PAPER

Atom based grain extraction and measurement of geometric properties

View the article online for updates and enhancements.

Atom based grain extraction and measurement of geometric properties

Gabriel Martine La Boissonière[1] ![ORCID] and Rustum Choksi

Department of Mathematics and Statistics, McGill University, 805 Rue Sherbrooke O, Montréal, QC H3A 0B9, Canada

E-mail: gabriel.martine-laboissoniere@mail.mcgill.ca and rchoksi@math.mcgill.ca

Abstract

We introduce an accurate, self-contained and automatic atom based numerical algorithm to characterize grain distributions in two dimensional Phase Field Crystal (PFC) simulations. We compare the method with hand segmented and known test grain distributions to show that the algorithm is able to extract grains and measure their area, perimeter and other geometric properties with high accuracy. Four input parameters must be set by the user and their influence on the results is described. The method is currently tuned to extract data from PFC simulations in the hexagonal lattice regime but the framework may be extended to more general problems.

Keywords: grain recognition, phase field crystal, Voronoi diagram, grain growth

(Some figures may appear in colour only in the online journal)

1. Introduction

The polycrystalline structure of materials is one of the main deciding factors in determining physical properties such as hardness and ductility. A thorough understanding of the conditions that lead to the formation of a desirable crystalline structure is essential in many technological applications. To study such conditions and their impact on the evolution of polycrystalline materials, it is necessary to obtain experimental data throughout the structure forming processes. For example, extensive studies into the evolution of grains and stagnation in thin Al and Cu metallic films have been carried out in [1]. Such investigations are experimentally challenging so simulations are desirable to investigate the evolution of crystals. Many theoretical models have been proposed to do so, such as the well-known

---

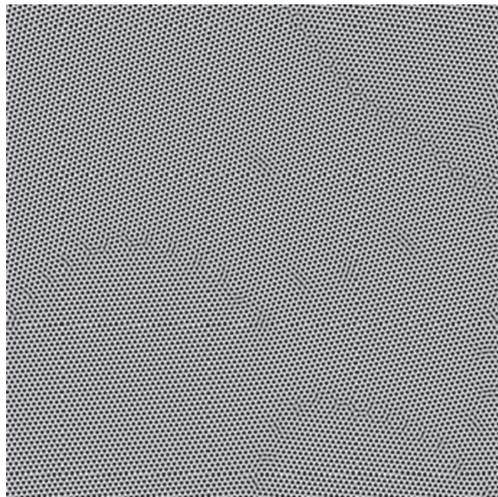[1] Author to whom any correspondence should be addressed.

Figure 1. Detail of atomic positions in a simple PFC simulation. Each atom is represented as a 'bump' in luminosity. Away from grain boundaries, these bumps are uniform through space and time and arrange into regular hexagonal lattices.

Mullins model [2]. More recently, the Phase Field Crystal (PFC) [3, 4] model has been successful in modeling a wide variety of crystalline phenomena. In particular, grain size distributions predicted by the 2D PFC model agree remarkably well with experimental data [5]. While these results are encouraging, to our knowledge, a thorough statistical assessment of geometric properties of grain boundaries in the PFC model has yet to be performed. To do so, one needs a verifiably accurate method for extracting grain boundaries and geometric features from input data containing millions of atoms (see figure 1).

In this article, we present and assess a simple and self-contained method specifically adapted to characterize the geometric properties of grain distributions extracted from 2D hexagonal PFC simulations. Attempts to numerically characterize such grain networks are not new. Much experimental data comes from digital images of thin metallic films or tomographic maps of 3D crystals in which grains are very large compared to atoms. This typically leads to very sharp and clear grain boundaries that can be recognized readily both manually or numerically, as in [6, 7]. Grain area can then be measured by counting the number of pixels within each connected region. Identifying the boundary network obtained from atomistic simulations proves to be a much more complicated task. Unlike images where orientation is essentially a piecewise constant map in space, the orientation in atomistic simulations must first be extracted from atomic positions. Such an orientation map may be extracted using variational techniques as in [8–11]. Alternatively, the map may be obtained by projecting and interpolating geometrically computed local orientations, obtained for example with the atomistic visualization tool OVITO [12] and see [13] for details on specific approaches. In both cases, the gradient of the orientation map may then be thresholded to create a 'skeleton' of grain boundaries. Grain area may then be characterized by counting the number of pixels in each region enclosed by a boundary. Recently, an automated atomistic technique to identify grains whose barycenter can be tracked in time has been developed in [14].

We propose a grain extraction and measurement procedure which is also based on using atomic positions instead of working with a projection of the orientation map. We show that our method is capable of measuring grain area, perimeter, Grain Boundary Character Distributions (GBCD) [15] and related geometric properties with high accuracy. Our approach

uses the Voronoi region of each atom to characterize defects and grain boundaries, creating a stencil with which one identifies grains on the graph of atoms. Once atoms belonging to a grain have been identified, its geometry may be computed using the geometry of the atomic cells. All steps rely on very simple geometric routines that can be implemented efficiently. For simplicity and because our main goal is to extract grain distributions arising from the basic PFC equation [3], our algorithm is adapted to 2D hexagonal lattices. The proposed framework may be extended to more interesting scenarios once the underlying atomistic characterization is extended to other lattice types or to 3D as in [14].

The outline of the article is as follows. First, we detail our atom based grain extraction and measurement procedure. Then, we validate the accuracy of both the numerically computed grain network and the extracted properties by comparing the numerical segmentation with hand segmentations and artificially constructed grain distributions. The choice and influence of the input parameters is also discussed. Finally, we present a few sample results from PFC simulations.

2. Atom based grain extraction and measurement

Let us suppose we are given an image of atomic positions representing a polycrystalline material as in figure 1. A grid based approach to characterizing this material would be the use of variational techniques to convert the input into a map of local lattice orientations. The gradient of this map may then be used to construct rough grain boundaries from which grains can be recognized as enclosed regions, represented as connected sets of pixels on a grid. The geometric properties of these sets can be computed with image processing techniques. Another approach, unified in the proposed method, is to first extract atomic positions then repeat the steps outline above on the graph of connections between neighbor atoms which we call the 'graph of atoms'. The geometric properties of these sets are then to be computed using the geometric properties of the atomic cells they contain. While the implementation of both approaches is quite different, their respective steps are in a natural parallel.

We now present the later approach, the atom based grain extraction and measurement algorithm. For simplicity and to analyze basic PFC simulations, we focus on characterizing 2D simulations in which atoms have arranged into suitably large hexagonal lattices with interatomic distance d. The full procedure is outlined as follows:

- Atomic positions are extracted from an input image (2.1).
- A lattice orientation is assigned to each atom (2.2).
- The local lattice structure of each atom is probed to determine which atoms are close to grain boundaries, forming a grain boundary mask (2.2).
- Grains are identified in a fashion similar to a flood-fill algorithm (2.3).
- Atoms in the grain boundary mask are assigned to the closest grain (2.3).
- Grain properties are measured (2.4).
- Spurious grains are detected and removed (2.5).

