PAPER • **OPEN ACCESS**

On a probabilistic model for martensitic avalanches incorporating mechanical compatibility

OPEN ACCESS
IOP Publishing | London Mathematical Society Nonlinearity
Nonlinearity **34** (2021) 4844–4896 https://doi.org/10.1088/1361-6544/abfca9

On a probabilistic model for martensitic avalanches incorporating mechanical compatibility

Francesco Della Porta[1], Angkana Rüland[2], Jamie M Taylor[3] and Christian Zillinger[4,*]

[1] Max-Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany
[2] Ruprecht-Karls-Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
[3] Basque Center for Applied Mathematics, Mazarredo, 14, 48009 Bilbao Basque Country, Spain
[4] Karlsruhe Institute of Technology, Englerstrasse 2, 76131 Karlsruhe, Germany

E-mail: DellaPorta@mis.mpg.de, Angkana.Rueland@uni-heidelberg.de, taylor@bcamath.org and christian.zillinger@kit.edu

Received 6 September 2020, revised 8 March 2021
Accepted for publication 28 April 2021
Published 30 June 2021

Abstract

Building on the work by Ball *et al* (2015 *MATEC Web of Conf.* **33** 02008), Cesana and Hambly (2018 A probabilistic model for interfaces in a martensitic phase transition arXiv:1810.04380), Torrents *et al* (2017 *Phys. Rev. E* **95** 013001), in this article we propose and study a simple, geometrically constrained, probabilistic algorithm geared towards capturing some aspects of the nucleation in shape-memory alloys. As a main novelty with respect to the algorithms by Ball *et al* (2015 *MATEC Web of Conf.* **33** 02008), Cesana and Hambly (2018 A probabilistic model for interfaces in a martensitic phase transition arXiv:1810.04380), Torrents *et al* (2017 *Phys. Rev. E* **95** 013001) we include *mechanical compatibility*. The mechanical compatibility here is guaranteed by

Recommended by Dr Claude Le Bris.

*Author to whom any correspondence should be addressed.

using *convex integration building blocks* in the nucleation steps. We analytically investigate the algorithm's convergence and the solutions' regularity, viewing the latter as a measure for the fractality of the resulting microstructure. We complement our analysis with a numerical implementation of the scheme and compare it to the numerical results by Ball *et al* (2015 *MATEC Web of Conf.* **33** 02008), Cesana and Hambly (2018 A probabilistic model for interfaces in a martensitic phase transition arXiv:1810.04380), Torrents *et al* (2017 *Phys. Rev. E* **95** 013001).

Keywords: convex integration, avalanches, mechanical compatibility, shape-memory alloys, probabilistic, geometrically constrained nucleation

Mathematics Subject Classification numbers: 74N05, 74A50, 74N15, 74B20, 74G65, 35Q74, 60G99.

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

1. Introduction

Shape-memory alloys are materials displaying a striking thermodynamical behaviour on the one hand and a rich mathematical structure on the other hand. Physically, these materials undergo a first-order, diffusionless, solid-solid phase transformation in which symmetry is reduced upon the passage from the high temperature phase, *austenite*, to the low temperature phase, *martensite*. This reduction of symmetry gives rise to various *variants of martensite* in the low temperature regime.

Mathematically, these materials have been successfully described within the calculus of variations by minimization problems of the form

$$\text{minimize} \int_{\Omega} W(\nabla \mathbf{y}, \vartheta)\, \mathrm{d}x, \tag{1}$$

for instance, with prescribed displacement boundary conditions [Bal04, BJ92, Bal02, BJ89, Bha03, Mül99]. Here $\Omega \subset \mathbb{R}^n$ is the reference configuration, $\vartheta : [0,\infty) \to [0,\infty)$ denotes temperature and $\mathbf{y} : \Omega \to \mathbb{R}^n$ is the deformation describing how the reference configuration is deformed. Defining $\mathbb{R}^{n\times n}_+$ to be the set of n by n matrices with positive determinant, the stored energy function $W : \mathbb{R}^{n\times n}_+ \times [0,\infty) \to \mathbb{R}_+$ describes the energetic cost of a deformation at a given temperature. Physical requirements on it are *frame indifference*, i.e. the fact that

$$W(\mathsf{F}) = W(\mathsf{QF}) \ \text{ for all } \mathsf{F} \in \mathbb{R}^{n\times n}_+,\ \mathsf{Q} \in SO(n),$$

and *material symmetry*, i.e. the fact that

$$W(\mathsf{F}) = W(\mathsf{FH}) \ \text{ for all } \mathsf{F} \in \mathbb{R}^{n\times n}_+,\ \mathsf{H} \in \mathcal{P},$$

where $\mathcal{P}$ denotes the (discrete) symmetry group which sends the austenite lattice into itself.

In particular, the zero set—or physically the set of *exactly stress-free strains*—associated with W is typically of the form

$$K(\vartheta) = \begin{cases} \alpha(\vartheta)SO(n)Id & \text{for } \vartheta > \vartheta_c, \\ \bigcup_{j=1}^{N} SO(n)\mathsf{U}_j(\vartheta_c) \cup SO(n)Id & \text{for } \vartheta = \vartheta_c, \\ \bigcup_{j=1}^{N} SO(n)\mathsf{U}_j(\vartheta) & \text{for } \vartheta \leqslant \vartheta_c. \end{cases}$$

Here the matrices $\mathsf{U}_j(\vartheta) \subset \mathbb{R}^{n\times n}_+$ are obtained through conjugation of $\mathsf{U}_1(\vartheta)$ by elements from $\mathcal{P}$ and represent the N variants of martensite, while $\alpha(\vartheta) : \mathbb{R}_+ \to \mathbb{R}_+$ models the thermal expansion of the underlying lattice depending on temperature, with the convention that $\alpha(\vartheta_c) = 1$. In order to study low energy configurations of (1), a common strategy [BJ89, Bha03, CDK07, DM95, Kir03, Kir98, Rül16a, Rül16b, Sim17] is to first study exactly stress-free deformations by investigating the differential inclusion

$$\nabla \mathbf{y} \in K(\vartheta). \tag{2}$$

While the study of the minimization problem (1) has proved very successful and influential, e.g. in predicting interfaces between variants of martensite and scaling laws [BJ92, KM94, KKO13, KK11, Con00, CO12, CO09, Rül16b], it is often the case that the dynamics of the phase transition play an important role in the formation of the complex microstructures observed in experiments (see e.g., [SCD13] and related comments in [DP19a]). Indeed, observed microstructures are often the result of different smaller microstructures, nucleating at different points of the domain and expanding. In order to preserve continuity of the deformation, and hence compatibility, these microstructures refine and become more complex at the interfaces at which they encounter. However, they do not globally minimise an energy functional penalising interfaces between martensitic junctions. Such complex evolution has been observed both with optical microscopy, a common tool to analyse martensitic microstructure, and by phonon emission measurements, a second method based on the observation that every nucleation event is accompanied by an acoustic emission (see e.g., [BBG20]). In particular, both methods have thus been used for tracking the dynamics of nucleation phenomena. High time resolution measurements of the described type display strongly intermittent behaviour and the presence of 'avalanches' [PMV13, SKR09] with 'universal', power law behaviour for central statistical quantities.

Based on these and related observations, it has been the objective of several recent works to study simplified dynamic models of phase transformations in shape-memory alloys (see also section 1.2 for further, more detailed physical models and explanations): on the one hand, the continuum mechanical models in [DP19a, DP19b] seek to capture the evolution of the microstructures and the mechanical effects that the dynamics may have on them based on *optical* microscopy observations. On the other hand, in parallel, simplified probabilistic, geometrically constrained dynamic models have been proposed and investigated in the literature [BCH15, CH18, TIVP17]—both in the mathematical and the physics community. The latter aim at predicting the above mentioned *acoustic* observations and at deriving an improved understanding of the 'universal', power law behaviour for central statistical quantities. As shown in [TIVP17, figure 1], these probabilistic, geometrically constrained models sometimes also successfully reflect the 'wild', 'random', irregular microstructures observed in optical microscopy.

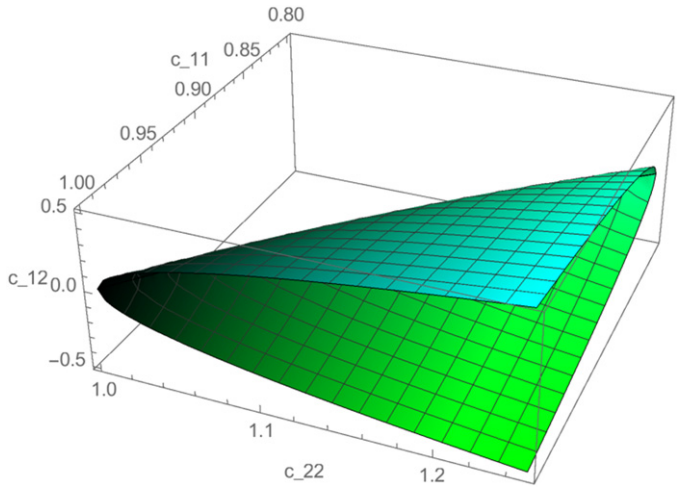


Figure 1. The quasiconvex hull K^{qc}, i.e. the set of all macroscopically realizable deformations, associated with the set K, depicted in Cauchy–Green space, see section 3 for more detailed definitions. Here the wells from (3) correspond to the two corners of the paraboloid (and are coloured cyan and green, respectively). All other matrices in the depicted set are obtained as Cauchy–Green tensors of first or second order laminates (corresponding to the boundaries of the paraboloid and its relative interior, respectively). The colour coding here is the colour coding which is used for vertical twins (see the explanations below). In order to illustrate the difference between horizontal and vertical twins (in the sense of [DPR20]), we use a second colour scheme (see figure 3).

In order to capture the avalanching phase transformation dynamics, the models proposed in [BCH15, CH18, TIVP17] take into account two key features which are believed to be characteristic of many martensitic phase transformations:

- During the phase transformation a domain which has transformed from austenite to martensite does *not* transform back (see also the moving mask hypotheses in [DP19a]).
- The nucleation domains are given by long (needle-like) domains ('plates') which are oriented according to the rank-one connections which are present between the wells (see e.g., the experimental results in [IHM13] and cf once more the moving mask hypotheses in [DP19a]).

Based on this, the models in [BCH15, CH18, TIVP17] roughly propose the following simplified, geometrically constrained nucleation mechanisms:

(a) Choose a point randomly out of the sample/reference configuration and choose a direction (out of the possible rank-one directions, i.e. out of the directions of compatibility between austenite and a martensitic plate) randomly.
(b) Nucleate a martensitic plate in the chosen direction through the chosen point until it hits another plate or the boundary of the sample.
(c) Iterate this.

