sankt Augustin, Germany*

<sup>b</sup>*Institute for Numerical Simulation, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany*

---

## Abstract

The Poisson-Voronoi tessellation is commonly used as an approximation to the microstructure of polycrystalline material. Although simple, this approximation fails to respect basic physical properties observed empirically, including the generally lognormal distribution of grain sizes. Stochastic approximations such as genetic algorithms can be used to adjust a Poisson-Voronoi tessellation to better reflect such a distribution. We apply techniques from multilevel optimisation to give a new approach to the evolutionary generation of polycrystalline structures, in a way that allows approximation of a target lognormal grain size distribution with unit mean and arbitrary variance. Results obtained through this method indicate almost perfect distribution fitting, show up to two orders of magnitude improvement in the number of evolutionary steps required for an acceptable fit, and suggest a reduction of the overall problem complexity from  $\Theta(N^3)$  to  $\Theta(N^2)$ .

---

## 1. Introduction

### 1.1. Existing Approaches to Polycrystalline Structure Generation

Polycrystalline materials are, by definition, composed of many inhomogenous *grains*. The characteristics of such materials are deeply related to the collective properties of these grains; their size, shape, relative orientation, etc. It follows that the accuracy and validity of any numerical simulation of the behaviour of a polycrystalline material is founded upon the grain properties entailed by its initial conditions; the properties of the initial model of the structure should match, as closely as possible, those of an empirical sample of the material.

Techniques for the characterisation of such samples have historically been laborious, particularly those that consider all three dimensions [1, 2]. Although recent developments show promising reductions in processing time and can deal with larger samples [3–5], physical analyses can still require several days to complete. Therefore, it is desirable to have an efficient method for the numerical generation of polycrystalline structures, in such a way that the produced microstructure displays statistical similarity to a sample of the material in question.

A common first approximation to a natural grain structure is the Voronoi tessellation of a (periodic) Bravais unit cell  $X \subset \mathbb{R}^3$ , for a finite, uniform-randomly chosen set of *seed points*  $\mathbf{x} = (\mathbf{x}_i \in \mathbb{R}^3)_{i=1}^N$  within that cell [6, 7]. This tessellation associates a subset  $X_i$  of the unit cell with each seed point  $\mathbf{x}_i$ ; these *Voronoi cells* have the defining property that  $X_i = \{\mathbf{y} \in X \mid d(\mathbf{x}_i, \mathbf{y}) \leq d(\mathbf{x}_j, \mathbf{y}) \forall j \neq i\}$ , where  $d(\cdot)$  is an appropriate metric on  $X$ . The metric chosen is usually standard Euclidean distance, calculated to obey the minimum image convention. The Voronoi cells thus produced are convex polytopes, disjoint everywhere except on their boundaries, and are treated as idealised grains for the purposes of simulation.

Although this method is straightforward to implement, the properties of the grain structures it generates are not a good fit to those observed in nature. In particular, the grain sizes generated by a set of uniformly-random seed points follow the Poisson-Voronoi distribution [8], while the grain sizes of polycrystalline structures generally follow a lognormal distribution [9–12]. To address this problem,

---

\*Corresponding author.

Email address: [james.barker@scai.fraunhofer.de](mailto:james.barker@scai.fraunhofer.de) (James Barker)

several stochastic techniques have been proposed to progressively optimise an initial guess at seed point locations, until the grain size distribution of the derived grain structure matches one that would be expected in reality to some “acceptable” degree, as measured by the minimisation of a suitable penalty function.

By nature, grain structures modelled with the Voronoi tessellation cannot capture a number of properties that can be empirically observed in real polycrystalline structures [6]. For example, Voronoi tessellations produce planar grain boundaries, and cannot represent oblique or non-convex grains [13]. A number of approaches have been described that represent grain structures using alternative tessellations [13–16], generally created by optimisation against more sophisticated statistical properties than the grain size distribution. Although we restrict ourselves to a discussion of techniques using the standard Voronoi tessellation and optimising against grain size distribution only, we note that many of these alternatives are also reliant upon randomly-chosen parameters and could potentially replace the standard Voronoi tessellation in the method we describe below.

Gross and Li apply an inverse Monte Carlo technique to approximate a target (lognormal) distribution; they begin with a uniformly-random set of seed points, and iteratively adjust the location of individual seeds by a small amount [10]. After each such adjustment, the quality of the set of seed points is (re-)evaluated by a penalty function  $\chi^2 : \mathbb{R}^{3N} \rightarrow \mathbb{R}_{\geq 0}$ , constructed such that  $\chi^2(\mathbf{x}') < \chi^2(\mathbf{x})$  implies that the distribution of the grain sizes entailed by the set of seed points  $\mathbf{x}'$  is in some sense closer to the target distribution than that of the seed points  $\mathbf{x}$ . If an adjustment yields a sufficiently significant improvement in the penalty function, or if it degrades the distribution fit by less than a random amount chosen under a low, linearly-decreasing ceiling, then it is retained; otherwise, it is discarded. (The allowance of degradation serves to enable the algorithm to escape local minima in the search space, particularly during early phases of the run.) The process continues until the penalty function falls below a preset threshold, at which point the target distribution is considered to be sufficiently approximated. This technique is capable of generating a population of grains exhibiting a lognormal size distribution; however, it is computationally expensive, in that a large number of individual adjustments are required, each of which necessitates a costly recalculation of the Voronoi tessellation; it is also inherently sequential, which limits its performance on modern-day parallel architectures.

Inspired by the approach of Gross and Li, Suzudo and Kaburaki apply a variant of the genetic algorithm to the problem of grain-size distribution fitting [11, 17]. They begin with a *population* of multiple sets of seed points, each distributed in a random fashion across a cubic lattice spanning the unit cell, and progressively “evolve” the population over some number of generations. At each generation, penalty function values are calculated for each member of the population; a fitness function is additionally defined, as the inverse of the penalty function. A new population is then formed, by drawing elements with replacement from the old population in proportion to their relative fitness values.

Some of the seed points in each element of the new population are subjected to mutation, with mutation probability proportional to the size of the surrounding grain. Mutation of a seed point takes one of two forms: either the seed point is relocated randomly to another lattice point within the unit cell, or it is shifted to an adjoining location on the lattice. An elitism strategy is also applied: to ensure that existing good solutions are not discarded, the best few elements of the old population are included in the new population without mutation. The process of generating new populations continues until an element of the new population has a penalty function lower than a preset threshold, in which case convergence is deemed to have occurred, or until a large number of generations have elapsed without convergence.

In this way, Suzudo and Kaburaki are able to generate a grain structure with approximately lognormally-distributed sizes in a more efficient manner than Gross and Li. Their method is also trivially parallelisable in the population size, as the calculation of the penalty function for each member of a population (which requires the calculation of a Voronoi tessellation, the most computationally-intensive component of the algorithm) is completely independent on that of the other members.

However, the basic genetic-algorithm approach still requires a sizeable number of generations to converge, and does not work well on larger sets of seed points. Here, we propose an extension to the approach of Suzudo and Kaburaki, that uses the multilevel transform described in Section 1.2 below to rapidly accelerate the speed of convergence, and attempt to evaluate in a more rigorous way the quality of the resulting grain structure.

## 1.2. Multilevel Optimisation and the Multilevel Transform

In general, a global optimisation problem is characterised by a  $d$ -dimensional objective function  $f : X^d \rightarrow Y$  which takes a global minimum at some (not necessarily unique) point  $\mathbf{x}^* \in X^d$ . The goal of such a problem is then to find the point  $\mathbf{x}^*$  that minimises the value of the objective function; or, failing that, to find an approximate point  $\mathbf{x}$  such that  $f(\mathbf{x})$  is as close to the global minimum as possible. Clearly, the problem outlined in Section 1.1 above is one of global optimisation: we seek to find a set of seed points that best approximates a global minimum of  $\chi^2$ .

Multilevel optimisation is a well-established approach to global optimisation. Consider the family of smooth maps  $C^\infty(X^d \rightarrow X^d)$ . For any initial guess  $\mathbf{x}^{(0)} \in X^d$ , there must be at least one such map  $T^*$  that takes  $\mathbf{x}^{(0)}$  to  $\mathbf{x}^*$ ; that is, there exists some  $T^*$  such that  $f(T^*(\mathbf{x}^{(0)})) = \operatorname{argmin}_{\mathbf{y} \in X} f(\mathbf{y})$ .<sup>1</sup> The original optimisation problem is then equivalent to that of finding either an optimal map  $T^*$ , or another such smooth map  $T$  that offers a sufficiently good approximation to some  $T^*$ .