A MATLAB implementation of our procedure is capable of analyzing an 8192^2 image containing roughly a million atoms in about 7 min on a standard machine. We note that an implementation of the grid based approach would follow the same general outline with the first steps replaced by variational filtering techniques, converting the input image into an orientation map and an appropriate grain boundary mask. Subsequent steps follow naturally with the pixels playing the role of atoms.

Our input data typically comes from simulations over periodic domains; it is therefore necessary for the implementation to allow grains to be identified even if they cross periodic
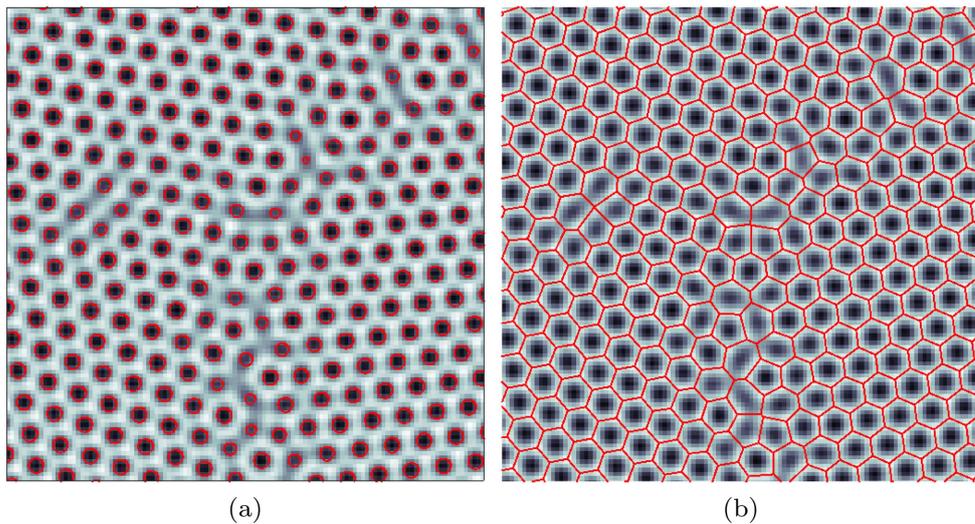
(a)   (b)

Figure 2. Detail of a PFC simulation with contours used to find atomic positions (a) and their Voronoi diagram (b). Note how grain boundaries can be identified as the set of cells that are not perfect hexagons.

boundaries. Average orientations are also often required, a possibly ill defined quantity given the identification between $0° = 60°$ in a hexagonal lattice. To compute an isotropic average, convert a set of orientations into position vectors on the unit circle such that $(1, 0)$ corresponds to both $0°$ and $60°$. Then, the angle of the average *vector* can be mapped on the same range and taken as the average orientation. This agrees with the Euclidean average for narrow distributions but avoids any ambiguity near the identification.

2.1. Atom recognition

Given an input image where atoms are represented as pixelated circular bumps, we use a contouring algorithm to find the intersection points between the input and a height level h. The points of each contour are then fit to a circle whose center is used to define atoms as in figure 2(a). Note that the input parameter h must be chosen so that atoms at grain boundaries are also recognized even if they are not as well formed as those within lattices. This is not an issue since the appropriate range is quite wide.

2.2. Computation of atomic Voronoi properties and grain boundaries

We now identify defects and atoms with a local environment that is far from the expected lattice configuration. This traces a preliminary boundary network while other atoms will be assumed to be part of a grain. Our approach is to compute the Voronoi tessellation [16] of all atomic positions and determine how closely the Voronoi or Wigner–Seitz cell of an atom matches a regular hexagon, see figure 2(b) for an illustration. If the cell is hexagonal, we may also compute the local lattice orientation of an atom by averaging the angle of the vector pointing to the six Voronoi vertices modulo the $60°$ symmetry. Near boundaries, this computation will give inaccurate results. When enough neighbor atoms have similar orientations, they form a coherent lattice and will be considered to form a grain.
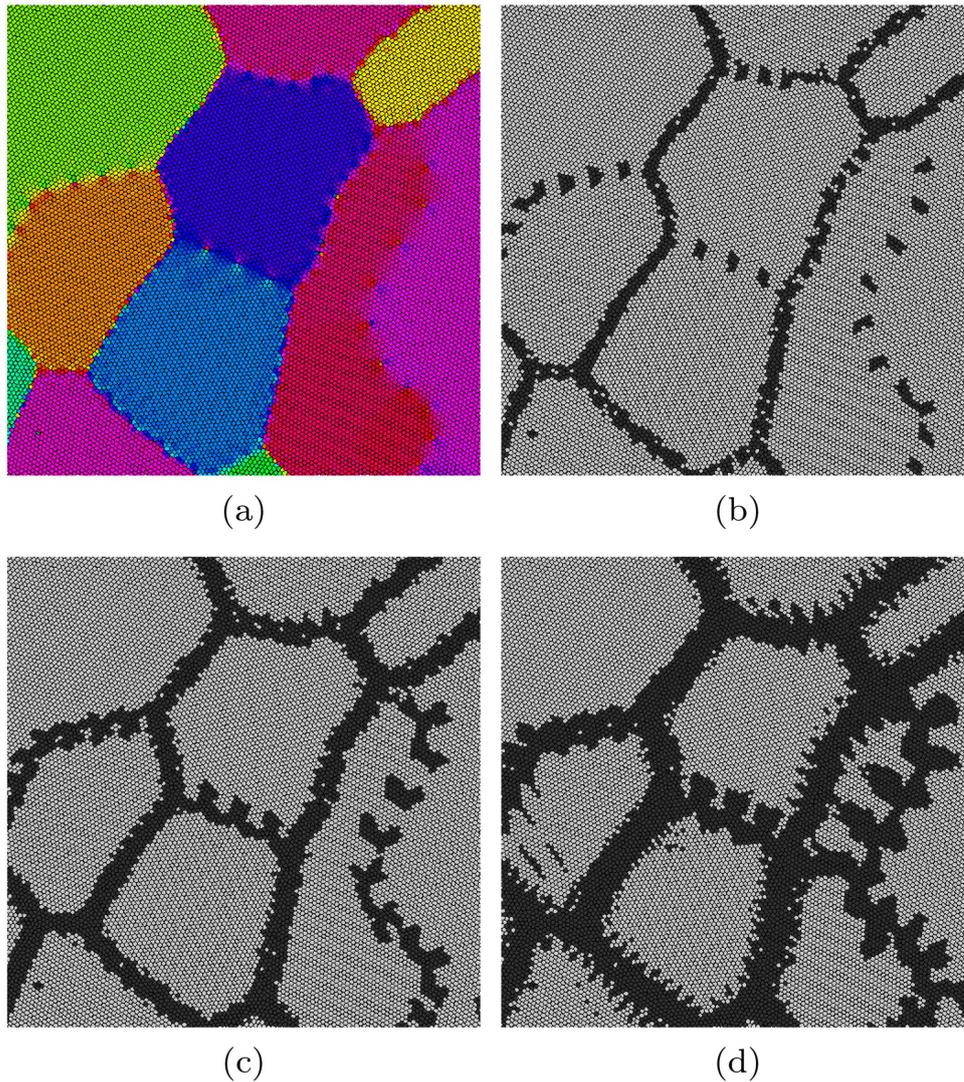
Figure 3. Visualization of local lattice orientations in a PFC simulation (a). Grain boundary masks using γ equal to 0.002, 0.0005 and 0.0002 in (b), (c), (d) respectively.