We emphasize that this leads to a purely 'scalar' model which is not formulated on the level of the deformation gradients and, in particular, does *not* take into account any compatibility of the associated deformation gradients beyond the fact that the nucleated plates should roughly be aligned with the rank-one directions. Numerical simulations of these dynamics lead to highly fractal, self-organized, 'wild' structures in the martensitic materials. Based on the

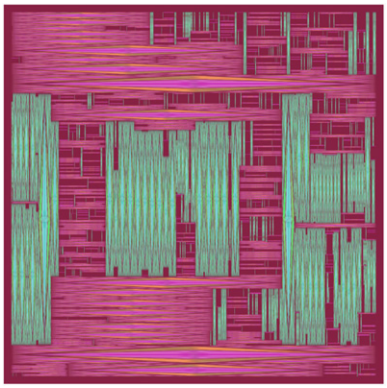


Figure 2. The random convex integration solutions produced by (a variant) of our algorithms (the picture here is generated by means of the modifications explained in section 8.1). The colour coding uses cyan and green for vertical and magenta and orange for horizontal twins. In addition to the fractal behaviour originating from the random (greedy type) covering which is observed in [BCH15, CH18, TIVP17], we here have a second source of fractality originating from the use of the convex integration building blocks within the random rectangle covering (see section 3 below for more comments on this). In this simulation we have used the parameters $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$.

described dynamics, in their analysis and simulations in [BCH15, CH18, TIVP17] the authors derive properties of the statistical distribution of martensitic plates and deduce self-similarity and power law behaviour in certain regimes. This may indicate that, in spite of the drastic simplifications, the geometrically imposed constraints could indeed provide insights into the experimentally measured universal exponents in the nucleation experiments.

It is the objective of this article, to propose and investigate an *intermediate* model building on the geometrically constrained, probabilistic models from [BCH15, CH18, TIVP17], capturing both the random, geometrically constrained, self-organizing behaviour and including the key mechanical aspect of *compatibility*. In the previous works on purely geometrically constrained, probabilistic models the latter had only been taken into account in terms of fixing the orientation of the martensitic plate and not in terms of the associated deformation (see also the more complex physical models discussed in section 1.2 from the mechanics community). As in [BCH15, CH18, TIVP17] we are also interested in studying the universality properties of solutions. In contrast to these results we however focus on the *regularity of solutions* as a measurement for this and the 'wildness' of the microstructure, and interpret regularity as the main quantity from which statistical properties could be deduced (see section 1.3 below). We further link this to our recent investigation of deterministic 'wild' microstructures obtained through the method of convex integration (figures 1 and 2).

1.1. The model and the main results

In the sequel, as a model setting, we focus on the geometrically nonlinear, two-dimensional two-well problem. Extensions to other models would not pose any difficulties as pointed out in our discussion below. Fixing temperature below the transformation temperature, we thus consider

$$K = SO(2)\mathsf{F}_0 \cup SO(2)\mathsf{F}_0^{-1} \tag{3}$$

where $\mathsf{F}_0, \mathsf{F}_0^{-1} \in \mathbb{R}^{2\times 2}$ are respectively given by

$$\mathsf{F}_0 = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}, \qquad \mathsf{F}_0^{-1} = \begin{bmatrix} 1 & -\gamma \\ 0 & 1 \end{bmatrix}, \tag{4}$$

and $\gamma > 0$.

In the present work we propose dynamics which are strongly inspired by the ones in [BCH15, CH18, TIVP17] but which take *mechanical compatibility* into account. More precisely, essentially our nucleation algorithms still follow the steps (a)–(c) from above, with the main difference that condition (b) is now formulated on the level of the full deformation gradients (instead of the scalar order parameters from [CH18, TIVP17]). Therefore, the plates which are nucleated in (b) are now prescribed in a *compatible* way (in the sense of not creating any stresses). This is achieved by relying on convex integration building blocks which are exactly stress-free solutions to the differential inclusion (2) at a fixed temperature (in our case below the critical temperature) and with prescribed displacement boundary conditions. As in [BCH15, CH18, TIVP17] for the setting of the two-well problem this gives rise to two specific orientations of the martensitic plates which are however now *exact solutions* to the differential inclusion. In the infinite iteration/time limit, we thus obtain exactly stress-free solutions resembling those of [BCH15, CH18, TIVP17] which now however are defined on the level of the deformations and in particular include compatibility and (up to a set of measure zero) fully transform Ω. The precise algorithms used in our dynamics are described in algorithms 3.3 and 3.4 in section 3 below. Let us remark that such a behaviour reminds of experimental observations in TiNbAl (see e.g., [Ia]) where, after the phase transition, it is possible to observe different colonies of 'wild' microstructures.

As in [BCH15, CH18, TIVP17] we seek to show that these dynamics give rise to 'power-law' behaviour and self-organized structures in a probabilistic sense (see section 8.3 for some numerical evaluations of the length scale statistics). As already indicated above, we do not aim at proving direct power-law distributions for the present lengths scales but view the *regularity* of solutions as a proxy for this which encodes important statistical information (e.g. in terms of the solutions' heavy tailed Fourier distribution etc).

Our main analytical result for these dynamics is summarized in the following theorem. Here algorithms 3.3 and 3.4 are two variants of the algorithm with the steps (a)–(c) from above, now including compatibility. The detailed models will be outlined in section 3.

Theorem 1. *Let K be as in (3)–(4) and let $\tilde{\Omega} = (0,1)^2 \subset \mathbb{R}^2$. Let $\{\mathbf{y}_k\}$ denote the sequences obtained in algorithms 3.3 or 3.4 (defined in section 3) and let μ denote the corresponding probability measure (constructed in detail in section 4). Then there exists $\theta_{A,B} > 0$ such that for all $s \in (0,1)$, $p \in (1,\infty)$ with $sp < \theta_{A,B}$ and for all $\mathsf{M} \in \operatorname{int} K^{\mathrm{qc}}$ and μ-almost every sequence $\{\mathbf{y}_k\}$ there exists a deformation $\mathbf{y} : \tilde{\Omega} \to \mathbb{R}^2$ such that $\mathbf{y}_k \to \mathbf{y}$ in $W^{1,1}(\tilde{\Omega})$ and*

$$\begin{aligned} \nabla \mathbf{y} &\in K \ \ \text{a.e. in } \tilde{\Omega}, \\ \mathbf{y} &= \mathsf{M}\mathbf{x} \ \ \text{on } \partial\tilde{\Omega}, \end{aligned} \tag{5}$$

with $\mathbf{y} \in W^{1,\infty}(\tilde{\Omega};\mathbb{R}^2) \cap W^{1+s,p}(\tilde{\Omega};\mathbb{R}^2)$.

Remark 1.1. It would be possible to extend the result to domains $\tilde{\Omega} \subset \mathbb{R}^2$ which are more complicated (e.g. domains which can be written as controlled (in-)finite unions of rectangles). In order to avoid dealing with the associated issues and as the domain geometry does not constitute our main focus in this article, we restrict to the above model setting in which $\tilde{\Omega} = (0,1)^2$.

We emphasize that essentially all sequences (in terms of μ) produced in our dynamics in the infinite iteration/long time limit lead to exactly stress-free solutions of the differential inclusion. Moreover, they have a certain fractality (and are in this sense self-organized and not completely random) as encoded in the higher Sobolev regularity result with $\nabla \mathbf{y} \in W^{s,p}(\tilde{\Omega}, \mathbb{R}^2)$ (see also the remarks below). Numerical evaluations of the algorithms are presented in section 8, in which we also discuss the lengths scale statistics involved in the solutions and highlight the different convergence rates. We hope that this may eventually allow for comparisons with the measured length scale (and avalanche) distributions in experimental settings.

1.2. Context

Self-organized, critical systems and cellular automata have attracted substantial interest in systems undergoing phase transformations (see [BTW87] and the large amount of literature building on this). Also in the mechanical literature there have been substantial endeavours towards understanding this more precisely and various explanations have been proposed for the origins of these effects. This includes pinning–depinning transitions [2005], coevolving disorder associated with transformation-induced slip [PRTZT16], inertia-induced nucleation [AA01] or the presence of quenched disorder [PRV04]. We also refer to [PRTZ08, PRTZ07, PRTZ09, BBB15, PRTZT16, BUZZ16] and the references therein. We emphasize that both scalar and vectorial models have been considered in these references. In the context of martensitic phase transformations and self-similarity we highlight the early works [RSS95, MRSS95, PLKK97] in which random, geometrically constrained models had been proposed and analysed in the study of self-organized structures in martensitic phase transformations. Already in these, the emergence of self-similar, fractal microstructures was observed.

The models proposed in this article follow the line of ideas introduced in [RSS95, MRSS95, PLKK97, BCH15, CH18, TIVP17]. It is our main objective to explore how simplified, geometrically constrained, probabilistic dynamics may lead to universal power law behaviour in nucleation processes as observed, for instance, through acoustic emission measurements. In restricting ourselves to these models, compared to the above cited approaches from the mechanics literature, we substantially simplify the mechanisms in the problem, losing various physically-relevant effects. This for instance neglects the relevant time scales (which are simply prescribed and normalized) or the relevance of long-range effects (which are captured only minimally by the history of the stochastic process but not made explicit in the form of non-local elastic energy contributions). Physical effects like pinning–depinning transitions, disorder due to transformation-induced slip are not resolved but simply subsumed in the randomness of the algorithms. Moreover, we focus our attention on global minima although certainly local minimization and metastability are of physical significance in avalanche models. While we are aware of these simplifications, we emphasize that *in spite* of the outlined reduction of the complexity of our model, it still displays 'universal behaviour' in the sense of bounds on the frequency distributions in Fourier space. Indeed, *due to* its simplicity, it is possible to establish the Fourier bounds rigorously. Thus, in our model randomness should be viewed as a proxy for partially capturing the outlined physical aspects which we do not resolve. It would be interesting to combine ingredients of our model with the resolution of additional physical effects.

While the earlier probabilistic, geometrically constrained models from [BCH15, CH18, TIVP17] did not take into account mechanical *compatibility* conditions, by connecting the probabilistic models from above with convex integration building blocks, our model does take this into account. In particular it allows us to link the 'self-organized' model dynamics from [BCH15, CH18, TIVP17], convex integration schemes [MŠ98, MS01, MŠ99]—which have a

natural dynamic interpretation—and the recently obtained higher Sobolev regularity results for convex integration solutions [RZZ19, RZZ18, RTZ18, DPR20]. It is our hope that with further simulations, experiments and analytical investigations these connections can be strengthened and that eventually the obtained regularity exponents can be compared to the observed universal exponents of the (length scale) statistics in the experiments. From a mathematically point of view, the connection of the proposed model and 'random', average convex integration algorithms in which only the *average* instead of *tailor-made packings* are considered also seems to be of independent interest (we also refer to [Kir03] and [Pom] for random walk interpretations of convex integration procedures). We emphasize that our model should be viewed as a *hybrid* model connecting the ideas from [PRTZ08, PRTZ07, PRTZ09, BBB15, PRTZT16, BUZZ16] and from convex integration with higher Sobolev regularity from [RZZ19, RZZ18, RTZ18, DPR20]. For the sake of mathematical simplicity in this first treatment of probabilistic models involving convex integration we separate the two ingredients, the probabilistic point of view and the convex integration scheme as much as possible. Building on this, as next steps, possibly slightly more natural algorithms could include a simultaneous iteration of the convex integration schemes and the random choice of the nucleation spots and the building block directions. These (possibly energetically more justified) models however lead to significantly more complicated analytical problems. Seeking to introduce a coupling between the ideas of convex integration (and thus of *compatibility*) and the random, geometrically constrained (and thus *self-organized*) structures from [PRTZ08, PRTZ07, PRTZ09, BBB15, PRTZT16, BUZZ16], we here focus on the simplest possible setting, but plan to study the indicated, more complex structures in future projects.