In the context of the problem of grain-size distribution optimisation, we consider maps of the form  $T \in C^\infty(\mathbb{T}^{3N} \rightarrow \mathbb{T}^{3N})$ , where  $\mathbb{T}^3$  is the 3-torus formed from  $[0, 1]^3$ . Any such map can be represented as a collection of  $N$  functions  $(T_m : \mathbb{T}^3 \rightarrow \mathbb{T}^3)_{m=1}^N$ , which act elementwise on a collection of seed points  $S = (\mathbf{s}_m \in \mathbb{T}^3)_{m=1}^N$ , so that  $T(S) = (T_m(\mathbf{s}_m))_{m=1}^N$ . These seed points represent points in a Bravais unit cell  $X \subset \mathbb{R}^3$ , specified in fractional coordinates with respect to some set of lattice parameters. We seek to find or approximate an optimal map  $T^*$ , which takes some initial set of seed points  $S^{(0)}$  to an optimal collection of seed points  $S^*$ , which is mapped to a set of points  $\mathbb{R}^{3N}$  that minimises  $\chi^2 : \mathbb{T}^{3N} \rightarrow \mathbb{R}_{\geq 0}$ .

Because the order of seed points within the collection is arbitrary, we assume the penalty function  $\chi^2$  is invariant under permutation; that is, for any permutation  $P \in \Sigma_N$ , we have  $\chi^2(S) = (\chi^2 \circ P)(S)$ . We are interested only in maps that preserve this property, such that  $(\chi^2 \circ P \circ T)(S) = (\chi^2 \circ T \circ P)(S)$  for all  $P$ . One way of ensuring this is to assume that  $T_m = T_n$  for any choice of  $m$  and  $n$ , so that the same function is applied to each seed point, regardless of order; we denote this function  $t : \mathbb{T}^3 \rightarrow \mathbb{T}^3$ , and write  $t(\mathbf{s}) = (t_1(\mathbf{s}), t_2(\mathbf{s}), t_3(\mathbf{s}))$  where  $\mathbf{s} \in \mathbb{T}^3$  is any seed point and  $t_j : \mathbb{T}^3 \rightarrow \mathbb{T}$ . Given an optimal set of distinct seed points  $S^*$ , and an arbitrary initial guess  $S^{(0)}$ , at least one optimal  $T^* = (t^*)_{m=1}^N$  is guaranteed to exist.<sup>2</sup>

Under the further assumption that  $s_j \neq 0$ , the functions

$$g_j(\mathbf{s}) = \frac{t_j(\mathbf{s})}{s_j} - 1 \quad (1)$$

are well-defined on  $\mathbb{T}^3$ , and can be written as three-dimensional Fourier expansions

$$g_j(\mathbf{s}) = \sum_{\mathbf{k} \in \mathbb{Z}^3} \mathbf{c}_{\mathbf{k},j} e^{2i\pi \mathbf{k}^T \mathbf{s}}, \quad (2)$$

where each  $\mathbf{c}_{\mathbf{k}} \in \mathbb{C}^3$  is the vector of Fourier coefficients corresponding to the multi-dimensional index  $\mathbf{k}$ . Rewriting, we obtain

$$t_j(\mathbf{s}) = s_j \left( 1 + \sum_{\mathbf{k} \in \mathbb{Z}^3} \mathbf{c}_{\mathbf{k},j} e^{2i\pi \mathbf{k}^T \mathbf{s}} \right), \quad (3)$$

or alternatively, using  $\odot$  to indicate component-wise multiplication,

$$t(\mathbf{s}) = \mathbf{s} + \mathbf{s} \odot \sum_{\mathbf{k} \in \mathbb{Z}^3} \mathbf{c}_{\mathbf{k}} e^{2i\pi \mathbf{k}^T \mathbf{s}}. \quad (4)$$

Expressed this way,  $t$  has the desirable property of ensuring that a set of zero-valued Fourier coefficients provides the identity transformation.<sup>3</sup> The original optimisation problem is then equivalent to finding a set of Fourier coefficients that minimises the difference between  $\chi^2(T(S))$  and  $\chi^2(S^*)$ .

---

<sup>1</sup>The constant map  $T^*(\mathbf{x}) = \mathbf{x}^*$  is a trivial example.

<sup>2</sup>Fix  $j \in \{1, 2, 3\}$ . Let  $\varphi_j = \{\varphi_{mj} : \mathbb{T}^3 \rightarrow \mathbb{T} \mid 1 \leq m \leq N\}$  be a set of smooth bump functions chosen such that  $\varphi_{mj}(\mathbf{s}_m^{(0)}) = s_{mj}^*$ . Because the seed points are distinct, these functions can be chosen such that  $\operatorname{supp}(\varphi_m) \cap \operatorname{supp}(\varphi_n) = \emptyset$  for  $m \neq n$ . Now set  $t_j(\mathbf{s}) = \sum_{m=1}^N \varphi_{mj}(\mathbf{s})$ . Because each  $\varphi_{mj}$  is smooth, so is  $t_j$ , as is  $t(\mathbf{s}) = (t_1(\mathbf{s}), t_2(\mathbf{s}), t_3(\mathbf{s}))$ . If  $T = (t)_{m=1}^N$ , then by construction,  $T(S^{(0)}) = S^*$ , and  $T$  is optimal.

<sup>3</sup>This property is “desirable” in the context of several other related applications of multilevel optimisation; we apply this representation here for consistency. There are, of course, many other choices for the representation of  $t$ .

In order for the problem to be computationally tractable, we must consider truncated expansions of the Fourier series in the functions  $t_j$ ; that is, we restrict the indices  $\mathbf{k}$  to an index set of level  $K$ ,

$$I_K = \{\mathbf{k} \in \mathbb{Z}^3 \mid \|\mathbf{k}\|_\infty \leq K\}, \quad (5)$$

and write

$$t(\mathbf{s}) \approx t^K(\mathbf{s}) = \mathbf{s} + \mathbf{s} \odot \sum_{\mathbf{k} \in I_K} \mathbf{c}_\mathbf{k} e^{2i\pi \mathbf{k}^\top \mathbf{s}}. \quad (6)$$

Results in sparse grid methods [18–21] have shown that rather than using the  $\|\cdot\|_\infty$  norm in Equation (5), acceptable accuracy can be obtained simply with the  $\|\cdot\|_1$  norm. Applying this refinement to the construction of the index set  $I_K$  allows non-trivial computational savings.

A large amount of computation can be further avoided by noting that because the codomain of  $t$  is  $\mathbb{R}^3$ , it must be the case that  $\mathbf{c}_\mathbf{k} = \overline{\mathbf{c}_{-\mathbf{k}}}$ ; this allows us to write

$$t^K(\mathbf{s}) = \mathbf{s} + \mathbf{s} \odot \mathbf{c}_0 + \mathbf{s} \odot \sum_{\mathbf{k} \in I'_K} 2\mathbf{c}_\mathbf{k} e^{2i\pi \mathbf{k}^\top \mathbf{s}} \quad (7)$$

where  $I'_K = \{\mathbf{k} \in I_K \mid -\mathbf{k} <^d \mathbf{k}\}$ , using  $<^d$  to denote the standard lexicographic (total) ordering on  $\mathbb{Z}^3$ . We refer to the map  $T^K = (t^K)_{m=1}^N$  as the *multilevel transform of order  $K$* .

Varying the order  $K$  of a multilevel transform enables different types of topological transformations to be applied to the input set of seeds  $S$ . At order  $K = 0$ , the transform describes a linear scaling and translation. As the order  $K$  increases, the transform acts more and more distinctly on each individual point in the initial guess. As  $K$  approaches infinity, every point in the initial guess can be considered to move (or not move) in space in a manner almost completely decoupled from the other points. Such a transform is sufficient to capture the behaviour of an iteration of the genetic algorithm described above, regardless of whether each seed point is moved slightly, randomised, or left static. Therefore, the set of all possible multilevel transforms of any order (or equivalently, the set of all possible coefficient matrices of any size) can be considered a superset of all possible outcomes of an iteration of the genetic algorithm.