A simple measure of closeness between the Voronoi cell and the regular hexagon is as follows. The area A and perimeter P of the Voronoi regions can be computed using methods such as the Shoelace formula [17]. From these properties, one can compute the isoperimetric ratio $Q = \frac{4\pi A}{P^2}$ of the region. This parameter is positive and less than 1 by the isoperimetric inequality, that value being achieved only by a perfect circle. A similar result is that within the class of N-sided polygons, the regular N-gon achieves the maximum isoperimetric ratio $Q_N = \frac{\pi/N}{\tan(\pi/N)}$. This means that the cell of an atom with six neighbors is close to a regular hexagon if Q is close to Q_6. To characterize defects and atoms close to grain boundaries, we will label any atom for which $|Q_6 - Q|$ is larger than some threshold γ as 'bad'. In contrast, if an atom has six neighbors and its region is close enough to a perfect hexagon, it is called a

'good' atom. The value of γ then sets how atoms close to boundaries are expected to behave compared to atoms deep in a perfect lattice. This 'bad atom mask' is illustrated in figure 3: notice how decreasing γ from (b) to (d) progressively thickens the boundaries and closes most gaps between grains until the boundaries become too thick and start invading the grains proper. From this figure, the partition between good and bad atoms is clear: it gives a preliminary mask that can be used to ignore all misleading orientations near grain boundaries, improving the accuracy in the detected grain angle and boundaries.

Since this approach is similar to the Voronoi Analysis detailed in [13], it suffers from issues that reduce its applicability to other lattice types or to 3D. As in [14], this step may be overhauled with similar methods such as Common Neighbor Analysis or Centrosymmetry Parameter.

2.3. Flood fill procedure

Grains must now be identified by connecting neighboring good atoms based on their orientation. This step is similar to extracting connected components on a square grid, working instead on the graph of atoms given by connecting neighboring atoms. These may be detected by connecting atoms that share a Voronoi vertex or by connecting atoms that share a triangle in the Delaunay triangulation [16] of the atomic positions.

A grain is found by picking a random good atom and probing its six neighbors. If a neighbor is good and its orientation is sufficiently close to that of the starting atom, it is accepted into the grain. How small this misorientation can be is set by the misorientation threshold θ. The grain having grown by one shell of neighbors, a preliminary grain orientation is computed by averaging the orientations of accepted atoms. This process iterates for all new atoms with the difference that orientations are now compared to the preliminary grain orientation. When the process ends, the number of atoms is computed and if it is smaller than an atom number threshold α, the grain is deleted and its atoms are marked as bad. This ensures that small spurious clusters of good atoms are not falsely recognized as grains.

The procedure iterates until all good atoms have been associated to a grain or marked as bad, resulting in a partition of atoms in either grains or the thick grain boundary network similar to figure 3(c). As long as no true grain has been missed, we can assume the 'sharp' grain boundaries lies exactly in the middle of the region of bad atoms between preliminary grains. This is done approximately by connecting a bad atom to the grain that contains most of its neighbors, ignoring those connected only to bad atoms. This process repeats until all atoms belong to a grain. Since all grains acquire a few new atoms, the smallest grain detected will be slightly bigger than α. An appropriate choice for this threshold is then any value smaller than half the size of the smallest feature that is represented in the input. Since very small clusters of good atoms may be present within the boundary network, α must still be larger than some minimum value to ensure that not too many false positives are detected.

2.4. Measurement of grain properties

Measuring grain properties turns out to be quite simple once grains are identified and their atoms known. Let us call boundary atoms those atoms that have a neighbor inside a different grain. We now describe how to compute several grain properties using only the Voronoi area of atoms.

Grain area. The area of a grain is the sum of the area of the Voronoi regions of its atoms. This property is usually expressed through the normalized reduced area, computed by first taking the root of the areas then dividing by the average.
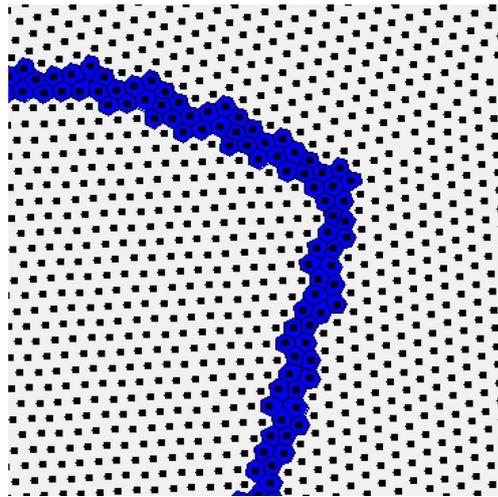
Figure 4. Detail of a PFC grain boundary with extracted atomic positions in black and the Voronoi region of boundary region atoms colored in blue.

Grain perimeter. The perimeter of a grain is the sum of the area of all boundary atoms that are connected to the grain, divided by a characteristic boundary thickness. This sum then includes all boundary atoms of the current grain plus the boundary atoms of grains that are in contact with it.

Several issues arise when attempting to compute grain perimeter since a boundary is an extended region ambiguously defined by discrete atomic positions. In addition, similarly to the coastline paradox, one could either trace a smooth line *somewhere* through a grain boundary or one could trace the line connecting boundary atoms one by one. While the first approach is more sensible from a materials science point of view, only the second is unambiguous. Naturally, the perimeters recorded vary between the two approaches. Since there is no clear notion of perimeter, we use a geometrically unambiguous quantity to approximate the length of grain boundaries. Taking PFC as a concrete example, grain boundaries are very thin with a thickness comparable to the interatomic distance d. The region defined by the Voronoi regions of boundary atoms is then roughly two atoms thick, illustrated in figure 4. Thus, dividing its area by $2d$ transforms the quantity into a one dimensional value akin to perimeter. While d is not necessarily fixed in space or time, we simply assume that all atoms have a perfectly hexagonal Voronoi region and equally divide the domain so that given the domain area A and the number of atoms N_a, one has that $d^2 \approx 2A/(\sqrt{3}N_a)$. This simple estimate gives reasonable results as will be shown in 3.2.