1.3. Regularity, self-similarity and power law length scale distributions

Last but not least, we seek to heuristically connect the regularity of solutions to (5) and the power-law behaviour of statistical quantities such as length scale distributions. Precise relations between the (maximal) regularity of solutions and scaling laws are deduced in [RTZ18]. On an L^2 based level the higher H^s Sobolev regularity of the deformation gradient $\nabla \mathbf{y}$ corresponds to the finiteness of the integral

$$\int_{\mathbb{R}^2} |\mathbf{k}|^{2+2s} |\mathcal{F}\mathbf{y}(\mathbf{k})|^2 \, \mathrm{d}\mathbf{k}. \tag{6}$$

In particular this implies that $\mathcal{F}\mathbf{y}(\mathbf{k})$ (the Fourier transform of $\mathbf{y}$ at $\mathbf{k}$) necessarily has a decay rate that (in an average sense) is determined by the Sobolev regularity of $\mathbf{y}$. Assuming that $\mathcal{F}\mathbf{y}(\mathbf{k})$ is of a power law distribution, i.e. that $|\mathcal{F}\mathbf{y}(\mathbf{k})| \sim |\mathbf{k}|^{-\alpha/2}$ for $|\mathbf{k}| \geqslant 1$ and some $\alpha \in \mathbb{R}$, the finiteness condition for (6) would imply a power law behaviour of the length scales involved in $\mathcal{F}\mathbf{y}(\mathbf{k})$ of the order at least $\alpha > 4 + 2s$. Combined with scaling laws for the associated elastic and surface energies, one would also be able to provide upper bounds on α as explained in [RTZ18]. In this sense, the Sobolev regularity captures the degree of self-organization in a precise sense. Similar, Fourier based considerations (for two-point functions) as a measure of the fractality or degree of self-organization of a solution can be found in [PLKK97].

1.4. Outline of the remainder of the article

The remainder of the article is organized as follows: after briefly collecting our most important notation in section 2, we present our models in section 3. In order to fix the precise setting these are complemented with the precise probabilistic set-up in section 4. A first convergence result for our algorithms is discussed in section 5. In sections 6 and 7 the higher Sobolev

regularity and the μ-almost everywhere convergence of the algorithms is studied. Last but not least, in section 8, we provide several illustrations of the numerical implementation of our algorithms—highlighting their different convergence properties—and their statistics. We hope that these are of use in eventually comparing our results with experimental data.

Complementing the results from the main body of the text, we have included four appendices: in appendix A a useful covering lemma is proved. In the appendix B we deduce a coupling between the random sequences generated by the algorithms 3.3 and 3.4; in particular proving that they generate equivalent limiting distributions. Moreover, in analysing the underlying fragmentation process in more detail in appendix C, we explain how (for a slight variant of) algorithm 3.3 it is possible to dispose of the non-degeneracy condition in the algorithms and to replace this by appropriate 'tail estimates'. We view this as an analytically important result of the article connecting our set-up from the main body of the text to the detailed analysis of fragmentation models. In appendix D we prove that the modifications of the numerical implementation do not affect the regularity of the resulting deformations.

2. Notation

For the convenience of the reader, we collect some of the notation which will be used in the following sections. We first collect the central notation from the algorithms 3.3 and 3.4 at step k:

- $\mathcal{V}_k$—this is the still not transformed part of the domain $(0,1)^2$ in the iteration step k, it consists of a finite union of open rectangles,
- $\mathcal{C}(\mathcal{V}_k)$—this is the set of connected components of $\mathcal{V}_k$,
- $\mathbf{p}_k$—this is the randomly chosen point in the algorithms,
- $\mathcal{C}(\mathcal{V}_k, \mathbf{p}_k)$—is the connected component of $\mathcal{V}_k$ containing $\mathbf{p}_k$,
- d_k—this is the randomly chosen orientation in the algorithms,
- $\mathbf{y}_k$—this is the current deformation,
- $\mathbf{z}_k^j$—this is the replacement building block given by theorem 2,
- $\mathcal{B}_k$, $\mathcal{B}_k^j$—these are the sets on which the current deformation is replaced by a deformation which is in the wells.

In our discussion of the probabilistic background we use the following notation:

- $\Omega^k := \Omega \times \cdots \times \Omega$ (k-times)—the k-fold Cartesian product of a set $\Omega \subset \mathbb{R}^2$,
- $\mathcal{B}(\Omega)$—the Borel sets on Ω,
- μ, μ_k, ρ_k—the measures constructed in lemmas 4.1 and 4.2,
- $\mathbb{E}$, $\mathbb{E}_k$—expectations with respect to the measures μ and μ_k, by construction μ is an extension of μ_k, so $\mathbb{E}$ reduces to $\mathbb{E}_k$ for finite iterations of our algorithms,
- $|A|$—Lebesgue measure of a Lebesgue measurable subset $A \subset \mathbb{R}^n$,
- $\mathcal{D}(D)$—the descendants of a set D, see definition 5.5.

3. The models

As a model setting, we consider the energy wells determined by the strains F_0 and F_0^{-1} from (3) and (4). We remark that, as shown in [BJ92, section 5], given two wells with two rank-one connections (and the physically natural condition of equal determinant), one can always reduce the problem to our case via an affine change of variables.

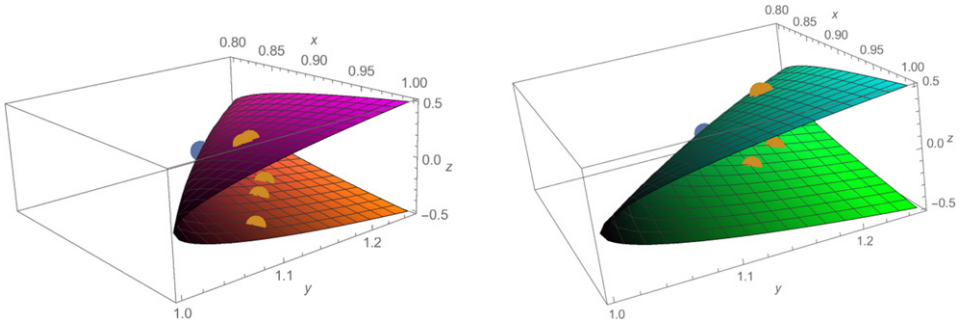


Figure 3. Representation of the iterative steps of the convex iteration algorithm in [DPR20] with boundary conditions $M \approx \mathrm{diag}(0.939, 1.064)$, $\gamma = 0.5$ are shown in terms of the Cauchy–Green tensors. Here the coordinates are $x = (\mathsf{M}^{\mathrm{T}}\mathsf{M})_{11} \in (0, 1)$, $y = (\mathsf{M}^{\mathrm{T}}\mathsf{M})_{22} \in (0, 1 + \gamma^2)$ for the plane directions and $z = (\mathsf{M}^{\mathrm{T}}\mathsf{M})_{12} = (\mathsf{M}^{\mathrm{T}}\mathsf{M})_{21} = \pm\sqrt{1 - xy}$ for the vertical direction. The closer the steps are to $\mathsf{F}_0, \mathsf{F}_0^{-1}$ (that is to the points with $(x, y) = (1, 1 + \gamma^2)$) the closer the algorithm is to convergence. Here we use a magenta–orange–black color coding for the horizontal replacement (horizontal rectangle) and cyan–green–black for the vertical one (vertical rectangle).

3.1. The building blocks: higher regularity convex integration solutions for the geometrically nonlinear two-well problem

Below we are going to rely on the following higher regularity convex integration theorem:

Theorem 2 (Theorem 1, [DPR20]). *Let K be as in (3)–(4). Let $\tilde{\Omega} \subset \mathbb{R}^2$ satisfy*

(*D*) *$\tilde{\Omega}$ is open, connected, and can be covered (up to a set of measure zero) by finitely many open disjoint triangles.*

Then there exists $\theta_0 > 0$ (independent of $\tilde{\Omega}$) such that for all $s \in (0, 1), p \in (1, \infty)$ with $sp < \theta_0$ and for all $\mathsf{M} \in \mathrm{int}\, K^{\mathrm{qc}}$ there exists a deformation $\mathbf{u} : \tilde{\Omega} \to \mathbb{R}^2$ such that

$$\nabla \mathbf{u} \in K \text{ a.e. in } \tilde{\Omega},$$

$$\mathbf{u} = \mathsf{M}\mathbf{x} \text{ on } \partial\tilde{\Omega},$$

$$\mathbf{u} \in W^{1,\infty}(\tilde{\Omega}; \mathbb{R}^2) \cap W^{1+s,p}(\tilde{\Omega}; \mathbb{R}^2).$$

We will use the solutions from theorem 2 as building blocks for our 'plates' (see step (b) of the probabilistic nucleation algorithms explained in the introduction, see also figure 5 for an illustration of a building block for $\mathsf{M} \approx \begin{pmatrix} 0.939 & 0 \\ 0 & 1.064 \end{pmatrix}$, $\delta = 0.05$ and $\gamma = 0.5$).

The solutions from theorem 2 are obtained iteratively through the method of convex integration, by iteratively deforming the current gradient distribution into an increasingly favourable one, eventually in the infinite iteration limit passing to a solution of the full differential inclusion (5) (see figure 3).

Here in each step we cover a rectangle in the given domain by 'rhombi-constructions' (see lemma 4.1 and figure 4 in [DPR20], building on the rhombi-constructions from the works [Con08, CT05, MŠ98]) which are needle-like basic building blocks (see figure 4).

We remark that in this iterative replacement of deformation gradients, there are two favoured orientations for the rhombi-constructions (and thus for building blocks). These correspond

Figure 4. Given the matrix decomposition, we employ 'rhombi-construction' on a diamond-shaped domain. The green colour coding of figure 3 right corresponds to vertical twins, the magenta colour coding of figure 3 left to horizontal twins. In the notation of [DPR20] these correspond to the coordinates F_1 and F_2, respectively.

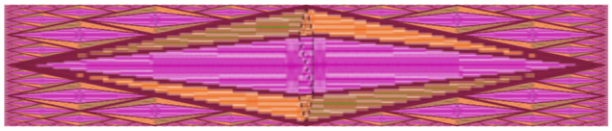


Figure 5. We use theorem 2 to construct building blocks in the shape of rectangles (which themselves are covered by rhombi-constructions in the form of diamonds, see figure 4). Different colours here correspond to different values of $\nabla \mathbf{u}$ with the colour coding given as in figure 3. For horizontal rectangles in the algorithms below we always begin with a decomposition along horizontal laminates, i.e. in the magenta–orange colour coding scheme. In particular from the colours in this figure it is clear that the underlying deformation is not yet a full solution (but only a subsolution, roughly speaking an approximate solution) to the differential inclusion. The construction of solutions to theorem 2 is iterative. We have here depicted a subsolution obtained after three iterations for the choice $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$.

to the horizontal and vertical rank-one directions which are present between the wells (see lemma 4.1 and figure 4 in [DPR20]). Thus, the choice of the orientation (of the needle-like nucleation domains) which was only heuristically justified in (ii) in [BCH15, CH18, TIVP17] now becomes a rigorously justified consequence of compatibility. In order to avoid additional difficulties in the covering estimates and to keep closer to the models from [BCH15, CH18, TIVP17], we do not directly work with the diamond-shaped rhombi-constructions as the basic building blocks but consider rectangles oriented according to the rhombi-constructions which are then themselves covered by rhombi-constructions (see figures 4 and 5).