Multilevel optimisation is often paired with other optimisation techniques, whereby the set of coefficients  $\{\mathbf{c}_\mathbf{k}\}$  is optimised with respect to the output of the penalty function, as applied to the same original input. In the context of genetic algorithms, we propose a different usage. At every generation, we allow some or all of the members of the new population to undergo a multilevel transform of some order  $K$ , instead of the standard mutation process of Suzudo and Kaburaki. This transform is described by a matrix of randomly-chosen coefficients. The intention is to allow a wider range of exploration options over the search space, and (particularly in the case of lower orders  $K$ ) to take advantage of potentially-beneficial topological changes which are vanishingly unlikely to occur through standard mutation.

## 2. Method

### 2.1. Basic Algorithm

We apply a modified variant of the standard genetic algorithm approach. A starting *population* is created, consisting of some number of *chromosomes*. Each chromosome is simply a set of seed points  $S$  as described in Section 1.2 above; initially, chromosomes are populated in a uniformly-random fashion from the 3-torus. Under transformation by the coordinate matrix composed of lattice vectors describing a Bravais unit cell  $X \subset \mathbb{R}^3$ , each chromosome maps to a set of  $N$  points  $\mathbf{x} = (x_i)_{m=1}^N \in X$ . The suitability of a chromosome can be evaluated according to a penalty function  $\chi^2 : \mathbb{T}^{3N} \rightarrow \mathbb{R}_{\geq 0}$ ; the precise choice of the penalty function is discussed in Section 2.3 below.

After each iteration, all candidate chromosomes are ranked according to a fitness function, defined as the inverse of the penalty function,  $1/\chi^2(\mathbf{x})$ . A pre-determined number of the best-ranked of these candidates are termed *elite*; these chromosomes are carried over into the new population unchanged. The remaining chromosomes in the new population are then created as *mutations* of members of the old population.<sup>4</sup>

---

<sup>4</sup>We concur with Suzudo and Kaburaki [11] that generating the new population through crossover provides no significant

The process of creating a new mutant is as follows. A chromosome from the old population is selected and replicated into the new population, with probability of selection proportional to its fitness function value. The selected chromosome is then subjected to either *standard* or *multilevel* mutation, according to a probability  $p_{\text{ml}}$ .

Under standard mutation, the individual seeds of each chromosome are either mutated or left unchanged, according to a probability  $p_{\text{mut}}$ , called the *mutation rate*. If mutated, they are either reset to a uniformly random location within the simulation volume, or shifted slightly by the addition of a “jitter vector”  $j \in \mathbb{R}^3$ , with  $j_i \sim \mathcal{N}(0, \sigma_{\text{jitter}}^2)$ . The probabilities of each mutation type occurring are given by  $p_{\text{jitter}}$  and  $p_{\text{randomise}}$ , with the obvious restriction that these two probabilities sum to unity.

Under multilevel mutation, the entire chromosome is mutated according to a randomised multilevel transformation of some order  $K$ . An  $|I'_K + 1| \times 3$  matrix of complex Fourier coefficients  $C$  is formed; the real and imaginary parts of these coefficients are uncorrelated, and are drawn separately from a normal distribution  $\mathcal{N}(0, \sigma_{\text{coeff}}^2)$ . The transformed chromosome  $S' = (T^K(\mathbf{s}_m))_{m=1}^N$  then becomes part of the new population.

Once the new population has been completely generated, the penalty function for each new element is evaluated. If the “best” (i.e. lowest) penalty function falls below some pre-set threshold  $\chi_{\text{conv}}^2$ , then the algorithm is considered to have “converged”, and terminates, returning the corresponding chromosome as the desired result; otherwise, a new population is created once more, and the process repeats.

## 2.2. Adaptive Mutation

There are several parameters in our mutation algorithm that can be tuned. They include:

1.  $p_{\text{mut}}$ , the per-seed mutation rate for standard mutation.
2.  $p_{\text{jitter}}$  and  $p_{\text{randomise}}$ , the probabilities of the different types of standard mutation.
3.  $\sigma_{\text{jitter}}$ , the standard deviation of the normal distribution used to generate jitter vectors.
4.  $p_{\text{ml}}$ , the likelihood of a new population member undergoing multilevel mutation instead of standard mutation.
5.  $K$ , the order of multilevel mutation.
6.  $\sigma_{\text{coeff}}$ , the standard deviation of the normal distribution used to generate multilevel coefficients.

It is useful to vary some or all of these parameters over the course of an evolutionary run. In particular, it is useful to apply larger, more aggressive mutations early in the run, to allow wider exploration of the search space from the initial populations, and to then switch to finer-scale mutations as the population elements approach good solutions.

In their test case, Suzudo and Kaburaki apply a change of (standard) mutation rate from  $p_{\text{mut}} = 0.1$  to  $p_{\text{mut}} = 0.001$  at the 500th generation; this rule was derived from repeated observation of the process of their algorithm. In general, it would be better to have an automatic approach, in the hope of assuring satisfactory performance without the need for repeated inspection of the behaviour of the algorithm on a particular case. To this end, we apply a strategy of *adaptive mutation*.

Under adaptive mutation, a series of threshold events are scheduled before runtime. As mutation proceeds, the algorithm maintains awareness of its past behaviour, particularly the previous best values of the penalty function  $\chi^2$ . The presence of at least one elite in each new population guarantees that the best value of  $\chi^2$  decreases monotonically over time; however, it is possible (and common) that multiple generations can pass before the best value of  $\chi^2$  changes. This is referred to as a *plateau*, and for a given set of parameters, becomes more likely over time. Once a plateau has continued for a preset number of generations, the algorithm assumes that no further benefit will be obtained from the current settings; the next threshold event is applied, adjusting one or more of the parameters above, usually towards a finer mutation type.

---

benefit. We hypothesise that this is due to the nature of the chromosome, which encodes information only about the location of seed points and not about their spatial interrelationships. As the size of the Voronoi cell surrounding each seed point is determined entirely by the locations of some subset of the other seeds within the chromosome (i.e. those which share a cell wall with the original seed), and no guarantees can be made about their position within the chromosome, any kind of crossover results in unpredictable changes to potentially all the Voronoi cells in the parent individuals. From this perspective, a crossover is simply another form of random mutation.

### 2.3. Choice of Penalty Function

The choice of penalty function  $\chi^2$  is critical to the behaviour of the algorithm. The original formulation provided by Gross and Li [10] is

$$\chi^2(S) = \frac{1}{N} \sum_{m=1}^N [P_r(d_m) - P(d_m)]^2, \quad (8)$$

where  $N$  is the number of seed points in the chromosome  $S$ ,  $d_m$  is the size of the grain surrounding the (transformed) seed point  $x_m$ ,  $P_r$  is the desired lognormal probability density function

$$P_r(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right], \quad (9)$$

with location and scale parameters  $\mu, \sigma \in \mathbb{R}$ ,  $\sigma > 0$ , and  $P$  is a probability density function representing the distribution of the grain sizes in the configuration. We generate tessellations using the Voro++ library [22], and follow the example of Gross and Li and Suzudo and Kaburaki by normalising grain sizes such that they have unit mean.

Both Gross and Li and Suzudo and Kaburaki appear to estimate  $P$  as a histogram of the obtained grain sizes, with frequencies normalised such that the total area of the bins is unity. Although simple, there are a number of issues with this approach, the most notable of which is the difficulty of choosing an appropriate bin size. If too large or too small a bin size is chosen, then the underlying probability distribution will be under- or oversampled respectively. In general, there is no single best choice of bin size for a histogram representing an unknown distribution, although a number of heuristics exist. This is also true under the assumption that the data approximate a lognormal distribution with unknown parameters.

As an alternative, we use as  $P_r$  and  $P$  the cumulative distribution function (CDF) of the target lognormal distribution and the empirical CDF (ECDF) of the obtained grain size distribution, given respectively as

$$\text{CDF}(d_m) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln(d_m) - \mu}{\sqrt{2}\sigma}\right), \quad (10)$$

and

$$\text{ECDF}(d_m) = \frac{1}{N} \sum_{n=1}^N a_{mn}, \quad (11)$$

where  $a_{mn}$  is unity if  $d_n \leq d_m$  and zero otherwise. This approach has the advantage of not requiring any *a priori* assumptions about optimal bin size, and is only marginally less efficient than the calculation of a histogram; the calculation of the ECDF of the grain size values requires  $O(N \log(N))$  operations and is still much cheaper than the evaluation of the Voronoi tessellations.