Any other method of relating the area of the diffuse boundary to the perimeter could be used and accuracy might be gained by using a relationship that is better informed by the exact geometry of expected interfaces. An alternative and unambiguous characterization of area and perimeter in atomistic simulations may simply be the number of atoms inside a grain and at the boundary respectively. While this would not as well extend to mesoscopic experiments, it could be used to compare different atomistic simulations directly and unambiguously.

Grain isoperimetric ratio. This ratio is computed from the area and the perimeter. The closer this value is to 1, the closer a grain is to a circle. In particular, oscillations in the boundary quickly decrease this value. This provides a measure not only of grain circularity but also of the roughness of its interface. Our method does not provide a geometric

isoperimetric ratio since area and perimeter do not represent the same curve, hence the ratio may be larger than 1.

Grain coordination number. The coordination number of a grain is the number of grains with which it makes contact.

Grain–grain area ratio. Given the areas A_1 and A_2 of two neighbor grains, the area ratio is defined as the minimum between A_1/A_2 and A_2/A_1 to keep the ratio below 1. This property is an attempt to capture geometric correlations between neighbors. For example, if a grain has grown in time, its area must have been part of neighboring grains, giving rise to local imbalances in area. The evolution in time of the distribution of area ratios can then reveal whether this process 'balances out' across the structure or whether large grains grow systematically to the detriment of very small neighbor grains.

Grain–grain interface length, misorientation and GBCD. As with the perimeter, the area of the Voronoi regions of *only* the boundary atoms connecting two neighbor grains may be summed and divided by $2d$, measuring the length of the interface between the grains. The misorientation between two grains may be computed using the periodic angular difference between the grain orientations. A derived property is then the GBCD [15] which is computed by summing the interface length between grains whose misorientation falls in a given angular bin. In contrast to the bare misorientation distribution, this produces a histogram weighted by interface length rather than number which highlights preferred misorientations according to a geometrical metric.

2.5. Post processing grains

Some issues in the input image may lead to the detection of spurious grains. This is a problem when using very small thresholds in the hope of recognizing all true grains. The principal cause of concern is the presence of defects in ambiguous low misorientation boundaries as shown in figure 5. In (a), it is quite ambiguous whether there are two purple grains or a single one. Upon closer inspection, the misorientation between the dark and light purple is $2.6°$ which is slightly above the chosen threshold $\theta = 2.5°$. However, there is a very smooth transition between the two orientations suggesting that this is a single grain under stress due to the large defect, hence, it should not be split. While the algorithm correctly identifies the whole smooth region as a single grain, the defect influences a large set of good atoms. Such regions around point defects are sometimes counted as spurious grains, especially in large grains, as in (b). These artefacts may be identified by detecting grains that have only one neighbor, which should be extremely rare in usual circumstances.

A similar issue sometimes arises at the boundary between two grains. Unlike defects, the problem is that part of the whole diffuse boundary is detected as an independent grain. Such cases cannot be detected on the basis of the coordination number but they usually have a very small misorientation to one of their neighbor. An example of such an artefact is shown in figure 5(c). While the diffuse region contains a dark blue and a purple grain, the misorientation between the blue half and the dark blue grain to its right is of only $1°$. The boundary marked by the white arrow in (d) is then spurious and must be deleted. Such cases are detected easily by finding all grain to grain misorientations smaller than θ which then becomes an exact threshold. In both cases presented above, the spurious grain is simply deleted and its atoms are reassigned to the relevant neighbor.
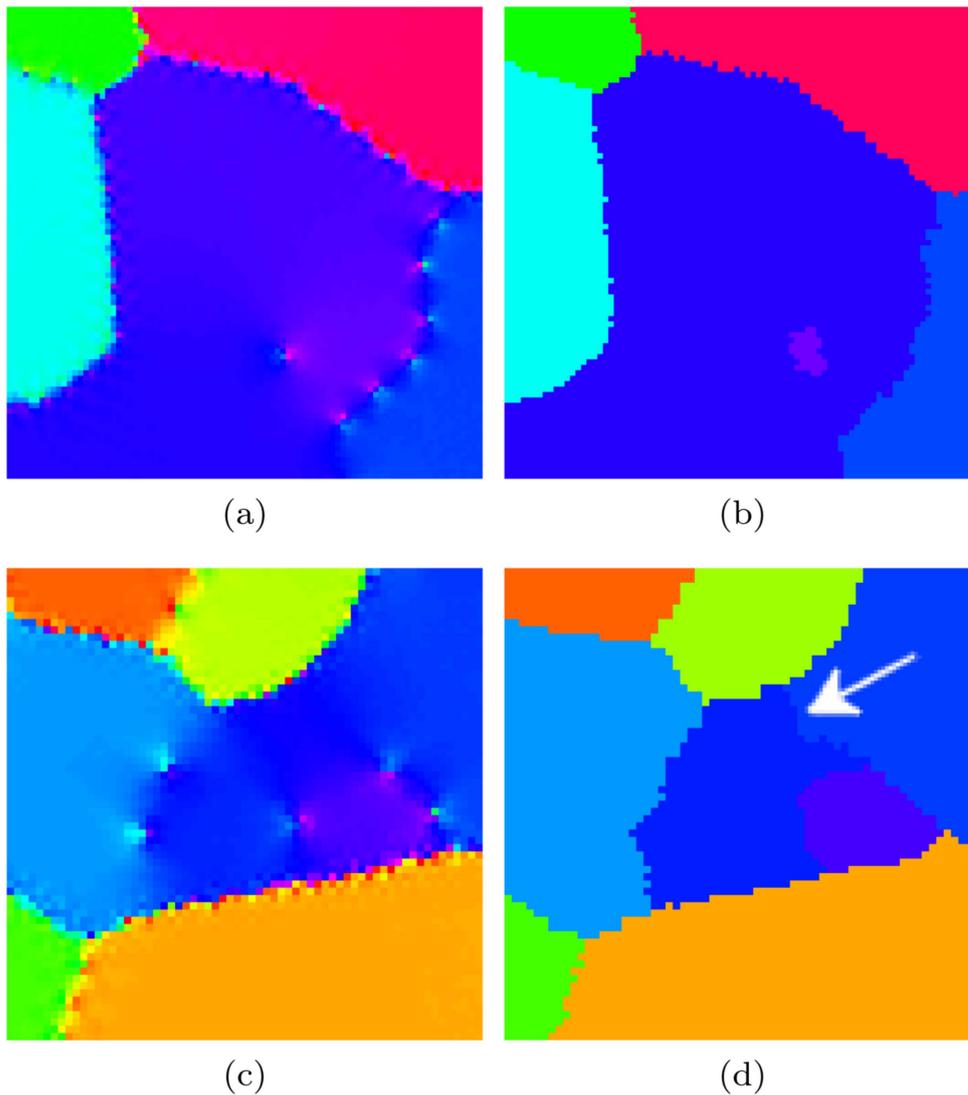
Figure 5. Visualization of local lattice orientations in a PFC simulation (a), (c). Corresponding visualizations of extracted grain orientations showing a single-neighbor grain (b) and a boundary with misorientation smaller than the threshold (d).