Remark 3.1. Instead of focusing on the geometrically nonlinear two-well problem, we could also have used the results in [RZZ19] or in [RZZ18] instead of theorem 1 in [DPR20]. As a consequence, all the results which are deduced below for the geometrically nonlinear two-well problem would similarly hold in these settings.

Remark 3.2. We remark that in (two-dimensional) situations with exceptionally high symmetry of the martensitic wells, it would also be possible to use the building block structures from [CKZ17] and [CDPR20] as the 'plates of nucleation' together with greedy covering results. It is however not clear whether such constructions are also available in three dimensions; hence we do not pursue this further in this article.

3.2. Two geometrically constrained, probabilistic nucleation algorithms with mechanical compatibility

With this background, we next introduce two possible models for simplified, geometrically constrained, *mechanically compatible* nucleation dynamics. We emphasize that these dynamics are purely phenomenological and are not derived from first principles. Their main objective is to provide further insight into the observed phenomena of universal exponents in martensitic phase transformation by means of simplified dynamics now *including compatibility*. Further we seek to indicate how convex integration algorithms could naturally play a role in these types of dynamics.

3.2.1. Model A. We first present an algorithms which nucleates a new martensitic plate in each connected component of austenite in each iteration step.

Algorithm 3.3 (Model A). Let $\Omega = (0,1)^2$, $\delta \in (0,1)$ and:

- $\mathsf{M} \in K^{\mathrm{qc}}$;
- $\mathbf{y}_0 := \mathsf{M}\mathbf{x}$ in Ω;
- $\mathcal{V}_0 = \Omega$.

Then, for any $k \in \mathbb{N}$

- Let $\mathcal{C}(\mathcal{V}_{k-1})$ be the set of connected components of $\mathcal{V}_{k-1}$ (these are at most 2^k rectangles)
- Let $\mathbf{p}_k : \mathcal{C}(\mathcal{V}_{k-1}) \to \mathbb{R}^2$ be a function associating to each element $D^j_{k-1} \in \mathcal{C}(\mathcal{V}_{k-1})$ a point $\mathbf{p}^j_k$ chosen uniformly at random in D^j_{k-1}
- Let $d_k : \mathcal{C}(\mathcal{V}_{k-1}) \to \{1,2\}$ be a function associating to each $D^j_{k-1} \in \mathcal{C}(\mathcal{V}_{k-1})$ an orientation d^j_k (horizontal vs vertical; modelled by the numbers 1, 2) which is equal to 1 with probability $p \in (0,1)$ and equal to 2 with probability $1-p$. We define $(d^j_k)^\perp := \{1,2\}\backslash d^j_k$
- For each $D^j_k \in \mathcal{C}(\mathcal{V}_{k-1})$ (which is a rectangle of sides-length ℓ^j_1, ℓ^j_2) we set

$$\mathcal{B}^j_k := \begin{cases} \left\{ \mathbf{x} \in D^j_k : \mathbf{x}\cdot \mathbf{e}^\perp_{d^j_k} \in \left(\mathbf{p}^j_k \cdot \mathbf{e}^\perp_{d^j_k} - \delta\delta^j_k \ell^j_{d^j_k}, \mathbf{p}^j_k \cdot \mathbf{e}^\perp_{d^j_k} + \delta(1-\delta^j_k)\ell^j_{d^j_k} \right) \right\}, \\ \qquad\qquad \text{if } \delta\ell^j_{d^j_k} < \ell^j_{(d^j_k)^\perp}, \\ D^j_k, \qquad\quad \text{if } \delta\ell^j_{d^j_k} \geqslant \ell^j_{(d^j_k)^\perp}, \end{cases}$$

where

$$\delta^j_k := \quad \arg\min \left\{ \left| s - \frac{1}{2} \right| : s \in (0,1) \text{ and both } \mathbf{p}^j_k - s\ell^j_{d^j_k}\mathbf{e}^\perp_{d^j_k}, \mathbf{p}^j_k + (1-s)\ell^j_{d^j_k}\mathbf{e}^\perp_{d^j_k} \in D^j_k \right\}$$

- We set $\mathcal{V}_k := \mathcal{V}_{k-1}\backslash \overline{\bigcup_j \mathcal{B}^j_k}$ and

$$\mathbf{y}_k := \begin{cases} \mathbf{y}_{k-1}, & \text{on } \Omega\backslash \overline{\bigcup_j \mathcal{B}^j_k}, \\ \mathbf{z}^j_k, & \text{on } \mathcal{B}^j_k, \end{cases}$$

where $\mathbf{z}^j_k \in W^{1,\infty}(\mathcal{B}^j_k; \mathbb{R}^2)$ is given by theorem 2.

Let us comment on this algorithm and its dynamics: we begin with a sample Ω which represents our material at the beginning of the nucleation process (e.g. with the sample being in the austenite phase or possibly also being under some prestrain). As illustrated in figure 6, in each iteration step of the algorithm, in each connected component of $\mathcal{V}_k$ we randomly choose

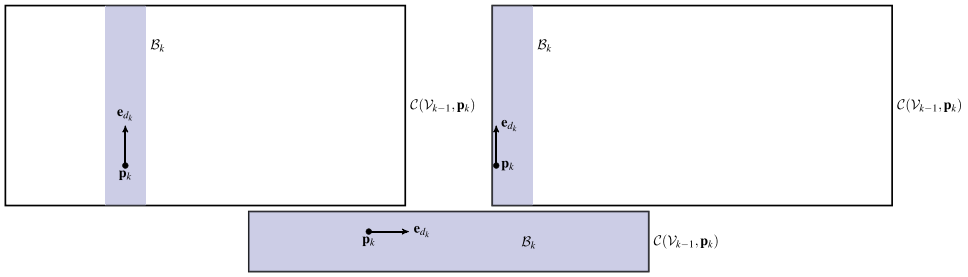


Figure 6. Given a point $\mathbf{p}_k \in \mathcal{C}(\mathcal{V}_k, \mathbf{p}_k)$ and a direction $\mathbf{e}_{d_k}$ we insert a maximal rectangle $\mathcal{B}_k \ni \mathbf{p}_k$ of aspect ratio $1 : \delta$. Here, generically the rectangle is centered around $\mathbf{p}_k$ (left). If $\mathbf{p}_k$ is too close to the boundary, we instead shift the rectangle $\mathcal{B}_k$ to touch the boundary (right). If the domain $\mathcal{C}(\mathcal{V}_k, \mathbf{p}_k)$ is too narrow (pictured on the bottom), we instead pick $\mathcal{B}_k = \mathcal{C}(\mathcal{V}_k, \mathbf{p}_k)$.

a point and an orientation, and consider a set $\mathcal{B}_k^j$ (ideally centered at the chosen point and oriented in the chosen direction, see figure 6) on which we replace the current deformation $\mathbf{y}_k$ by a deformation $\mathbf{z}_k^j$ which itself is given by theorem 2. We iterate this infinitely many times, eventually obtaining a deformation which is increasingly close to being a solution to (5) (and being an exact solution in the limit $k \to \infty$).

We remark that the main idea of the dynamics of the described algorithm is very similar to the ones proposed and analysed in [RSS95, MRSS95, PLKK97, BCH15, CH18, TIVP17]. One main difference here is that instead of just 'declaring' the domains $\mathcal{B}_k^j$ to be filled with martensite, our domains $\mathcal{B}_k^j$ are actually filled with martensite by replacing the deformation $\mathbf{y}_{k-1}$ from the previous step by the new deformation $\mathbf{z}_k^j$ which is obtained by virtue of theorem 2. With respect to the algorithms from [RSS95, MRSS95, PLKK97, BCH15, CH18, TIVP17] by prescribing the precise deformation, our algorithm thus takes care of an *additional layer of complexity* which had been ignored in the previous models.

We remark that there are several natural ways of achieving this. In our algorithm the domains $\mathcal{B}_k^j$ are immediately completely covered by a stress-free martensite configuration. As a consequence, the fully transformed sets $\mathcal{B}_k^j$ will never be modified by the algorithm again (the material is already in the energy wells). As an alternative one could, for instance, have considered an algorithm in which the diamond-shaped rhombi-constructions (see figure 4) are iteratively applied and which thus improve the stress distribution but do not directly yield completely stress-free configurations. In this scenario, one would then try to improve the strain distribution in the sets $\mathcal{B}_k^j$ iteratively again in later steps of the algorithm. Mathematically the latter model would thus correspond to a 'full, random convex integration model', while our algorithm is rather a 'hybrid, random convex integration model', where the convex integration part is taken as a full, black-box building block as a consequence of theorem 2. Due to the additional difficulties in combining the probabilistic perspective and the detailed convex integration estimates, we postpone the study of 'full, random convex integration algorithms' to future work.

In studying the length scale distribution in the sense of understanding the regularity of the final solution $\mathbf{y}$ (in expectation or μ-almost everywhere), we thus need to combine an analysis of the covering algorithm (determined by the generation of the sets $\mathcal{B}_k^j$) which is essentially a probabilistic fragmentation process (and thus related to the problems in for instance [FGRV95, Ber06] and the references therein) with the regularity of the building blocks from theorem 2.

We remark that our constructions in the domains $\mathcal{B}_k^j$ are of a prescribed length-to-width ratio. In the described model, this length scale is introduced in an ad hoc manner. Ideally, it would be

derived as a scale in the interaction of elastic and surface energy contributions as for instance in nucleation results from the literature [KKO13, KK11, CDMZ20]. We further stress that in our definition of the 'nucleation sets' $\mathcal{B}_k^j$ we allow for *degenerate* sets as long as their long axis is oriented perpendicular to the long axis of the sets which are introduced through nucleation. We however exclude degenerate, too long, thin sets, if their long axis is oriented in the same direction as the sets which are inserted in the nucleation step (see the second condition in the definition of the sets $\mathcal{B}_k^j$ which is a non-degeneracy condition). From a technical point of view this allows us to estimate the *gain in volume fraction* in each iteration step without discussing *tail estimates* which originate from increasingly degenerate domains. For these the perimeter would still be controlled, the gain in the volume would however not *a priori* yield exponential gains in the sense of propositions 5.1 and 5.4. From a physical point of view, the degenerate vs non-degenerate choice of the rectangles $\mathcal{B}_k^j$ at this point is ad hoc. However, we believe that in more sophisticated models control on the possible degeneracies can be deduced from surface energy constraints, thus giving some credence to these type of simplifications. As an indication in the direction of being able to derive sufficiently strong tail estimates which allow us to drop the non-degeneracy assumption, in the appendix C we establish such estimates for a slightly modified algorithm. We believe that with some further effort similar results could also hold for the unmodified algorithm (see remark C.7).

3.2.2. Model B. Let us next discuss a second variant of our nucleation mechanism.

Algorithm 3.4 (Model B). Let $\Omega = (0,1)^2$, $\delta \in (0,1)$ and:

- $\mathsf{M} \in K^{\mathrm{qc}}$;
- $\mathbf{y}_0 := \mathsf{M}\mathbf{x}$ in Ω;
- $\mathcal{V}_0 = \Omega$.