## 3. Results

### 3.1. Test Cases

We tested our algorithm over three test cases. The first was that used by Gross and Li and Suzudo and Kaburaki: optimisation of the grain structure of a simple-cubic polycrystal, to fit a lognormal grain-size distribution with parameters  $\mu = 0$  and  $\sigma = 0.35$ .

The second test case was a similar optimisation, but against a lognormal distribution with  $\mu = -0.06125$  and  $\sigma = 0.35$ . The non-zero choice of  $\mu$  sets the mean of the target lognormal distribution to be precisely unity, rather than  $\exp\left[\frac{0.35^2}{2}\right] \approx 1.063$  as is the case with  $\mu = 0$  – the significance of this is explained in Section 3.2 below.

The third test case was a triclinic polycrystal, with lattice parameters  $a = 7.48\text{\AA}$ ,  $b = 9.95\text{\AA}$ ,  $c = 7.68\text{\AA}$ ,  $\alpha = 111.65^\circ$ ,  $\beta = 115.383^\circ$ ,  $\gamma = 69.433^\circ$ . These parameters match one informal presentation of the crystalline structure of turquoise[23] – the precise choice of parameters is unimportant, as the goal of the test case is only to examine the behaviour in a sample triclinic structure. The grain structure of this cell was fitted against a target lognormal distribution with  $\sigma = 0.4$  and  $\mu = -0.08$ , again with  $\mu$  chosen to ensure unit mean.

N. Seeds	500	1000	1500	2000	2500	3000
Crit. Value	0.087183	0.061648	0.050335	0.043592	0.038989	0.035592

Table 1: Critical values for Kolmogorov-Smirnov goodness-of-fit test,  $\alpha = 0.001$ , by number of seeds.

Test cases were fitted for numbers of seeds  $N$  between 500 and 3000, increasing in increments of 500. Power-of-two population sizes were tested, from 8 up to 128. In every case, the number of elites was set at one-eighth of the population size.

Four mutation strategies were tested. The first strategy was that given by Suzudo and Kaburaki, i.e. no multilevel mutation, equal likelihood of jitter and randomisation, and mutation rate set at 0.1 for the first 500 generations and 0.001 thereafter.

The second strategy was adaptive, with a plateau length of 10 generations, using only standard mutation (so  $p_{\text{ml}} = 0$ ). Mutations are performed with  $p_{\text{jitter}} = 0.8$ , and  $p_{\text{randomise}} = 0.2$ ; both  $p_{\text{mut}}$  and  $\sigma_{\text{jitter}}$  are varied over the range 0.1 to 0.001, decreasing by an order of magnitude after each plateau with  $\sigma_{\text{jitter}}$  varying more frequently (that is,  $p_{\text{mut}} = 0.1$  while  $\sigma_{\text{jitter}}$  varies from 0.1 down to 0.001, then again with  $p_{\text{mut}} = 0.01$ , and so on).

The third strategy was also adaptive, with an identical plateau length. Both multilevel and standard mutations were performed, with the probability of multilevel mutation set at  $p_{\text{ml}} = 0.25$ . Multilevel mutations are performed from order 1 up to order 5, with  $\sigma_{\text{coeff}}$  for each varying over the range 0.01 to 0.0001, decreasing by an order of magnitude after each plateau. Beyond order 5, the size of the index set  $I'_K$  increases dramatically and multilevel mutations become prohibitively expensive in both computation and memory. Standard mutations are performed as in the strategy above.

The fourth strategy was similar to the third, but with only multilevel mutation allowed to take place (so  $p_{\text{ml}} = 1$ ).

All test cases were allowed to run for  $10^4$  generations, with no prior convergence threshold set (i.e.  $\chi^2_{\text{conv}} = 0$ ). At every generation, the normalised grain sizes of the population member with the lowest penalty functions were tested for goodness-of-fit against the target distribution using the one-sample Kolmogorov-Smirnov test (as implemented in the SciPy package for scientific computing [24]), generating two-sided  $p$ -values against the null hypothesis that the obtained grain sizes follow the target distribution. Significance was set at  $\alpha = 0.001$ , leading to critical values for the various numbers of seeds tested as given in Table 1.

Each test case was repeated over an ensemble of 32 individually-seeded populations; all results (best penalty functions, etc.) are given as mean values over the elements of the ensemble unless otherwise specified.

### 3.2. Discussion

We begin by investigating the behaviour observed for the Gross-Li/Suzudo-Kaburaki parameter set ( $\sigma = 0.35$ ,  $\mu = 0$ ). Graphs displaying the ensemble-averaged best penalty function for  $N = 2000$  can be seen in Figure 1. In all cases and for all values of  $N$ , the best penalty function decreased from a starting value slightly above  $10^{-2}$  to a final value between  $10^{-3}$  and  $10^{-4}$  after  $10^4$  generations. The two mutation strategies involving multilevel mutation (cf. the lower two of the four graphs) show rapid decrease up to  $10^2$  generations or slightly less, and only gradual decrease thereafter; by contrast, the non-multilevel mutation strategies show more consistent rates of decrease throughout the generations. The size of the population appears not to have a significant impact on either the best penalty function value or the rate of decrease.

The Kolmogorov-Smirnov (KS) test results for this case do not indicate good behaviour in terms of distribution fitting. (For ensemble-averaged KS test values for  $N = 2000$ , cf. Figure 2.) For the standard-only mutation strategies, the average  $p$ -values obtained increase slowly and roughly linearly from zero to between  $10^{-2}$  (for 500 seeds) and  $10^{-6}$  (for 3000 seeds) from the thousandth generation onwards; by comparison, the average  $p$ -values for the multilevel strategies show similar increase between  $10^1$  and  $10^2$  generations, remain roughly static until  $10^3$  generations, and generally begin to slowly decrease again. Cases for higher numbers of seeds increase more slowly; for example, for  $N = 3000$ , KS test values are still increasing at the  $10^4$ th generation. There are very few places in the evolutionary process where these values exceed the critical values in Table 1, and even then only for the case of 500 seeds; therefore,

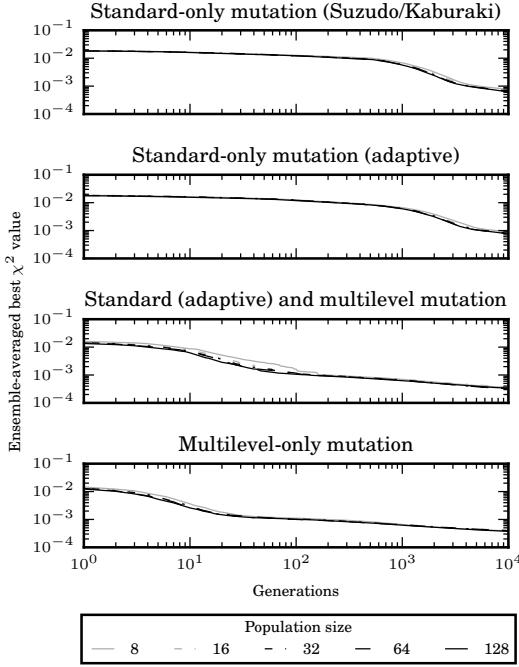


Figure 1: Progression of ensemble-averaged best  $\chi^2$  penalty function value over time, fitting 2000 seeds in a unit cell to a target grain-size lognormal distribution with  $\sigma = 0.35$ ,  $\mu = 0$ , using different mutation strategies. Note that the results for different population strategies are almost indistinguishable to the naked eye.

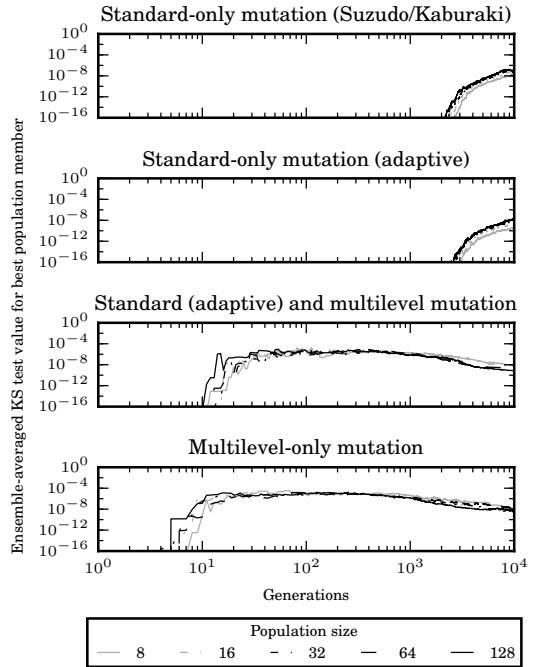


Figure 2: Progression of ensemble-averaged Kolmogorov-Smirnov  $p$ -values over time, corresponding to the population elements with best  $\chi^2$  values in Figure 1. Note that the lower vertical range of the graph is set at  $10^{-16}$ , roughly machine-precision.

the hypothesis that the obtained grain-size distributions follow the target distribution can almost always be rejected outright.