3. Validation

In this section, we validate the results of our atom based grain extraction procedure. The first issue is to verify that we properly account for almost all grains present in the input and identify as few spurious grains as possible. The definition of a grain is obviously subject to interpretation hence our goal is to ensure that the numerical scheme essentially reproduces a human manual segmentation. The second issue is that once all grains are properly identified, geometric properties must be faithful to the input with the caveat discussed before in the case of interface lengths. Our source of validation is two-fold: first, we use results from PFC

simulations, which we briefly summarize, to have a 'natural' grain distribution that can be hand segmented. Second, we produce artificial grain distributions using known geometric shapes whose number and properties are unambiguous. In both validation sources, there are targets for the grain structure and geometric properties that the numerical scheme must recover. We shall use cumulative density functions (CDF) to compare between data sets as there is no guarantee of a one-to-one map between recognized and actual grains. Unlike a histogram, a CDF is unique and can be used even for very small data sets.

3.1. PFC evolution

While many more elaborate PFC-like models have been developed, ranging from the prototypical evolution of [3] to those modeling dendritic solidification [18] and even graphene [19], its simplest form gives the evolution of a phase field u according to the partial differential equation

$$u_t = \nabla^2((\nabla^2 + q_0^2)^2 u + u^3 - \beta u).$$

Such an evolution conserves the average phase $\langle u \rangle = m$. The parameter β can be thought of as an inverse temperature while q_0 sets an atomic lengthscale. In 2D, the phase diagram of the system in (m, β) can be divided into three regions: a 'liquid' state where $u = m$, a 'rolls' state $u = m + A\sin(qx)$ and finally a state where u is a superposition of three rolls misoriented by $60°$ forming a hexagonal lattice. In this regime, an initially noisy phase field quickly evolves to form local 'bumps' representing atoms. These arrange locally into small hexagonal clusters with interatomic distance $d(q_0) = 4\pi/(\sqrt{3}\,q_0)$. The subsequent evolution is at a macroscopic level as clusters become grains and coarsen. This PFC system gives a prototypical evolution with a single crystalline lattice type and is thus the ideal test for our method.

To simulate the evolution, we use the unconditionally stable semi-implicit numerical scheme developed in [20]. We fix a square domain with periodic boundary conditions and choose the same PFC parameters $(m, \beta) = (0.07, 0.025)$ along with the regularization parameter $C = 2\beta = 0.05$ and $\tau = 1000$. Fixing $q_0 = 1$, the numerical domain size is given by $L = 1024d(1) \approx 7429$ with 8192 grid points. Defined this way, the domain supports roughly 1024 atoms in a dimension aligned with the lattice and each atom is resolved by approximately 8^2 pixels. These simulations were processed with $h = -0.035$, $\gamma = 0.001$, $\theta = 2.5°$ and $\alpha = 40$, choices justified in 3.3.

3.2. Comparison with PFC grain distributions

Our first comparison comes from the evolution of a noisy phase field according to the PFC equation. The phase was saved after 228 and 40 000 time steps, corresponding to roughly 1.25 million atoms divided in 1830 and 130 grains respectively. For clarity, we shall call these the early and late distributions. To manually segment these, atomic orientations were projected and interpolated on a grid over which grain boundaries were carefully traced as polygons without overfitting atomic structural details. Because low misorientation boundaries are ambiguous, especially in large grains, it is not realistic to trace *all* boundaries that are visible or implied by a line of defects; one should ideally consider the misorientation between the two grains as well as the openness of the line of defects. Without accepting such poorly defined porous boundaries, the manual boundary network in the late distribution is shown in figure 6(a) and contains 129 grains with an average perimeter of 2426. Note that average area is redundant since it equals the area of the domain divided by the number of grains. Since the early distribution contains so many grains, only 459 grains were extracted which represents a
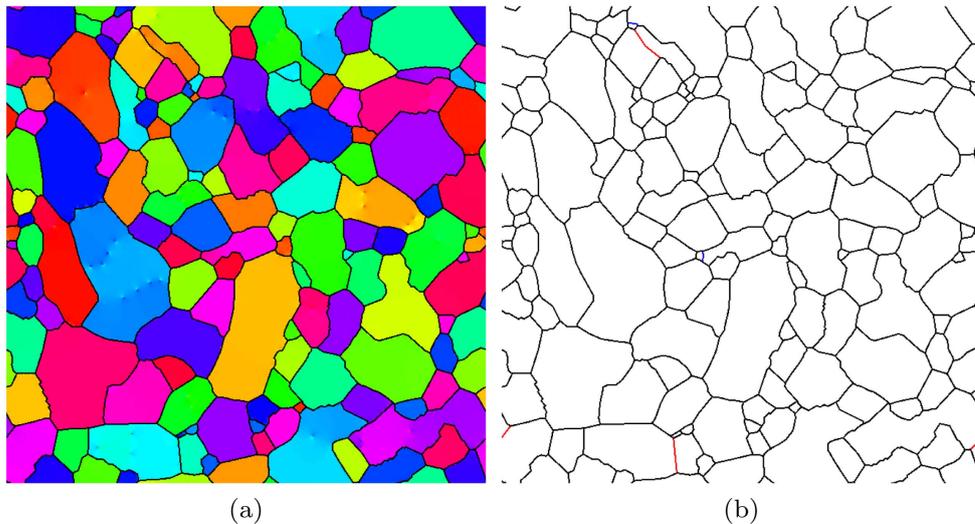
Figure 6. Visualization of local lattice orientations in the late distribution with manually segmented grain boundary network in black (a). Comparison between the manual and numerical segmentations (b); common boundaries are drawn in black, two spurious boundaries in blue and three undetected boundaries in red.

little more than a quarter of the entire domain. The target number of grains should therefore be close to 1830 with an average perimeter of 680. In both cases, the only regions prone to ambiguity are those in which misorientations are smaller than about 5° as otherwise grain boundaries are very sharp and clear.