Then, for any $k \in \mathbb{N}$

- Let $\mathbf{p}_k$ be a point chosen uniformly at random in $\mathcal{V}_{k-1}$ and we define $\mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k)$ to be the connected component of $\mathcal{V}_{k-1}$ containing $\mathbf{p}_k$ (we remark that $\mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k)$ is always a rectangle of size $\ell_1 \times \ell_2$, with $\ell_1, \ell_2 \in (0,1)$)
- Let $d_k \in \{1,2\}$ be equal to 1 with probability $p \in (0,1)$ and be equal to 2 with probability $1-p$. We define $d_k^\perp := \{1,2\} \backslash d_k$
- We set

$$\mathcal{B}_k := \begin{cases} \left\{ \mathbf{x} \in \mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k) : \mathbf{x} \cdot \mathbf{e}_{d_k}^\perp \in \left(\mathbf{p}_k \cdot \mathbf{e}_{d_k}^\perp - \delta_k \ell_{d_k}, \mathbf{p}_k \cdot \mathbf{e}_{d_k}^\perp + (1-\delta_k)\ell_{d_k} \right) \right\}, \\ \qquad\qquad\qquad\qquad \text{if } \delta \ell_{d_k} < \ell_{d_k^\perp}, \\ \mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k), \qquad\qquad \text{if } \delta \ell_{d_k} \geqslant \ell_{d_k^\perp}, \end{cases}$$

where

$$\delta_k := \arg\min \left\{ \left| s - \frac{1}{2} \right| : s \in (0,1) \right. \\ \left. \text{and both } \mathbf{p}_k - s\ell_{d_k}\mathbf{e}_{d_k}^\perp, \mathbf{p}_k + (1-s)\ell_{d_k}\mathbf{e}_{d_k}^\perp \in \mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k) \right\}$$

- We set $\mathcal{V}_k := \mathcal{V}_{k-1} \backslash \overline{\mathcal{B}}_k$ and

$$\mathbf{y}_k := \begin{cases} \mathbf{y}_{k-1}, & \text{on } \Omega \backslash \mathcal{B}_k, \\ \mathbf{z}_k, & \text{on } \mathcal{B}_k, \end{cases}$$

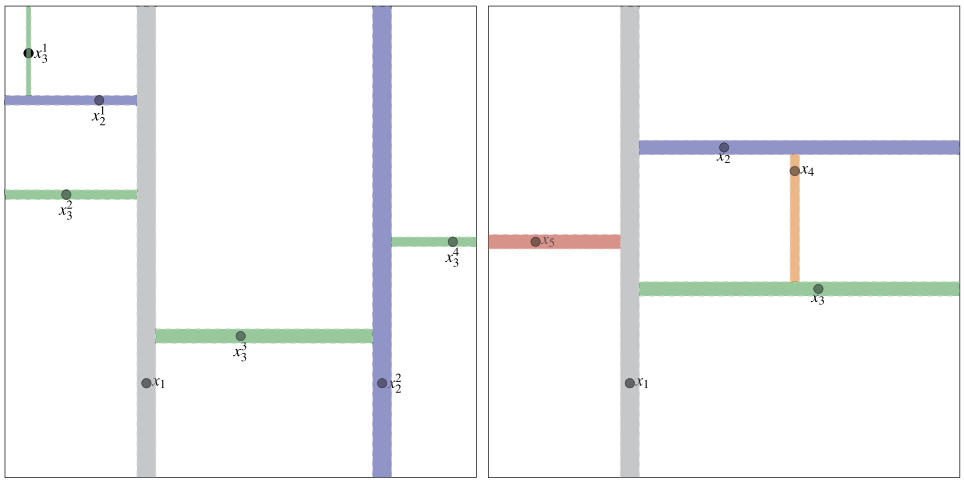


Figure 7. In algorithm 3.4 (pictured on the right) in each step we randomly pick a point in the remaining area according to the normalized Lebesgue measure and insert a maximal rectangle $\mathcal{B}$ containing this point. In algorithm 3.3 (pictured on the left) in each step we independently pick a random point for each connected component. In this schematic illustration of our algorithms the colours of the rectangles correspond to the iteration step k of our algorithm. In particular, we observe that in algorithm B (right) only one set $\mathcal{B}_k$ is introduced in the step k while in the algorithm A (left) we introduce 2^k new sets $\mathcal{B}_k^j$ in the kth iteration step. As a consequence, on average, the microstructure produced in algorithm 3.4 provides a much more uniform covering than the one from algorithm 3.3, see also figures 11 and 12 in section 8.

where $\mathbf{z}_k \in W^{1,\infty}(\mathcal{B}_k;\mathbb{R}^2)$ is given by theorem 2.

In contrast to the algorithm 3.3 this algorithm does not nucleate a new martensitic plate in each connected component of $\mathcal{V}_k$ but considers the more realistic (but mathematically slightly more involved) situation of a single nucleation event in each step. The position of the nucleation here is determined by the volume of the largest undeformed piece in the sample (see figure 7 for an illustration of the differences between the two algorithms).

While it may intuitively be apparent that the limiting deformations produced in algorithms 3.3 and 3.4 agree, the speed of convergence and in particular the behaviour after finitely many iteration steps are very different in the two algorithms. We illustrate this in section 8 in which we implement (numerically more tractable) variants of both algorithms. Thus, in the main text, we treat both algorithms in parallel. In the appendix B we prove that there is indeed a bijection (up to reordering) between the two outlined algorithms which yields 'equivalent' probability distributions.

In the following sections we thus analyse algorithms 3.3 and 3.4, study their convergence properties (in expectation) and the regularity of the resulting deformations.

4. Probability spaces and extensions

In the following we define the probability spaces associated to the algorithms 3.3 and 3.4 for each finite step $k \in \mathbb{N}$ and a common probability space $(X, \mathcal{F}, \mu)$ which includes all finite steps. We thus consider our convex integration algorithms as stochastic processes with k being interpreted as a discrete time step. In section 5 we then study the convergence of the algorithm

in L^p by computing expectations of various norms of (differences of) the sequences obtained in the constructions. In sections 6 and 7 we further study higher regularity of the solutions and in particular show that our algorithms μ-almost surely produce a $W^{1+s,p}$ regular solution of the differential inclusion.

Our probability spaces consider the sequences of points $x_k \in \Omega$ (produced by $\mathbf{p}_k$) and directions $d_k \in \{1,2\}$ chosen in the algorithms. To each such sequence we may then associate a sequence of sets

$$\mathcal{V}_k = \mathcal{V}_k(x_1, d_1, x_2, d_2, \ldots, x_k, d_k)$$

by constructing the rectangles $\mathcal{B}_k(x_1, d_1, \ldots, x_j, d_j)$ as prescribed in the algorithm. We will show that this function is measurable and that $(\mathcal{V}_k)_{k\in\mathbb{N}}$ can therefore be considered a random variable.

4.1. The probabilistic set-up for algorithm 3.4

For simplicity of notation in the following we first discuss algorithm 3.4 where k steps correspond to choosing k points $(x_1, \ldots, x_k) \in \Omega^k$ and directions $(d_1, \ldots, d_k) \in \{1,2\}^k$.

Lemma 4.1. *Consider the sequences of points $(x_1, x_2, \ldots)$ and directions $(d_1, d_2, \ldots)$ generated by algorithm 3.4 as a stochastic process. Then the corresponding (pullback probability) measure can be expressed as a density. More precisely, for each $k \in \mathbb{N}$ there exists a probability density*

$$\rho_k : (\Omega \times \{1,2\})^k \to [0,\infty),$$

such that for every Borel set $B \in \mathcal{B}(\Omega^k)$ and every $(d'_1, \ldots, d'_k) \in \{1,2\}^k$ the probability that algorithm 3.4 produces a sequence with $(x_1, \ldots, x_k) \in B$ and $(d_1, \ldots, d_k) = (d'_1, \ldots, d'_k)$ is given by

$$\mu_k(B \times \{d'_1, \ldots, d'_k\}) := \int_B \rho_k(x_1, \ldots, x_k; d'_1, \ldots, d'_k) \mathrm{d}x_1 \ldots \mathrm{d}x_k \prod_{j=1}^{k} P(d'_j),$$

where $P(1) = p$ and $P(2) = 1 - p$. That is, our probability measure can be written as a density with respect to the Lebesgue measure and a series of Bernoulli trials determining the direction of the rectangles. Furthermore, it holds that for any $k > 1$

$$\begin{aligned} p \int_\Omega \mathrm{d}x_k\, \rho_k(x_1, \ldots, x_k; d_1, \ldots, d_{k-1}, 1) + (1-p) \int_\Omega \mathrm{d}x_k\, \rho_k(x_1, \ldots, x_k; d_1, \ldots, d_{k-1}, 2) \\ = \rho_{k-1}(x_1, \ldots, x_{k-1}; d_1, \ldots, d_{k-1}), \end{aligned} \tag{7}$$

that is μ_{k-1} is given by the marginal of μ_k.

Proof. The first point x_1 generated in algorithm 3.4 is chosen uniformly at random in Ω (with respect to the Lebesgue measure) and the direction $d_1 \in \{1,2\}$ is chosen independently with probability $(p, 1-p)$. Thus, in this case

$$\mu_1(x_1, d_1) = \begin{cases} \dfrac{p}{|\Omega|}\, \mathrm{d}x_1 & \text{if } d_1 = 1, \\ \dfrac{1-p}{|\Omega|}\, \mathrm{d}x_1 & \text{if } d_1 = 2. \end{cases} \tag{8}$$

Given a point and direction $(x_1, d_1) \in \Omega \times \{1,2\}$, in algorithm 3.4 we obtain a rectangle $\mathcal{B}_1(x_1, d_1)$ and choose x_2 uniformly at random (with respect to the Lebesgue measure) in $\Omega \backslash \mathcal{B}_1(x_1, d_1)$ with probability density:

$$\frac{1}{|\Omega \backslash \mathcal{B}_1(x_1, d_1)|}(1 - 1_{\mathcal{B}_1(x_1,d_1)}(x_2))\mathrm{d}x_2, \tag{9}$$

and choose d_2 independently. Thus, given d_1', d_2', we may compute

$$\mu_2(x_1, x_2; d_1', d_2') = P(d_1')P(d_2')\frac{1}{|\Omega \backslash \mathcal{B}_1(x_1, d_1')|}(1 - 1_{\mathcal{B}_1(x_1,d_1')}(x_2))\mathrm{d}x_1\, \mathrm{d}x_2, \tag{10}$$

where we note that $\mathcal{B}_1(x_1, d_1')$ is prescribed in a measurable way.