This behaviour is unsurprising. By construction, observed grain-size distributions are normalised to have unit mean; however, the target lognormal distribution possesses a slightly non-unit mean, as outlined above. It is therefore unreasonable to expect good convergence of the latter distribution with any example of the former, regardless of the optimisation technique used. Given this, we are unable to account for the results reported by Gross and Li and Suzudo and Kaburaki, as both also normalise grain sizes to unit mean.

The behaviour of both the penalty function and the KS test changes when the second set of test cases is considered, i.e. seeds in a unit box fitted against lognormal distributions with  $\sigma = 0.35$  and  $\mu = -0.06125$ . Similar graphs to those given for the first test case can be found in Figures 3 and 4. Over the course of  $10^5$  generations, the best penalty function values decrease to between  $10^{-5}$  and  $10^{-7}$ . The standard-only mutation strategies appear steady until well past  $10^3$  generations, particularly in the cases with larger numbers of seeds; by contrast, the strategies that make use of multilevel mutation show their most rapid rates of decrease between 10 and 100 generations, are less marked thereafter, and generally show less difference as the number of seeds is varied. The impact of different choices for population size is noticeable, with larger populations showing slightly quicker rates of decrease in the penalty function; however, the difference is never large enough to justify the added computational expense of population sizes larger than 32, and in some cases 16. (Although we did not perform exhaustive timing tests due to limits on computational resources, we note anecdotally that doubling the population size had a roughly equivalent effect on the computation time per generation; as calculation is dominated by Voronoi tessellation, which is independent for each population member, this is not surprising.) A sample visualisation displaying slices taken from both an initial (Voronoi-Poisson) tessellation and a fitted tessellation can be seen in Figure 5.

The  $p$ -values obtained through KS tests for the second set of test cases also show a marked change. Both of the standard-only mutation strategies display eventual progression towards unity in a more-or-less

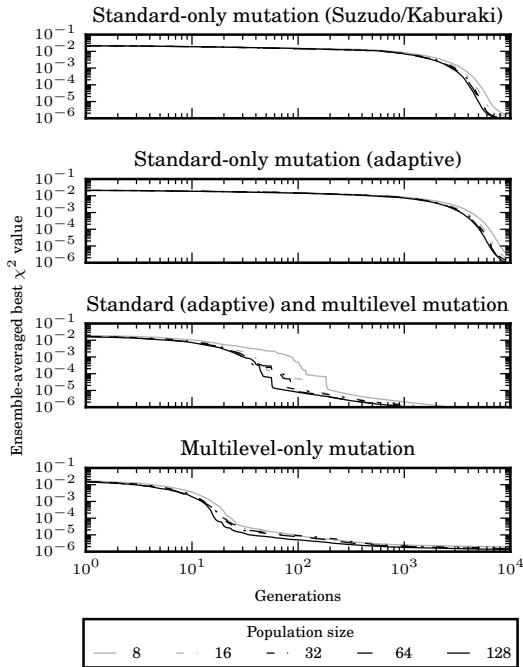


Figure 3: Progression of ensemble-averaged best  $\chi^2$  penalty function value over time, fitting 2000 seeds in a unit cell to a target grain-size lognormal distribution with  $\sigma = 0.35$ ,  $\mu = -0.06125$ , using different mutation strategies.

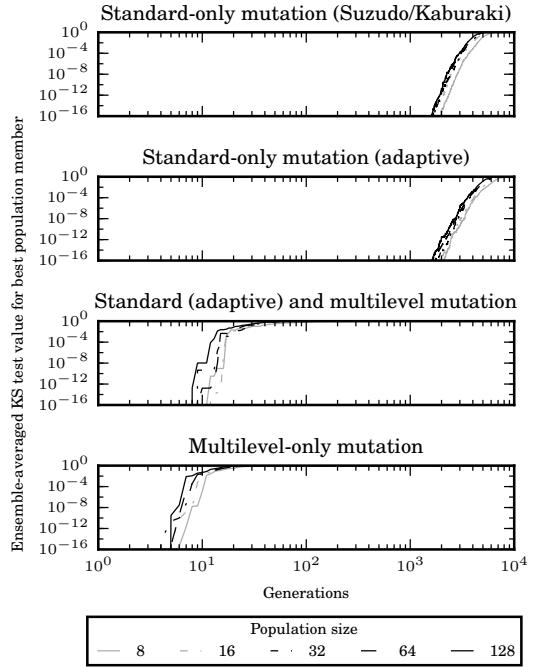


Figure 4: Progression of ensemble-averaged Kolmogorov-Smirnov  $p$ -values over time, corresponding to the population elements with best  $\chi^2$  values in Figure 3. Again, the lower vertical range of the graph is set at roughly machine-precision.

monotonic fashion, and eventually pass the criticality thresholds given in Table 1. The  $p$ -values of both the mutation strategies involving multilevel mutation also progress towards unity and pass criticality; however, they do so much more rapidly than the standard-only strategies, in all cases passing into the range  $[10^{-1}, 1]$  between 10 and 100 generations, as opposed to between 1000 and 10,000 generations. For standard-only mutation types, the KS test values for larger numbers of seeds remain below machine precision much longer than those for smaller numbers of seeds (for 500 seeds, machine precision is passed between 10 and 100 generations, while for 3000 seeds, around 2000 generations are required); however, for multilevel mutation strategies, machine precision is passed in all cases by the 100th generation.

The results for the third set of test cases, i.e. those in the triclinic system, display similar behaviour as those in the second (cf. Figure 6 and Figure 7 for results with  $N = 2000$ ). Best penalty function values for standard-only strategies stay relatively stable at first, before decreasing rapidly in later generations, with larger numbers of seeds showing delay before reduction. Interestingly, the best penalty functions obtained for standard-only mutation strategies display more consistent decay across runtime; however, the final penalty function values stabilise between  $10^{-5}$  and  $10^{-7}$ , as for the previous case. KS test behaviour is also similar, although standard-only strategies do begin to increase above machine precision earlier than for the unit-cell case, corresponding to the earlier point of penalty function decrease. Again, larger population sizes do result in slightly more rapid decrease in penalty function and increase in KS test value, although again, the difference is not marked enough to recommend the consistent use of larger population sizes.

We note that Kolmogorov-Smirnov test values past criticality do not, strictly speaking, provide affirmative evidence that the obtained grain-size distributions follow the target distributions; rather, they only serve to disallow the rejection of that hypothesis under the test conditions. Nevertheless, these  $p$ -values allow us to make an informed definition about a sensible choice of the convergence threshold in the general case. In all our measurements, obtained penalty function values below  $10^{-5}$  correspond to  $p$ -values greater than or equal to 0.95, well above criticality for any reasonable choice of  $\alpha$  or number of seeds and corresponding to very tight agreement between the target and observed CDFs.

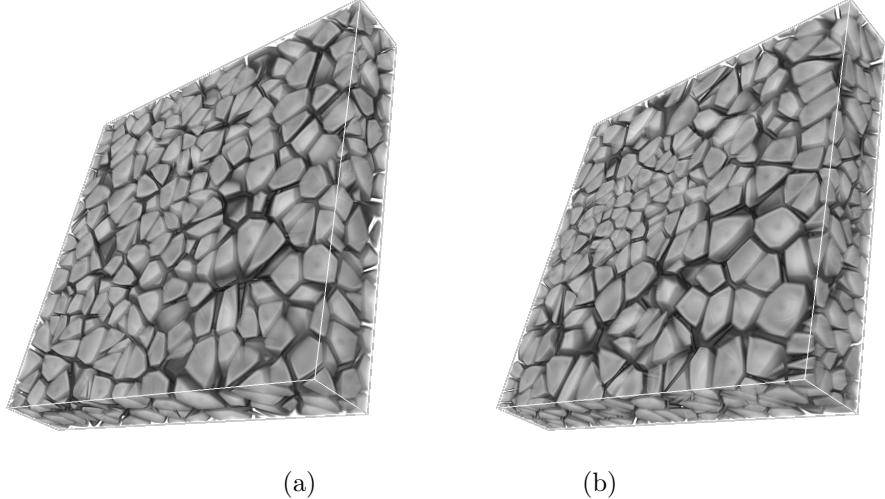


Figure 5: Comparison of slices from polycrystalline structures with 2500 grains each; (a) an initial (Voronoi Poisson) distribution of grain sizes, and (b) lognormal distribution ( $\sigma = 0.35, \mu = -0.06125$ ), obtained only through the use of adaptive multilevel mutation.