Since the flood-fill approach uses a random atom as its starting point, the results may differ each time the algorithm is run. With the parameters chosen before, the numerical scheme detects between 128 and 130 grains with an average perimeter of 2360.4–2399.6 in the late distribution and between 1810 and 1827 grains with an average perimeter of 656.3–660.2 in the early distribution, representing a 2% variability.

Let us first compare the boundary networks in the late distribution, shown in figure 6(b). With 128 detected grains, five boundaries are in disagreement, three being found only in the hand segmentation and two spurious grains or artefacts being found numerically. These boundaries are all ambiguous and porous with a misorientation lying between 2° and 3°. The scheme is then able to detect almost all hand segmented grains while identifying a minimal number of artefacts. A similar comparison is shown in figure 7 where 1822 grains were extracted numerically. The agreement shown in (b) is quite good considering the small working scale. One very small grain has been missed while two more actual grains were detected, meaning that the hand segmentation could have been further refined in a few cases.

It remains to compare the geometric properties. We focus on comparing the area, perimeter and isoperimetric ratio since other properties mostly depend on the networks being well recovered. The agreement between the CDFs of the manual and numerical segmentations is shown in figure 8 for the late distribution. While the area CDFs agree remarkably well, there are some differences in comparing the perimeter and isoperimetric ratios that can be traced to very small grains. The perimeter of these grains is generally underestimated, thus causing the shift in the perimeter comparison below 2000 and that in the isoperimetric ratio above 0.75.
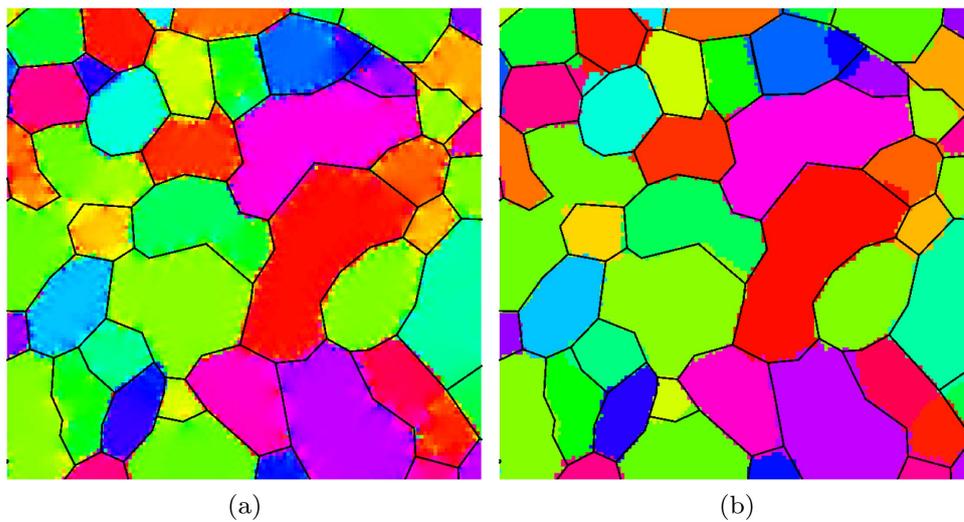
(a)   (b)

Figure 7. Detail of the visualization of local lattice orientations in the late distribution with manually segmented grain boundary network in black (a). Comparison between the manual and numerical segmentations (b); the numerically extracted grains are visualized in color while the manual segmentation is drawn in black.
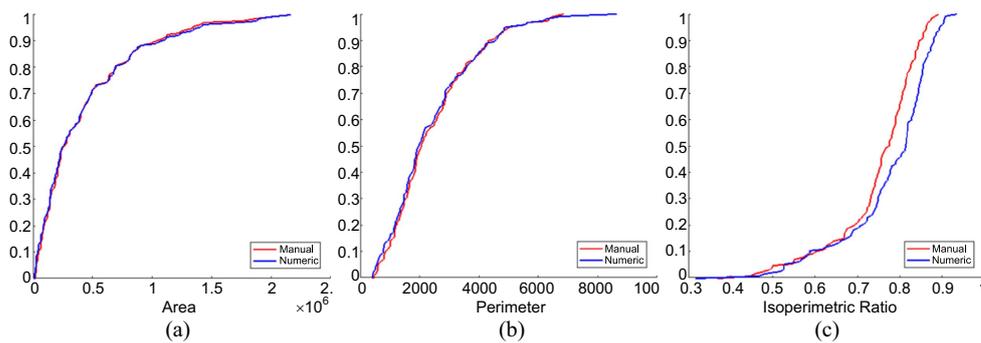

(a)   (b)   (c)

Figure 8. Comparison between the area (a), perimeter (b) and isoperimetric ratio (c) CDFs of the hand segmentation in red with the numerical extraction procedure in blue. Note that in (b), the grain with the largest numerical perimeter corresponds to the magenta/purple bottom right grain that was not properly split in figure 6.

Matching the numerically detected grains to their counterpart in the hand segmentation and ignoring outliers, for example when two grains have been merged, grain area is captured with less than 1% error while the perimeters are on average underestimated by about 2%. This is especially the case for very small grains whose perimeter is sometimes underestimated by as much as 20%. In the early distribution, areas are also recovered almost exactly while perimeters are underestimated by about 4%, highlighting a slight underestimation of perimeters depending on grain size.

Table 1. Dependence of results on γ (a), θ (b) and α (c).

(a)	Early distribution		Late distribution	
γ	Grains	Average perimeter	Grains	Average perimeter
0.0004	1698	698.4	128	2395.5
0.0006	1754	676.3	128	2387.1
0.0008	1804	661.9	126	2387.4
0.0010	1824	658.3	128	2375.5
0.0012	1841	653.7	129	2365.4
0.0014	1852	653.0	132	2333.5
0.0016	1866	651.9	129	2394.6
(b)	Early distribution		Late distribution	
θ (deg.)	Grains	Average perimeter	Grains	Average perimeter
1.0	2100	599.9	204	1697.2
1.5	1999	618.3	158	2043.5
2.0	1924	634.6	136	2283.6
2.5	1822	658.7	128	2383.1
3.0	1708	685.9	121	2467.3
3.5	1601	716.2	118	2538.1
4.0	1516	742.1	111	2684.7
(c)	Early distribution		Late distribution	
α	Grains	Average perimeter	Grains	Average perimeter
10	2062	602.4	150	2074.8
20	1898	639.6	134	2305.3
30	1878	644.8	131	2349.6
40	1817	659.0	128	2386.7
50	1785	667.5	130	2367.6
60	1743	678.2	127	2400.9
70	1692	690.9	130	2359.9
Target	1830	680	129	2426

3.3. Influence of the parameters

While the chosen parameters give rise to good agreement, their influence and choice is important to accurately measure grain properties. The atom extraction level h is chosen to maximize the radius of good atoms while recognizing poorly formed bad atoms. From visual inspection, this happens around -0.035 in our simulations. Other values of h miss atoms at boundaries, so the *next* shell of boundary atoms have very large Voronoi areas, thus introducing large overestimates in the perimeter computation. For our PFC parameters, this happens when h is roughly outside $[-0.01, -0.06]$. The effect of the three other parameters is more subtle. In the following tables, only one parameter was varied at a time from $\gamma = 0.001$, $\theta = 2.5°$ and $\alpha = 40$.