More generally, given $(x_1, \ldots, x_k)$, $(d_1', \ldots, d_k')$, and a set $B = B_1 \times \cdots \times B_k$ the conditional probability for the choice of the point x_{k+1} is given by the normalized Lebesgue measure on Ω with k rectangles $(R_i(x_1, \ldots, x_k, d_1', \ldots, d_k'))_{i \in \{1,\ldots,2^k\}}$ removed and, for product sets,

$$\begin{aligned}
&\mu_{k+1}(B \times B_{k+1} \times \left(d_1', \ldots, d_k', d_{k+1}'\right)) \\
&\quad = P(d_{k+1}') \int_{B_{k+1}} \frac{1_{\Omega \backslash \cup R_i(x_1,\ldots,x_k,d_1',\ldots,d_k')}(x_{k+1})}{|\Omega \backslash \cup R_i(x_1, \ldots, x_k, d_1', \ldots, d_k')|}\, \mathrm{d}x_{k+1} \\
&\qquad \times \mu_k(B \times \{d_1', \ldots, d_k'\}).
\end{aligned} \tag{11}$$

In particular, as the conditional probabilities are normalized, the marginal property (7) immediately follows. □

Having constructed probability spaces for each finite k, we now construct an extension $((\Omega \times \{1,2\})^{\mathbb{N}}, \mathcal{F}, \mu)$ which includes all these measures as restrictions. In the case of independent measures this would correspond to identifying the above measures with a premeasure on cylinder sets, constructing the product σ algebra and using Caratheodory's extension theorem. For our case we rely on the following more general extension theorem for discrete time stochastic processes.

Theorem 3 (Theorem 3.3.6 in [CZ01]). *Let $m, n \in \mathbb{N}$, $1 \leqslant m < n$ and define π_{mn} to be the embedding map of the Borel σ algebra $\mathcal{B}^m$ on $\mathbb{R}^m$ into $\mathcal{B}^n$ given by*

$$\forall\ B \in \mathcal{B}^m : \pi_{mn}(B) = \{(x_1, \ldots, x_n) : (x_1, \ldots, x_m) \in B\}.$$

Suppose that for each $n \in \mathbb{N}$, μ_n is a probability measure on $(\mathbb{R}^n, \mathcal{B}^n)$ such that

$$\forall\ m < n : \mu_n \circ \pi_{mn} = \mu_m. \tag{12}$$

Then there exists a probability space $(X, \mathcal{F}, \mu)$ and a sequence of random variables X_j such that for each n, μ_n is the n-dimensional probability measure of the vector $(X_1, \ldots, X_n)$.

Following the argument in [CZ01], we may apply this extension theorem to the sequence of probability measures generated by algorithm 3.4, obtaining a probability measure on the space of sequences $(\Omega \times \{1,2\})^{\mathbb{N}}$.

Lemma 4.2. *Let μ_k be the sequence of probability measures on $(\Omega \times \{1,2\})^k$ as in lemma 4.1 with the product Borel σ algebra for each k. Let $X = (\Omega \times \{1,2\})^{\mathbb{N}}$ be the Cartesian product equipped with the product σ algebra. Then there exists a measure μ on X and a sequence of random variables X_j such that μ_k is the probability measure of the vector $(X_1, \ldots, X_k)$.*

Proof. We consider probability measures on $\Omega \times \{1,2\}$, which can be considered as a (two-dimensional) subset of $\mathbb{R}^3$. The marginal property (12) is satisfied by (7) and we hence conclude by applying theorem 3. □

4.2. The probabilistic set-up for algorithm 3.3

It remains to discuss algorithm 3.3. Here, the choice of (x_1, d_1) is identical to algorithm 3.4, but in the kth step we choose not just one point x_k but rather 2^k points $(x_k^i)_{i=1}^{2^k}$, one for each connected component.

Lemma 4.3. *Let μ_k be the sequence of probability measures on $(\Omega \times \{1,2\})^{2^k}$ generated by k steps of algorithm 3.3. Then there exists a measure μ on $(\Omega \times \{1,2\})^{\mathbb{N}}$ and a sequence of random variables X_j such that μ_k is the probability measure of the vector $(X_1, \ldots, X_{2^k})$.*

Proof. We note that given the points obtained in step k, the algorithm picks all these points independently at the same time. In view of the extension of theorem 3 we further construct a sequence of intermediate measures

$$\mu_k^1((x_{k-1}^i), x_k^1), \mu_k^2((x_{k-1}^i), x_k^1, x_k^2), \ldots,$$

where we pick them sequentially from the connected components (since these points are chosen independently we may pick in any order). Each such measure can be written in terms of a density expressing conditional probabilities as in (11), where instead of all of $\Omega \backslash \cup R_i(x_1, \ldots, x_k^i, d_1, \ldots, d_k^i)$, we now consider the (Lebesgue) normalized densities on each connected component.

With this convention the measures considered in algorithm 3.3 correspond to the subsequence $\mu_k := \mu_k^{2^k}$. As the points are chosen independently according to a probability measure (which is normalized), the sequence of measures $\mu_k^1, \ldots \mu_k^{2^k}, \mu_{k+1}^1, \ldots, \mu_{k+1}^{2^{k+1}}, \mu_{k+2}^1, \ldots$ satisfies the marginal property and hence μ can be obtained by applying theorem 3. □

5. Convergence of the algorithms

In this section we study the convergence of the algorithms 3.3 and 3.4 with respect to L^p norms. More precisely we show that the expected value (with respect to the measure μ of section 4) of the Lebesgue measure of the sets $\mathcal{V}_k$ tends to zero as $k \to \infty$. In section 6 we further show that the expected value of the BV norms of the associated characteristic functions does not grow too quickly and that, as a result, the expectations of the $W^{1+s,p}$ norms of the differences $\nabla \mathbf{y}_{k+1} - \nabla \mathbf{y}_k$ form a Cauchy sequence (in $\mathbb{R}$). In section 7 we then pass from statements about expectations to statements about sequences and in particular establish convergence and higher regularity for μ-almost every sequence.

5.1. Convergence of model A

In this section we prove the following result:

Proposition 5.1. *Consider the algorithm 3.3 (model A), let μ be the probability measure constructed in lemma 4.3 and let $\mathbb{E}(\cdot)$ denote the expectation with respect to μ. Then for each $k \geqslant 0$, $\mathcal{V}_k$ is a random variable with respect to μ and it holds that*

$$\mathbb{E}\left(|\mathcal{V}_k|\right) \leqslant \tilde{c}_A^k |\Omega|,$$

where $\tilde{c}_A := \max\{p+(1-p)(1-\delta),(1-p)+p(1-\delta)\} \in (0,1)$ *and* $p \in (0,1)$ *is as in algorithm* 3.3.

We note that $\mathcal{V}_k$ only depends on $((x_1,d_1),(x_2,d_2),\ldots)$ in terms of the points and directions chosen up to step k. Hence, the expectation $\mathbb{E}$ may equivalently be computed in terms of the measures μ_k in which case we work with $\mathbb{E}_k$ (see section 2 for the notation).

Proof. Let $D_k^j \in \mathcal{C}(\mathcal{V}_{k-1})$. We notice that

$$\mathbb{E}_{k+1}\left(|D_k^j \backslash B_k^j| \,\middle|\, D_k^j\right) \leqslant \begin{Bmatrix} p(1-\delta)|D_k^j| + (1-p)|D_k^j|, & \text{if } \ell_1^j > \ell_2^j \\ (1-p)(1-\delta)|D_k^j| + p|D_k^j|, & \text{if } \ell_1^j \leqslant \ell_2^j \end{Bmatrix}$$
$$\leqslant \tilde{c}_A |D_k^j|,$$

since the new rectangle covers a fraction δ of the area if a favourable orientation is chosen by the algorithm. Here, $\mathbb{E}(V|D_k^j)$ corresponds to the conditional expectation with $(x_1,d_1,\ldots,x_k,d_k)$ prescribed (see (11) for the corresponding probability density).

Integrating this estimate with respect to $(x_1,d_1,\ldots,x_k,d_k)$ (and μ_k) we obtain the expected value inequality

$$\mathbb{E}_{k+1}\left(|D_k^j \backslash B_k^j|\right) \leqslant \tilde{c}_A \mathbb{E}_k\left(|D_k^j|\right).$$

Thus, taking the union over j and exploiting the fact that $\mathcal{V}_0 = \Omega$ and that $|\mathcal{V}_k| = \bigcup_j |D_k^j \backslash B_k^j|$ proves the claim. □

In particular, proposition 5.1 implies the following convergence result:

Corollary 5.2. *Algorithm* 3.3 (*model A*) *converges in expectation, i.e.,*

$$\lim_{k\to\infty} \mathbb{E}\left(|\mathcal{V}_k|\right) = 0, \qquad \lim_{(k,l)\to(\infty,\infty)} \mathbb{E}\left(\|\mathbf{y}_k - \mathbf{y}_l\|_{L^\infty(\Omega)}\right) = 0.$$

Remark 5.3. We emphasize that this corollary only constitutes the very first step of our analysis of the generated sequences $\{\mathbf{y}_k\}_{k\in\mathbb{N}}$. In particular, the corollary does not yet ensure the convergence of the sequence $\{\mathbf{y}_k\}_{k\in\mathbb{N}}$ to a solution of the differential inclusion (2).

Proof. The first statement is clear from proposition 5.1. For the second statement, we just notice that (supposing without loss of generality that $k < l$)

$$\mathbb{E}\|\mathbf{y}_k - \mathbf{y}_l\|_{L^\infty(\Omega)} = \mathbb{E}\|\mathbf{y}_k - \mathbf{y}_l\|_{L^\infty(\mathcal{V}_k \backslash \mathcal{V}_l)}$$
$$\leqslant c\mathbb{E}(\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_l\|_{L^\infty(\Omega)} |\mathcal{V}_k \backslash \mathcal{V}_l|^{\frac{1}{2}}) \leqslant c\tilde{c}_A^{\frac{k}{2}} |\Omega|,$$

for some $c > 0$, and where we have used that, since K, K^{qc} are bounded, $\nabla \mathbf{y}_j$ is bounded in $L^\infty(\Omega)$ for each $l \geqslant 0$. The claim thus follows. □

5.2. Convergence of model B

In this section we prove the following result (which does not yet ensure that $\mathbf{y}_k$ converges to a solution of (2), see theorems 4 and 6 for this).

Proposition 5.4. *Consider the algorithm* 3.4 (*model B*), *let* μ *be the probability measure constructed in lemma* 4.2 *and let* $\mathbb{E}(\cdot)$ *denote the expectation with respect to* μ. *Then for each* $k \geqslant 0$, $\mathcal{V}_k$ *is a random variable with respect to* μ *and it holds that*

$$\mathbb{E}\left(|\mathcal{V}_{2k+1}|\right) \leqslant \tilde{c}_B \mathbb{E}\left(|\mathcal{V}_k|\right),$$

where $\tilde{c}_B := \tilde{c}_A + (1-\tilde{c}_A)\frac{1+e^{-\frac{1}{2}}}{2} \in (0,1)$ *and* $\tilde{c}_A$ *is as in proposition* 5.1.

In order to work with a concise notation, we recall the concept of a descendant of a domain:

Definition 5.5 (Definition 3.3 in [RZZ18]). Let $\hat{D} \in \mathcal{C}(\mathcal{V}_k)$ for some $k \geqslant 0$. Then we say that $\check{D} \in \mathcal{C}(\mathcal{V}_l)$ for some $l \geqslant k$ is a descendant of $\hat{D}$ if $\check{D} \subset \hat{D}$. We denote the set of all descendants of $\hat{D}$ by $\mathcal{D}(\hat{D})$.