By setting  $\chi_{\text{conv}}^2 = 10^{-5}$  and replaying the tests, we can evaluate the speed of convergence of the various mutation strategies. Box-whisker plots displaying the range across the ensemble of generations required to attain convergence for the second and third test cases are given in Figures 8 and 9 respectively. For the second test case, we see that the standard-only mutation strategies behave similarly as the number of seeds increases, following a roughly linear progression. The spread of generations required is consistently between one and two thousand, although slightly higher for lower numbers of seeds; as the calculation of the Voronoi tessellation for each of the  $N$  chromosomes is at worse  $\Theta(N^2)$  [25], this implies that the cost of fitting a lognormal grain-size distribution with a standard-only mutation strategy is at worst case roughly  $\Theta(N^3)$  in the number of seeds, at least for the sizes we have examined here.

(Behaviour for standard-only strategies in the third test case is similar, although the slope of the linear increase is lower than that for the second case. We hypothesise that this is due to the change in the  $\sigma$  parameter, rather than that of the geometry of the unit cell, presumably giving a target distribution that is in some way ‘‘easier to reach’’ from the Poisson-Voronoi starting point; as Voronoi tessellation is being performed according to the minimum image convention, the geometry of the Bravais unit cell should be relatively unimportant to scaling.)

By contrast, the generations required for convergence in the multilevel strategies are significantly lower than those for the standard-only strategies, both for the second and third test cases. Full tables of results for the second and third test cases can be found in Table 2 and Table 3 respectively. The speedup in generations-until-convergence over standard-only strategies is of two orders of magnitude for most of our test cases (i.e. 1000 seeds and higher). However, it is more interesting to consider the scaling. Although extremely wide for 500 seeds, the range of generations requires tightens markedly for larger numbers of seeds, with the spread for 3000 seeds around a fifth of that for 500 seeds. The median and minimum generations required stays quite static for 1000 seeds and greater. These results suggest that the number of generations required to fit a target grain-size distribution with a multilevel strategy is approximately constant, and that the cost in the worst case is therefore roughly  $\Theta(N^2)$  in the number of seeds. Although current applications of polycrystal generation do not require more seeds than tested, it is not unreasonable to expect this to change in the future; as such, the reduction in order represents a significant improvement.

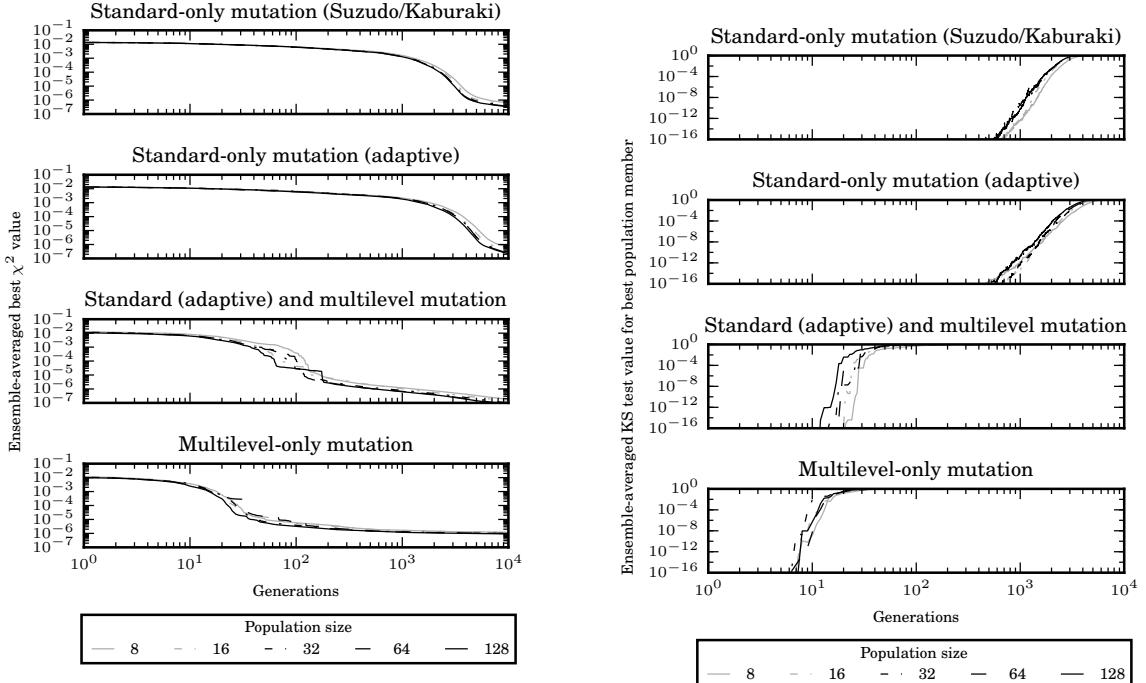


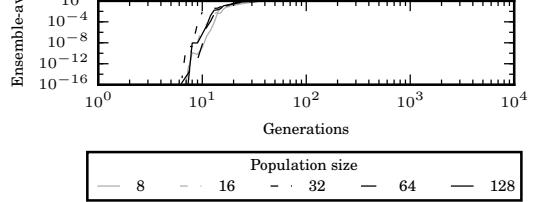
Figure 6: Progression of ensemble-averaged best  $\chi^2$  penalty function value over time, fitting 2000 seeds in a triclinic unit cell to a target grain-size lognormal distribution with  $\sigma = 0.4$ ,  $\mu = -0.08$ , using different mutation strategies.

#### 4. Conclusions and Future Work

We have presented an alternative and improved approach to the evolutional generation of initial polycrystalline structures for use in numerical simulation.<sup>5</sup> Through the application of the multilevel transform described in Section 1.2 as a mutation type, we observe a significant change in the rate with which the penalty function decreases, and correspondingly in the rate at which a target lognormal grain-size distribution can be approximated. This change is around two orders of magnitude in terms of generations required for convergence below a sensibly-chosen penalty-function threshold; the extra computational effort of the multilevel transform is not great, and does not account for a significant increase in per-generation raw computation time. Notably, use of the multilevel transform appears to change the scaling of the algorithm as larger numbers of grains are modelled; rather than increasing linearly, the number of generations required for convergence stays relatively constant. This implies that polycrystalline structures containing potentially indefinitely many grains can be modelled with the only change in computation cost coming from the Voronoi tessellation required to calculate their sizes; approaches without multilevel mutation also incurred a linearly-increasing cost in the number of grains. The overall effect is a reduction of the worst-case fitting time to a quadratic function of the number of seeds.

As noted in Section 1.1, the use of grain-size distribution alone is not sufficient to produce realistic models, nor is the standard Voronoi tessellation capable of the representation of arbitrary grain structures. However, the multilevel approach is agnostic with respect to the penalty function being optimised over, as well as the precise choice of tessellation. Therefore, we see no reason why it could not be applied to more complicated penalty functions that measure alternative characteristics of polycrystalline structures, represented in turn by more sophisticated tessellations. We suspect that the computational benefits in these cases may be similar or greater.

Figure 7: Progression of ensemble-averaged Kolmogorov-Smirnov  $p$ -values over time, corresponding to the population elements with best  $\chi^2$  values in Figure 6..



Population size  
— 8 - - 16 - · 32 — 64 — 128

<sup>5</sup>An implementation of our approach is included in the SCAITools addon of the software package Virtual NanoLab [26].