From table 1(a), γ has more impact on early results. Increasing γ increases the number of detected grains since smaller γ produce thicker bad atom masks that eventually block very small grains entirely. At 0.0014, more grains are found but these are generally small artefacts or spurious boundaries in ambiguous regions. Since the late results are stable, $\gamma = 0.001$ is chosen to prevent the introduction of artefacts. Part (b) shows that, as expected, less grains are

detected as θ is increased. At $2.0°$, a very large amount of spurious grains is detected suggesting that $2.5°$ is roughly the smallest threshold for which low misorientation boundaries can be characterized accurately. The results in (c) clearly show that increasing α decreases the number of detected small grains but does not affect the late distribution greatly. Below 40, almost all new grains that are detected are artefacts in both distributions. The value $\alpha = 40$ is again chosen to maximize the number of early grains without introducing many artefacts. Since the smallest features in the early distribution are about 60 atoms in size, choosing α as half the size of the smallest expected feature is a reasonable guideline.

The parameters chosen above are then reasonable to extract both the early and late distribution. We note however that accuracy could be improved slightly by further optimizing α and γ together. This might in principle produce optimal parameters that may differ between the two distributions. In such a case, it would be desirable to allow the parameters to vary dynamically as a function of grain size or other measured properties. Moreover, the optimal parameters could depend on the specifics of a given problem so similar tests should be done before extracting other types of data.

3.4. Comparison with Voronoi grain distributions

We now present a similar validation using an artificially constructed grain distribution so that areas and perimeters may be computed exactly. Given a target number N of grains on a domain of a given size, we compute the Voronoi tessellation of N randomly distributed points. Each Voronoi region is made to correspond to a grain whose angle is chosen randomly. Rotated lattices are placed onto the regions and passed through the PFC scheme to smooth out atoms at boundaries and the results are extracted. To compare with the scales used in PFC simulations, we let N vary between 125 and 2000 over the same grid size. Since grain misorientations below a certain threshold are ambiguous, we will not be able to recover all grains produced by the purely random Voronoi diagram. Instead, we can compute the proportion of grains in the diagram whose minimum misorientation is greater than the $2.5°$ threshold and take this number of grains as the target. The fraction of such grains is found to be independent of grain size and equals about 40%. Thus, roughly half of these grains will be merged to a neighbor, giving a target grain number of $0.8N$. Another approach, useful to compare the area and perimeter distributions directly, is to preemptively filter out neighbors with a small misorientation by changing one of their angle until all misorientations are greater than some threshold.

With the same extraction parameters as before, the algorithm is capable of finding the correct number of grains in both cases. When the angles are not filtered, an average of 79% grains are found across all lengthscales. Obviously, the average perimeter is overestimated at roughly 115%. When angles are pre-filtered, an average of 99% grains are detected since random Voronoi diagrams do not generate ambiguous boundaries and very small grains are rare. The average perimeter is recovered at 101% for $N = 125$ and at 98% for $N = 2000$. When matching grains, one finds that the error in the area is negligible and independent of grain size. On the contrary, the error in the perimeter depends on grain size and is consistent with the previous results: for $N = 125$, perimeter is recovered almost exactly while for $N = 2000$, grains are underestimated by 2%. Thus, the CDF of the isoperimetric ratios match almost exactly when there are 125 grains but are spaced by 0.03 when there are 2000 grains as shown in figure 9. These are quite slight deviations but when comparing the isoperimetric ratios on the same plot, this discrepancy does lead to confusion as the bias may be confused with evolution, meaning that isoperimetric ratio should only be compared qualitatively.
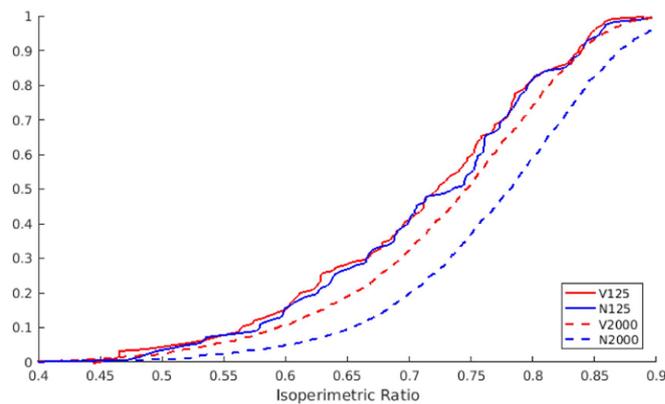
Figure 9. Comparison between the isoperimetric ratio CDFs of the 125 and 2000 grains Voronoi distributions in red and the numerically extracted results in blue. The solid and dashed lines correspond to the 125 and 2000 grains distribution respectively.

3.5. Comparison with simple geometric shapes

The accuracy of the scheme was also tested on simple geometric shapes. Grains with boundaries shaped like circles, squares and other shapes were created by embedding a misoriented grain of that shape within another and smoothing out atoms as before. The geometry of the inner grain was then extracted and compared to the expected value. This was done for various various sizes and misorientations but the later did not affect the results. Overall, the area was always recovered within ±0.5% accuracy while the perimeter was recovered with similar accuracy for curved shapes and otherwise underestimated with a slight bias consistent with the results found above. The reason the curved shapes are precisely matched in perimeter is because they can only be represented as polygons using atoms. Thus, the perimeter of the shape of the polygonal grain will be slightly larger than that of the true curved shape, canceling out the inherent bias of the scheme.

4. PFC sample results

Using the PFC equation, we evolved 18 initially noisy distributions over 40 000 time steps, saving the phase field at certain intervals. Geometric properties were extracted using the extraction parameters given before. All runs were added together to form much larger data sets containing more than roughly 2300 grains at the final time so that histograms become accurate in visualizing the evolution. We present the normalized reduced area, GBCD and area ratio distributions extracted using our algorithm in figure 10.