Proof of proposition 5.4. We know that, in the setting of model algorithm 3.4, $\mathcal{V}_k$ has at most $k+1$ connected components D_k^j. After $k+1$ iterations of the algorithm we thus obtain that

$$\begin{aligned}\mathbb{E}\left(|\mathcal{V}_{2k+1}||\,\mathcal{V}_k\right) &= \sum_j \mathbb{E}\left(\left|\mathcal{D}(D_k^j)\cap\mathcal{V}_{2k+1}\right|\Big|\,\mathcal{V}_k\right)\\ &\leqslant \sum_j \bar{p}_j \mathbb{E}\left(\left|\mathcal{D}(D_k^j)\cap\mathcal{V}_{2k+1}\right|\Big|\,\mathcal{V}_k \quad \text{and} \quad \mathbf{p}_l \in D_k^j \quad \text{for some}\right.\\ &\qquad\qquad \left. l \in \{k+1,\dots,2k+1\}\right)\\ &\quad + \sum_j (1-\bar{p}_j)|D_k^j|,\end{aligned}$$

where $\bar{p}_j$ is the probability that $\mathbf{p}_l \in D_k^j$ for some $l \in \{k+1,\dots,2k+1\}$. This can be computed by noticing that

$$1-\bar{p}_j = \prod_{l=k+1}^{2k+1}\left(1-\frac{|D_k^j|}{|\mathcal{V}_{l-1}|}\right) \leqslant \left(1-\frac{|D_k^j|}{|\mathcal{V}_k|}\right)^{k+1},$$

from which we obtain that

$$\bar{p}_j \geqslant 1-\left(1-\frac{|D_k^j|}{|\mathcal{V}_k|}\right)^{k+1} =: p_j.$$

Therefore, setting $\tilde{c}_A := \min\{p+(1-p)(1-\delta),(1-p)+p(1-\delta)\} \in (0,1)$ and arguing as in the proof of proposition 5.1, we deduce that

$$\mathbb{E}\left(\left|\mathcal{D}(D_k^j)\cap\mathcal{V}_{2k+1}\right|\Big|\,\mathcal{V}_k \quad \text{and} \quad \mathbf{p}_l \in D_k^j \quad \text{for some}\, l \in \{k+1,\dots,2k+1\}\right) \leqslant \tilde{c}_A|D_k^j|,$$

which allows us to estimate

$$\begin{aligned}\mathbb{E}\left(|\mathcal{V}_{2k+1}||\,\mathcal{V}_k\right) &\leqslant \sum_j \left(\bar{p}_j\tilde{c}_A + (1-\bar{p}_j)\right)|D_k^j|\\ &\leqslant \sum_j \left(p_j\tilde{c}_A + (1-p_j)\right)|D_k^j| \leqslant |\mathcal{V}_k| + (\tilde{c}_A-1)\sum_j p_j|D_k^j|\\ &\leqslant \tilde{c}_A|\mathcal{V}_k| + (1-\tilde{c}_A)\sum_j\left(1-\frac{|D_k^j|}{|\mathcal{V}_k|}\right)^{k+1}|D_k^j|.\end{aligned} \tag{13}$$

Let now $r_j := \frac{|D_k^j|}{|\mathcal{V}_k|}$. We now claim that

$$\sum_j \left(1 - r_j\right)^{k+1} r_j \leqslant \hat{c}_B \tag{14}$$

for some $\hat{c}_B \in (0,1)$. Indeed, let

$$J_1 := \left\{ j : r_j \geqslant \frac{1}{2(k+1)} \right\}, \qquad J_2 := \left\{ j : r_j < \frac{1}{2(k+1)} \right\},$$

and note that $J := J_1 \cup J_2$ has $k+1$ elements. We have

$$\sum_{j\in J_1} \left(1 - r_j\right)^{k+1} r_j \leqslant \sum_{j\in J_1} \left(1 - \frac{1}{2(k+1)}\right)^{k+1} r_j \leqslant \mathrm{e}^{-\frac{1}{2}} \sum_{j\in J_1} r_j,$$

and

$$\sum_{j\in J_2} \left(1 - r_j\right)^{k+1} r_j \leqslant \sum_{j\in J_2} r_j.$$

Since $\#J = k+1$,

$$\sum_{j\in J_2} r_j \leqslant \frac{1}{2(k+1)} \sum_{j\in J_2} 1 \leqslant \frac{1}{2(k+1)} \sum_{j\in J} 1 \leqslant \frac{1}{2},$$

and since $\sum_{j\in J_1} r_j = 1 - \sum_{j\in J_2} r_j$, we have

$$\begin{aligned}
\sum_{j\in J} \left(1 - r_j\right)^{k+1} r_j &\leqslant \sum_{j\in J_1} \left(1 - r_j\right)^{k+1} r_j + \sum_{j\in J_2} \left(1 - r_j\right)^{k+1} r_j \\
&\leqslant \mathrm{e}^{-\frac{1}{2}} \sum_{j\in J_1} r_j + \sum_{j\in J_2} r_j \\
&\leqslant \mathrm{e}^{-\frac{1}{2}} + \left(1 - \mathrm{e}^{-\frac{1}{2}}\right) \sum_{j\in J_2} r_j \leqslant \mathrm{e}^{-\frac{1}{2}} + \left(1 - \mathrm{e}^{-\frac{1}{2}}\right) \frac{1}{2},
\end{aligned}$$

which is (14) with $\hat{c}_B := \frac{1+\mathrm{e}^{-\frac{1}{2}}}{2}$. Therefore, combining (13) and (14) we deduce

$$\mathbb{E}\left(|\mathcal{V}_{2k+1}| \,|\, \mathcal{V}_k\right) \leqslant |\mathcal{V}_k|(\tilde{c}_A + (1 - \tilde{c}_A)\hat{c}_B) =: \tilde{c}_B |\mathcal{V}_k|.$$

We remark that $\tilde{c}_B \in (0,1)$ is independent of k. Thus, taking expectation, we deduce

$$\mathbb{E}\left(|\mathcal{V}_{2k+1}|\right) \leqslant \tilde{c}_B \mathbb{E}\left(|\mathcal{V}_k|\right).$$

□

In particular, proposition 5.4 implies the desired convergence result.

Corollary 5.6. *Algorithm 3.4 (model B) converges in expectation. That means*

$$\lim_{k\to\infty} \mathbb{E}\left(|\mathcal{V}_k|\right) = 0, \qquad \lim_{(k,l)\to(\infty,\infty)} \mathbb{E}\left(\|\mathbf{y}_k - \mathbf{y}_l\|_{L^\infty(\Omega)}\right) = 0.$$

Proof. Let $k \geqslant 1$, and $\bar{n}, \bar{k}$ be defined by

$$\bar{n} \coloneqq \sup\{n \in \mathbb{N} : 2^n - 1 \leqslant k\}, \qquad \bar{k} \coloneqq 2^n - 1.$$

Then, by proposition 5.4 we have

$$\mathbb{E}\left(|\mathcal{V}_k|\right) \leqslant \mathbb{E}\left(|\mathcal{V}_{\bar{k}}|\right) \leqslant \tilde{c}_B^{\bar{n}}|\Omega|,$$

which implies the first claim. The second claim follows by arguing as in the case of model A in corollary 5.2. □

6. Regularity of the solutions

After having discussed the convergence in expectation of the algorithms from model A and model B in the previous section, we now study their expected higher regularity properties.

Again we begin by considering model A first and then pass on to model B.

Theorem 4. *There exists $\theta_A \in (0,1)$ such that, for each $(s,p) \in (0,1) \times [1,\infty)$ satisfying $sp < \theta_A$ we have that $\mathbf{y}_k$ constructed as in algorithm 3.3 (model A) satisfies*

$$\mathbb{E}\left(\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)}\right) \leqslant C 2^{-k\alpha}$$

for some $C, \alpha > 0$ depending on θ_A, p and $\mathsf{M} \in K^{\mathrm{qc}}$ only.

Proof. Let us first recall that for any $f \in W^{s,p}(\Omega)$ we have

$$\|f\|_{W^{s,p}(\Omega)}^p = \|f\|_{L^p(\Omega)}^p + \int_\Omega \int_\Omega \frac{|f(\mathbf{x}) - f(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{2+sp}} \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}.$$

Let us start by assuming that $sp < \theta_0$, where θ_0 is as in theorem 2. On the one hand, since $\nabla \mathbf{y}_l(x) \in K$ for $x \in \Omega \backslash \mathcal{V}_l$ and any $l \geqslant 0$ (and as $\nabla \mathbf{y}_k$ will not be changed along the iteration on that set any more) and as both K and K^{qc} are compact, we have

$$\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}\|_{L^p(\Omega)}^p \leqslant c|\mathcal{V}_k \backslash \mathcal{V}_{k+1}| \leqslant c|\mathcal{V}_k|.$$

On the other hand, setting $\mathsf{v}_k \coloneqq \nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}$ and using that $\nabla \mathbf{y}_k = \nabla \mathbf{y}_{k+1}$ on $(\bigcup_j \mathcal{B}_k^j)^c$, we observe that

$$\begin{aligned}\int_\Omega \int_\Omega \frac{|\mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{2+sp}} \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} &\leqslant \sum_j \int_{\mathcal{B}_k^j} \int_{\mathcal{B}_k^j} \frac{|\mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{2+sp}} \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \\ &\quad + 2\sum_j \int_{\mathcal{B}_k^j} \int_{(\mathcal{B}_k^j)^c} \frac{|\mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{2+sp}} \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}.\end{aligned} \tag{15}$$

The first term in (15) can be bounded thanks to theorem 2. Indeed,

$$\sum_j \int_{\mathcal{B}_k^j} \int_{\mathcal{B}_k^j} \frac{|\mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{2+sp}} \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \leqslant \sum_j \left(\|\nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\mathcal{B}_k^j)}^p + \|\nabla \mathbf{y}_k\|_{W^{s,p}(\mathcal{B}_k^j)}^p \right).$$

Now building on the interpolation estimate (see [RZZ19, corollary 3])

$$\|u\|_{W^{s,p}(\Omega)} \leqslant \|u\|_{L^\infty(\Omega)}^{1-\frac{1}{p}} (\|u\|_{L^1(\Omega)}^{1-\tilde{\theta}_0} \|u\|_{BV(\Omega)}^{\tilde{\theta}_0})^{\frac{1}{p}}, \tag{16}$$

with $\tilde{\theta}_0 = sp$, as well as the estimates (see [DPR20, proposition 7.1] and [DPR20, lemma 7.1], where in the latter $|\Omega|$ has to be replaced by $\mathrm{Per}(\Omega)$)

$$\begin{aligned} \|\nabla \mathbf{u}_k\|_{BV(\Omega)} &\leqslant C2^k \,\mathrm{Per}(\Omega), \\ \|\nabla \mathbf{u}_k\|_{L^1(\Omega)} &\leqslant Cc^k |\Omega|, \end{aligned} \tag{17}$$

for some constant $c \in (0,1)$ which is independent of Ω and where $\mathbf{u}_k$ is the deformation from [DPR20], we obtain that for our deformation, by combining (16) and (17), we have for $sp < \theta_0$ (where $\theta_0 > 0$ is the regularity threshold from [DPR20] and theorem 2)

$$\|\nabla \mathbf{y}_k\|_{W^{s,p}(\mathcal{B}_k^j)} + \|\nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\mathcal{B}_k^j)} \leqslant C \,\mathrm{Per}(\mathcal{B}_k^j)^s |\mathcal{B}_k^j|^{\frac{1}{p}-s}. \tag{18}$$