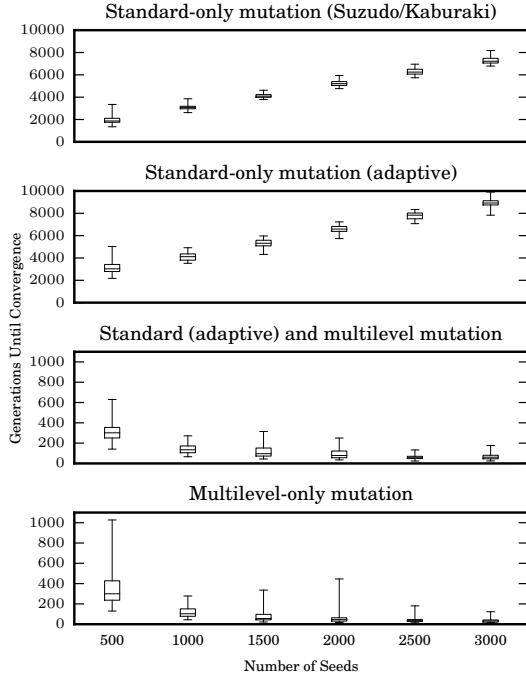


Figure 8: Generations until “convergence” below  $\chi^2 = 10^{-5}$  is achieved in the second test case, fitting varying numbers of seeds within a unit cell to a target grain-size lognormal distribution with  $\sigma = 0.35$ ,  $\mu = -0.06125$ , and using different mutation strategies (all with population size of 32).

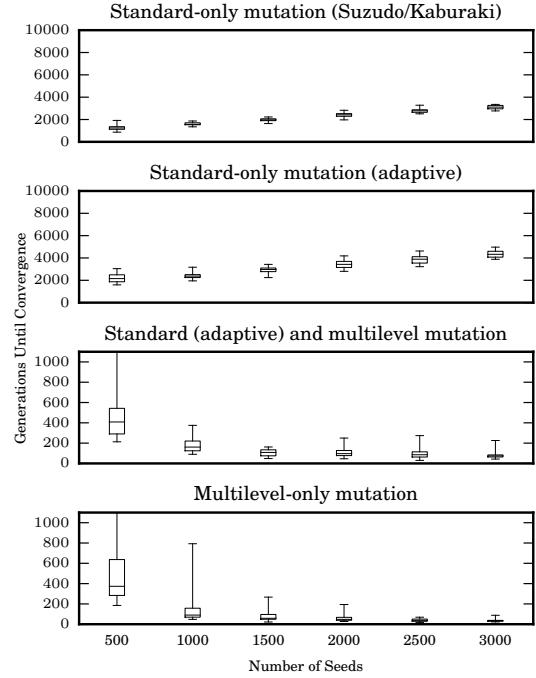


Figure 9: Generations until “convergence” below  $\chi^2 = 10^{-5}$  is achieved in the third test case, fitting varying numbers of seeds within a unit cell to a target grain-size lognormal distribution with  $\sigma = 0.4$ ,  $\mu = -0.08$ , and using different mutation strategies (all with population size of 32).

## 5. Acknowledgements

This work has been funded in part by the Bundesministerium für Bildung und Forschung (BMBF), under the Eurostars project E!6935 ATOMMODEL, and by the Hausdorff Center for Mathematics, Universität Bonn.

## 6. References

- [1] D. Rowenhorst, A. Gupta, C. Feng, G. Spanos, 3D crystallographic and morphological analysis of coarse martensite: Combining EBSD and serial sectioning, *Scripta Materialia* 55 (1) (2006) 11 – 16, viewpoint set no. 41 3D Characterization and Analysis of Materials Organized by G. Spanos. doi:<http://dx.doi.org/10.1016/j.scriptamat.2005.12.061>. URL <http://www.sciencedirect.com/science/article/pii/S135964620600039X>
- [2] A. Brahme, M. Alvi, D. Saylor, J. Fridy, A. Rollett, 3D reconstruction of microstructure in a commercial purity aluminum, *Scripta Materialia* 55 (1) (2006) 75 – 80, viewpoint set no. 41 3D Characterization and Analysis of Materials Organized by G. Spanos. doi:<http://dx.doi.org/10.1016/j.scriptamat.2006.02.017>. URL <http://www.sciencedirect.com/science/article/pii/S1359646206001229>
- [3] M. Groeber, B. Haley, M. Uchic, D. Dimiduk, S. Ghosh, 3D reconstruction and characterization of polycrystalline microstructures using a FIB-SEM system, *Materials Characterization* 57 (45) (2006) 259 – 273. doi:<http://dx.doi.org/10.1016/j.matchar.2006.01.019>. URL <http://www.sciencedirect.com/science/article/pii/S1044580306000623>
- [4] D. Rowenhorst, A. Lewis, G. Spanos, Three-dimensional analysis of grain topology and interface curvature in a  $\beta$ -titanium alloy, *Acta Materialia* 58 (16) (2010) 5511 – 5519. doi:<http://dx.doi.org/10.1016/j.actamat.2010.07.021>

No. Seeds	Pop. Size	Suzudo/Kaburaki			Stand. Adapt.			Dual Adapt.			ML Adapt.		
		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
500	8	1956	2870	4695	3936	5222	7708	315	672	2904	148	—	—
	16	1682	2175	3295	2622	3806	5268	134	404	1019	80	—	—
	32	1351	1942	3349	2176	3210	5027	141	315	630	130	368	1027
	64	1444	1760	2017	2109	2905	4282	152	278	553	127	339	1350
	128	1319	1688	2135	2214	2844	4794	132	217	406	116	261	522
1000	8	3347	3908	4702	4730	5431	6168	101	235	462	61	200	471
	16	2890	3260	3698	3767	4649	5254	54	199	414	37	133	307
	32	2618	3078	3858	3522	4101	4917	66	143	272	44	120	278
	64	2526	2886	3324	3099	3897	4534	59	130	207	17	266	5012
	128	2562	2874	3231	3124	3760	4424	59	112	194	30	130	307
1500	8	4324	5049	5602	6019	6758	7676	73	219	474	32	174	737
	16	3687	4422	5259	5128	5810	6267	66	139	257	28	119	546
	32	3795	4133	4619	4319	5323	5972	44	115	315	22	87	336
	64	3556	4002	4700	4453	5149	5831	40	86	237	20	59	153
	128	3479	3919	4443	4342	5064	6078	43	78	125	22	55	353
2000	8	5660	6307	7336	7045	7906	8895	66	190	339	32	93	553
	16	5028	5624	6418	6095	7010	8075	45	115	251	16	106	876
	32	4767	5220	5946	5736	6587	7231	35	90	250	17	84	446
	64	4572	5120	5593	5509	6338	7145	29	77	195	17	70	439
	128	4458	4909	5547	5317	6182	7381	26	70	228	14	34	155
2500	8	6931	7611	8472	8685	—	—	57	162	303	34	113	425
	16	5940	6676	7486	7404	8110	8711	41	94	199	14	91	634
	32	5741	6286	6954	7074	7763	8344	24	64	133	14	44	181
	64	5444	6107	6835	6921	7469	8538	30	61	201	14	68	636
	128	5535	5936	6574	6628	7257	8165	34	60	212	14	32	206
3000	8	7863	8671	9387	9602	—	—	50	149	392	17	78	277
	16	6917	7782	9036	8578	—	—	44	91	244	17	63	222
	32	6785	7289	8181	7832	8940	9905	25	64	176	16	38	123
	64	6450	7101	7675	7738	8484	9453	29	73	148	9	80	947
	128	6383	7033	7719	7436	8364	9230	24	61	323	13	49	563

Table 2: Minimum, mean and maximum generations elapsed until best penalty function converges below  $\chi^2 = 10^{-5}$  in test ensembles. Data is for fitting grain sizes in a unit box against a lognormal distribution with  $\sigma = 0.35$ ,  $\mu = -0.06125$ . Missing values indicate test ensembles where at least one case failed to converge within  $10^4$  generations.

org/10.1016/j.actamat.2010.06.030.