The normalized reduced area behaves as in [5], being at least approximately lognormal with a peak becoming flatter in time, stabilizing around 0.65. We note however that the relative proportion of very small and very large grains does continue to increase slightly as time goes on. Perimeter essentially tracks the reduced area while the isoperimetric ratio decreases, indicating that grains become qualitatively less regular in time. The absolute orientation distribution does not evolve in time and remains constant since the PFC equation is isotropic. On the other hand, the misorientation distribution and GBCD do evolve interestingly. The GBCD tracks the misorientation distribution but its features are more pronounced as there is some additional correlation between misorientation and interface length.
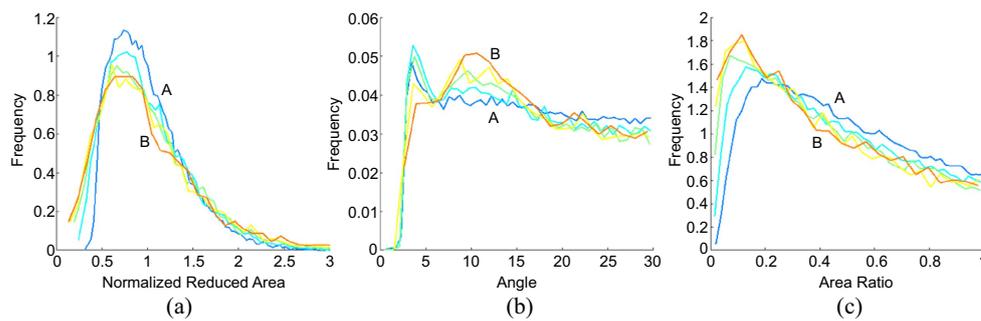
Figure 10. Evolution in time of the normalized reduced area (a), GBCD (b) and area ratio (c) CDFs with the PFC evolution. Lines correspond to the time steps {228, 752, 2828, 10637, 40000} with blue, A, being early in time and orange, B, late in time. Histograms have been normalized in area to make the comparison clearer.

Early in time, the GBCD is relatively flat away from the threshold but as time goes on, the misorientations around $10°$ become over represented to the detriment of small and large misorientations. The area ratio introduced earlier is also quite interesting. This property may be interpreted as follows: a flat area ratio distribution would indicate that the area of neighbor grains are completely uncorrelated. A peaked distribution close to 1 would indicate the neighbor grains are likely to be similar in size while a peak close to 0 would indicate that neighbors are likely to be very different in size. From figure 10(c), it is clear that the PFC evolution favors large area imbalances between neighbors. Finally, the coordination number does not evolve in time, peaking at 5 with an average of almost exactly 6 at all times.

5. Conclusion

Grain extraction being an important step in analyzing materials science data, there is a clear need for automated algorithms capable of extracting both the grain distribution and grain geometric properties. We detailed a simple, accurate and efficient atom based method to extract such grain networks along with grain area, perimeter and other properties. The detection accuracy of these measurements was tested using PFC grain distributions and surrogate artificial distributions. Overall, the boundary network and the number of grains may be extracted with very good precision. Grain area is measured with high accuracy while perimeter is underestimated by less than about 5% for very small grains.

The numerical method we have presented may then be used to automatically characterize the geometric properties of statistically significant data sets with good accuracy considering the difficulty and ambiguity inherent to the task. With modifications, our general atom based framework may be applied to 3D and to other crystal lattices. We shall use this scheme to analyze the PFC evolution in more detail in a future article.

Acknowledgments

ORCID iDs

Gabriel Martine La Boissonière ⬤ https://orcid.org/0000-0001-8848-8984

References

[1] Barmak K, Eggeling E, Kinderlehrer D, Sharp R, Taasan S, Rollett A D and Coffey K R 2013 Grain growth and the puzzle of its stagnation in thin films: the curious tale of a tail and an ear *Prog. Mater. Sci.* **58** 987–1055

[2] Mullins W W 1956 Two-dimensional motion of idealized grain boundaries *J. Appl. Phys.* **27** 900–4

[3] Elder K R, Katakowski M, Haataja M and Grant M 2002 Modeling elasticity in crystal growth *Phys. Rev. Lett.* **88** 245701

[4] Emmerich H, Löwen H, Wittkowski R, Gruhn T, Tóth G I, Tegze G and Gránásy L 2012 Phase-field-crystal models for condensed matter dynamics on atomic length and diffusive time scales: an overview *Adv. Phys.* **61** 665–743

[5] Backofen R, Barmak K, Elder K E and Voigt A 2014 Capturing the complex physics behind universal grain size distributions in thin metallic films *Acta Mater.* **64** 72–7

[6] Heilbronner R 2000 Automatic grain boundary detection and grain size analysis using polarization micrographs or orientation images *J. Struct. Geol.* **22** 969–81

[7] Berger A, Herwegh M, Schwarz J-O and Putlitz B 2011 Quantitative analysis of crystal/grain sizes and their distributions in 2d and 3d *J. Struct. Geol.* **33** 1751–63

[8] Singer H M and Singer I 2006 Analysis and visualization of multiply oriented lattice structures by a two-dimensional continuous wavelet transform *Phys. Rev.* E **74** 031103

[9] Berkels B, Rätz A, Rumpf M and Voigt A 2008 Extracting grain boundaries and macroscopic deformations from images on atomic scale *J. Sci. Comput.* **35** 1–23

[10] Elsey M and Wirth B 2014 Fast automated detection of crystal distortion and crystal defects in polycrystal images *Multiscale Model. Simul.* **12** 1–24

[11] Yang H, Lu J and Ying L 2015 Crystal image analysis using 2d synchrosqueezed transforms *Multiscale Model. Simul.* **13** 1542–72

[12] Stukowski A 2010 Visualization and analysis of atomistic simulation data with ovitothe open visualization tool *Model. Simul. Mater. Sci. Eng.* **18** 015012

[13] Stukowski A 2012 Structure identification methods for atomistic simulations of crystalline materials *Model. Simul. Mater. Sci. Eng.* **20** 045021

[14] Panzarino J F and Rupert T J 2014 Tracking microstructure of crystalline materials: a post-processing algorithm for atomistic simulations *J. Miner., Met. Mater. Soc.* **66** 417–28

[15] Holm E A, Rohrer G S, Foiles S M, Rollett A D, Miller H M and Olmsted D L 2011 Validating computed grain boundary energies in fcc metals using the grain boundary character distribution *Acta Mater.* **59** 5250–6

[16] Okabe A, Boots B, Sugihara K, Chiu S N and Kendall D G 2000 *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* 2nd edn (New York: Wiley)

[17] Braden B 1986 The surveyors area formula *Coll. Math. J.* **17** 326–37

[18] Ofori-Opoku N, Fallah V, Greenwood M, Esmaeili S and Provatas N 2013 Multicomponent phase-field crystal model for structural transformations in metal alloys *Phys. Rev. B* **87** 134105

[19] Seymour M and Provatas N 2016 Structural phase field crystal approach for modeling graphene and other two-dimensional structures *Phys. Rev.* B **93** 035447

[20] Elsey M and Wirth B 2013 A simple and efficient scheme for phase field crystal simulation *ESAIM: Math. Model. Numer. Anal.* **47** 1413–32