URL <http://www.sciencedirect.com/science/article/pii/S1359645410003952>

- [5] M. P. Echlin, A. Mottura, M. Wang, P. J. Mignone, D. P. Riley, G. V. Franks, T. M. Pollock, Three-dimensional characterization of the permeability of W-Cu composites using a new “TriBeam” technique, *Acta Materialia* 64 (2014) 307 – 315. doi:<http://dx.doi.org/10.1016/j.actamat.2013.10.043>.  
URL <http://www.sciencedirect.com/science/article/pii/S1359645413008069>
- [6] T. Xu, M. Li, Topological and statistical properties of a constrained Voronoi tessellation, *Philosophical Magazine* 89 (4) (2009) 349–374. arXiv:<http://dx.doi.org/10.1080/14786430802647065>, doi:[10.1080/14786430802647065](http://dx.doi.org/10.1080/14786430802647065).  
URL <http://dx.doi.org/10.1080/14786430802647065>
- [7] S. Zu, X. Liang, X. Xu, Out-of-plane thermal conductivity of polycrystalline silicon nanofilm by molecular dynamics simulation, *Journal of Applied Physics* 110 (054318). doi:<http://dx.doi.org/10.1063/1.3633232>.  
URL <http://scitation.aip.org/content/aip/journal/jap/110/5>
- [8] V. Lucarini, Three-dimensional random Voronoi tessellations: From cubic crystal lattices to Poisson point processes, *Journal of Statistical Physics* 134 (1) (2009) 185–206. doi:[10.1007/](http://dx.doi.org/10.1007/)

No. Seeds	Pop. Size	Suzudo/Kaburaki			Stand. Adapt.			Dual Adapt.			ML Adapt.		
		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
500	8	1194	1873	3369	1418	3572	4784	276	1106	4160	272	—	—
	16	1130	1418	1875	1484	2702	4470	218	656	1620	127	1295	7443
	32	866	1247	1919	1591	2225	3045	213	493	1201	185	549	2558
	64	874	1080	1414	597	1886	2946	205	370	1074	143	463	1875
	128	746	1034	1319	1306	1850	2776	133	340	675	105	—	—
1000	8	1540	2079	2497	2540	3359	4629	118	267	603	70	224	417
	16	1475	1720	2090	1455	2659	3292	102	203	397	68	306	3623
	32	1346	1597	1868	1950	2421	3172	90	182	375	48	154	793
	64	1211	1475	1745	1634	2277	2801	70	162	420	45	128	331
	128	1236	1475	1752	1505	2189	2799	55	135	234	42	94	249
1500	8	2017	2429	2833	2856	3554	4367	99	190	415	30	119	356
	16	1759	2134	2557	2254	3191	4019	68	161	348	37	90	276
	32	1638	1974	2228	2243	2933	3426	48	107	162	22	79	267
	64	1551	1966	2188	2098	2770	3447	62	98	158	19	65	166
	128	1654	1922	2230	2173	2777	3494	60	121	326	20	63	184
2000	8	2301	2764	3234	3626	4207	5183	78	205	472	27	104	389
	16	2326	2610	2935	2942	3668	4603	48	120	254	17	82	407
	32	1968	2407	2828	2805	3420	4187	46	113	250	27	64	194
	64	2099	2287	2581	2523	3280	3960	39	109	265	25	79	718
	128	1995	2264	2468	2683	3270	3885	38	78	131	21	40	117
2500	8	2855	3263	3679	3894	4633	5581	72	150	317	26	74	300
	16	2642	2902	3462	3334	4043	4695	44	119	262	30	56	123
	32	2510	2780	3280	3221	3857	4625	30	96	274	16	40	69
	64	2275	2709	3154	3114	3782	4461	45	79	180	19	48	182
	128	2344	2646	3054	3239	3747	4376	37	66	111	21	37	190
3000	8	3119	3558	4068	4417	5148	5951	42	138	255	30	63	257
	16	2876	3296	3956	3972	4583	5451	48	123	524	18	49	306
	32	2770	3102	3353	3882	4357	4971	43	87	226	23	36	88
	64	2742	3094	3439	3518	4270	5264	44	85	175	20	37	154
	128	2570	3049	3380	3231	4089	4925	32	69	197	20	31	60

Table 3: Minimum, mean and maximum generations elapsed until best penalty function converges below  $\chi^2 = 10^{-5}$  in test ensembles. Data is for fitting grain sizes in a triclinic cell against a lognormal distribution with  $\sigma = 0.4$ ,  $\mu = -0.08$ . Missing values indicate test ensembles where at least one case failed to converge within  $10^4$  generations.

s10955-008-9668-y.

URL <http://dx.doi.org/10.1007/s10955-008-9668-y>

- [9] C. G. Granqvist, R. A. Buhrman, Ultrafine metal particles, Journal of Applied Physics 47 (5) (1976) 2200–2219. doi:<http://dx.doi.org/10.1063/1.322870>.  
URL <http://scitation.aip.org/content/aip/journal/jap/47/5/10.1063/1.322870>
- [10] D. Gross, M. Li, Constructing microstructures of poly- and nanocrystalline materials for numerical modeling and simulation, Applied Physics Letters 80 (5) (2002) 746–748. doi:<http://dx.doi.org/10.1063/1.1432448>.  
URL <http://scitation.aip.org/content/aip/journal/apl/80/5/10.1063/1.1432448>
- [11] T. Suzudo, H. Kaburaki, An evolutional approach to the numerical construction of polycrystalline structures using the Voronoi tessellation, Physics Letters A 373 (48) (2009) 4484 – 4488. doi:<http://dx.doi.org/10.1016/j.physleta.2009.09.072>.  
URL <http://www.sciencedirect.com/science/article/pii/S0375960109012602>
- [12] M. Groeber, S. Ghosh, M. D. Uchic, D. M. Dimiduk, A framework for automated analysis and simulation of 3D polycrystalline microstructures.: Part 1: Statistical characterization, Acta Materialia 56 (6) (2008) 1257 – 1273. doi:<http://dx.doi.org/10.1016/j.actamat.2007.11.041>.  
URL <http://www.sciencedirect.com/science/article/pii/S1359645407007914>

- [13] K. Teferra, L. Graham-Brady, Tessellation growth models for polycrystalline microstructures, Computational Materials Science 102 (2015) 57 – 67. doi:<http://dx.doi.org/10.1016/j.commatsci.2015.02.006>.  
 URL <http://www.sciencedirect.com/science/article/pii/S0927025615000671>
- [14] A. Leonardi, P. Scardi, M. Leoni, Realistic nano-polycrystalline microstructures: beyond the classical Voronoi tessellation, Philosophical Magazine 92 (8) (2012) 986–1005. arXiv:<http://dx.doi.org/10.1080/14786435.2011.637984>.  
 doi:10.1080/14786435.2011.637984.  
 URL <http://dx.doi.org/10.1080/14786435.2011.637984>
- [15] A. Leonardi, M. Leoni, P. Scardi, Atomistic modelling of polycrystalline microstructures: An evolutionary approach to overcome topological restrictions, Computational Materials Science 67 (0) (2013) 238 – 242. doi:<http://dx.doi.org/10.1016/j.commatsci.2012.09.013>.  
 URL <http://www.sciencedirect.com/science/article/pii/S0927025612005526>
- [16] M. Groeber, S. Ghosh, M. D. Uchic, D. M. Dimiduk, A framework for automated analysis and simulation of 3D polycrystalline microstructures. part 2: Synthetic structure generation, Acta Materialia 56 (6) (2008) 1274 – 1287. doi:<http://dx.doi.org/10.1016/j.actamat.2007.11.040>.  
 URL <http://www.sciencedirect.com/science/article/pii/S1359645407007872>
- [17] T. Suzudo, personal communication (2015).
- [18] H.-J. Bungartz, M. Griebel, Sparse grids, Acta numerica 13 (2004) 147–269.
- [19] K. Matuschke, Trigonometrische Interpolation auf verallgemeinerten dünnen Gittern mit beliebiger Levelstruktur, Diplomarbeit, Institut für Numerische Simulation, Universität Bonn (2014).
- [20] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM Journal on Numerical Analysis 45 (3) (2007) 1005–1034.
- [21] M. Griebel, J. Hamaekers, Fast discrete Fourier transform on generalized sparse grids, in: Sparse grids and Applications, Vol. 97 of Lecture Notes in Computational Science and Engineering, Springer, 2014, pp. 75–108, iNS Preprint No. 1305.
- [22] C. Rycroft, Voro++: A three-dimensional Voronoi cell library in C++, Lawrence Berkeley National Laboratory.
- [23] Turquoise mineral data – turquoise crystallography, <http://webmineral.com/data/Turquoise.shtml>, retrieved 20/5/15.
- [24] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, [Online; accessed 2015-03-09] (2001–).  
 URL <http://www.scipy.org/>
- [25] M. J. Golin, H.-S. Na, On the average complexity of 3D–Voronoi diagrams of random points on convex polytopes, Computational Geometry 25 (3) (2003) 197 – 231. doi:[http://dx.doi.org/10.1016/S0925-7721\(02\)00123-2](http://dx.doi.org/10.1016/S0925-7721(02)00123-2).  
 URL <http://www.sciencedirect.com/science/article/pii/S0925772102001232>
- [26] Virtual NanoLab, QuantumWise A/S, <http://www.quantumwise.com>.