Hence, Cauchy–Schwarz and the fact that in the kth iteration step there are 2^k sets $\mathcal{B}_k^j$ in which $\mathbf{y}_k$ is modified, implies that

$$\begin{aligned} \sum_j \int_{\mathcal{B}_k^j}\int_{\mathcal{B}_k^j} \frac{|\mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} &\leqslant \sum_j \|\nabla \mathbf{y}_k\|^p_{W^{s,p}(\mathcal{B}_k^j)} \leqslant \mathrm{Per}(\Omega)^{sp} \sum_{j=1}^{2^k} |\mathcal{B}_k^j|^{1-sp} \\ &\leqslant \mathrm{Per}(\Omega)^{sp} 2^{ksp} |\mathcal{V}_k \setminus \mathcal{V}_{k+1}|^{1-sp}. \end{aligned}$$

Regarding the second term in (15), exploiting the boundedness of the v_k, we have for each j that

$$\begin{aligned} &\int_{\mathcal{B}_k^j}\int_{(\mathcal{B}_k^j)^c} \frac{|\mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x} \leqslant c\int_{\mathcal{B}_k^j}\int_{(\mathcal{B}_k^j)^c} \frac{1}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x} \\ &\leqslant c\int_{\mathcal{B}_k^j}\int_{\left(B\left(\mathbf{x},\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)\right)\right)^c} \frac{1}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x} \leqslant c\int_{\mathcal{B}_k^j}\int_{\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)}^{\infty} \frac{1}{r^{1+sp}}\,\mathrm{d}r\,\mathrm{d}\mathbf{x} \\ &\leqslant c\,\mathrm{Per}(\mathcal{B}_k^j)\min\{\ell_1^j,\ell_2^j\}^{1-sp} \leqslant c\,\mathrm{Per}(\mathcal{B}_k^j)^{sp}|\mathcal{B}_k^j|^{1-sp}. \end{aligned}$$

Here the estimate in the second to last line is a consequence of the following considerations: splitting

$$\begin{aligned} \int_{\mathcal{B}_k^j}\int_{\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)}^{\infty} \frac{1}{r^{1+sp}}\,\mathrm{d}r\,\mathrm{d}\mathbf{x} &= \sum_{i=1}^{2}\int_{\Delta_{j,k}^i}\int_{\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)}^{\infty} \frac{1}{r^{1+sp}}\,\mathrm{d}r\,\mathrm{d}\mathbf{x} \\ &\quad + \sum_{i=1}^{2}\int_{T_{j,k}^i}\int_{\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)}^{\infty} \frac{1}{r^{1+sp}}\,\mathrm{d}r\,\mathrm{d}\mathbf{x}, \end{aligned}$$

where $\Delta_{j,k}^i$ denote the triangles and $T_{j,k}^i$ the trapezoids from figure 8. Further, in the triangle $\Delta_{j,k}^i$ in figure 8 we estimate from above by the estimate over the rectangle $[0,\ell_2^j]\times[0,\ell_2^j]$:

$$\int_{\Delta_{j,k}^i}\int_{\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)}^{\infty} \frac{1}{r^{1+sp}}\,\mathrm{d}r\,\mathrm{d}\mathbf{x} \leqslant \ell_2^j \int_0^{\ell_2^j} x_1^{-sp}\,\mathrm{d}x_1 = \frac{1}{1-sp}\ell_2^j(\ell_2^j)^{1-sp},$$

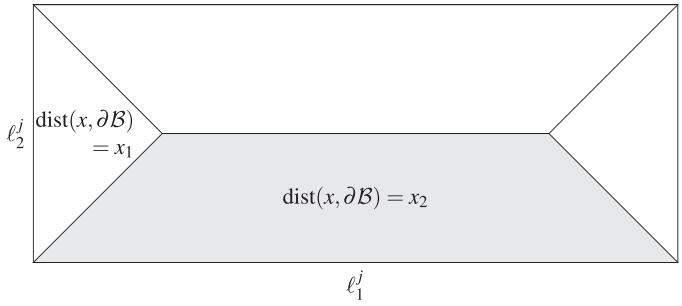


Figure 8. In order to estimate the $W^{s,p}$ seminorm over $\mathcal{B}_k^j \times \mathcal{B}_k^j$ we split the rectangle into the depicted four regions. In each of these regions the distance to the boundary is explicitly given in terms of the Cartesian coordinates.

where we used that the integral with respect to r is given by a constant times dist^{-sp}. Similarly, for the trapezoids on the bottom in figure 8 we similarly estimate by

$$\int_{T_{j,k}^i}\int_{\mathrm{dist}(\mathbf{x},\partial\mathcal{B}_k^j)}^{\infty}\frac{1}{r^{1+sp}}\,\mathrm{d}r\,\mathrm{d}\mathbf{x} \leqslant \ell_1^j\int_0^{\ell_2^j} x_2^{-sp}\,\mathrm{d}x_2 = \frac{1}{1-sp}\ell_1^j(\ell_2^j)^{1-sp}.$$

Thus, the integral is controlled by

$$\max(\ell_1^j,\ell_2^j)\min(\ell_1^j,\ell_2^j)^{1-sp} \leqslant \mathrm{Per}(\mathcal{B}_k^j)\min(\ell_1^j,\ell_2^j)^{1-sp}.$$

Thus, by Hölder's inequality and (18) we infer

$$\sum_j\int_{\mathcal{B}_k^j}\int_{(\mathcal{B}_k^j)^c}\frac{|\mathsf{v}_k(\mathbf{x})-\mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \leqslant c\sum_j \mathrm{Per}(\mathcal{B}_k^j)^{sp}|\mathcal{B}_k^j|^{1-sp}$$
$$\leqslant c\left(\sum_j \mathrm{Per}(\mathcal{B}_k^j)\right)^{sp}\left(\sum_j|\mathcal{B}_k^j|\right)^{1-sp} \leqslant c\big(2^k\big)^{sp}|\mathcal{V}_k|^{1-sp}.$$

By collecting all the above estimates we obtain for $c = c(\mathrm{Per}(\Omega)) > 0$

$$\|\nabla\mathbf{y}_k - \nabla\mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)}^p \leqslant c2^{ksp}|\mathcal{V}_k\backslash\mathcal{V}_{k+1}|^{1-sp} + c2^{ksp}|\mathcal{V}_k|^{1-sp}, \tag{19}$$

which by taking the expected value and by proposition 5.1 becomes

$$\mathbb{E}\left(\|\nabla\mathbf{y}_k - \nabla\mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)}^p\right) \leqslant c\big(2^k\big)^{sp}\big(\tilde{c}_A^k|\Omega|\big)^{1-sp}.$$

Therefore, choosing $\theta_A := sp \in (0,\theta_0)$ such that $2^{\theta_A}\cdot\tilde{c}_A^{1-\theta_A} < 1$, we deduce the existence of $\alpha > 0$ satisfying

$$\mathbb{E}\left(\|\nabla\mathbf{y}_k - \nabla\mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)}^p\right) \leqslant c2^{-\alpha k}.$$

□

Arguing similarly as for model A, regarding model B we have:

Theorem 5. *There exists $\theta_B \in (0,1)$ such that, for each $(s,p) \in (0,1) \times [1,\infty)$ satisfying $sp < \theta_B$ we have that $\mathbf{y}_k$ constructed as in algorithm 3.4 (model B) satisfies*

$$\mathbb{E}\left(\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{2k+1}\|_{W^{s,p}(\Omega)}\right) \leqslant C 2^{-\alpha \log_2(k+1)}$$

for any $k \geqslant 1$, and for some $C, \alpha > 0$ depending on θ_A, p and $\mathsf{M} \in K^{\mathrm{qc}}$ only.

We remark that for $k = 2^l - 1$ it holds that $2k+1 = 2^{l+1} - 1$ and $2^{-\alpha \log_2(k+1)} = 2^{-\alpha l}$. Hence, we may consider the subsequence $\tilde{y}_l := y_{2^l - 1}$ to obtain an estimate of the same form as in theorem 4.

Proof. The proof follows the approach devised in the proof of theorem 4. Again, we start by assuming that $sp < \theta_0$, where θ_0 is as in theorem 2. Then, by arguing as in the proof of theorem 4 we deduce

$$\begin{aligned}
\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{2k+1}\|^p_{W^{s,p}(\Omega)} &\leqslant c \operatorname{Per}(\Omega)^{sp} (k+1)^{sp} |\mathcal{V}_k \setminus \mathcal{V}_{2k+1}|^{1-sp} \\
&\quad + c \left(\sum_{j=k+1}^{2k+1} \operatorname{per}(\mathcal{B}_j) \right)^{sp} |\mathcal{V}_k|^{1-sp} \\
&\leqslant c \operatorname{per}(\Omega)^{sp} (k+1)^{sp} |\mathcal{V}_k \setminus \mathcal{V}_{2k+1}|^{1-sp} \\
&\quad + c (k+1)^{sp} |\mathcal{V}_k|^{1-sp},
\end{aligned}$$

where, in order to infer the estimate for the $W^{s,p}$ semi-norm, we bound the difference of $\nabla \mathbf{y}_k - \nabla \mathbf{y}_{2k+1}$ in the different $\mathcal{B}_j$, with $j \in \{k+1, \ldots, 2k+1\}$, rather than in the sets $\mathcal{B}^j_k$. Thus, taking the expected value and using the estimate from proposition 5.4, we arrive at

$$\begin{aligned}
\mathbb{E}\left(\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{2k+1}\|^p_{W^{s,p}(\Omega)}\right) &\leqslant c \Big((\tilde{c}_B^{\bar{n}} |\Omega|)^{1-sp} (k+1)^{sp} \\
&\quad + (k+1)^{sp} \left(\tilde{c}_B^{\bar{n}} |\Omega|\right)^{1-sp} \Big) \\
&\leqslant c \left(2^{\theta_B} \cdot \tilde{c}_B^{\frac{\bar{n}}{n}(1-\theta_B)} \right)^n,
\end{aligned}$$

where $n := \log_2(k+1)$, $\bar{n} := \lfloor n \rfloor$ and $\theta_B = sp$. Since $\frac{\bar{n}}{n} \geqslant \frac{1}{2}$, whenever $k \geqslant 1$, by choosing $\theta_B := sp \in (0, \theta_0)$ such that $2^{\theta_B} \cdot \tilde{c}_B^{\frac{1}{2}(1-\theta_B)} < 1$, we thus infer the claimed result. □

7. Almost sure convergence and higher regularity

In sections 4–6 we have established estimates on solutions, their regularity and their convergence in expectation. As a consequence of these results we further obtain convergence along sequences for μ-almost every sequence produced by the algorithms.

The following theorem converts the results on expectations of theorem 4 into a statement on μ-almost every sequence.

Theorem 6. *Consider algorithm 3.4 and let μ be as in lemma 4.2 (or algorithm 3.3 and μ as in lemma 4.3). Then there exists $\alpha' > 0$ such that for μ-almost every sequence $(x_1, d_1, \ldots)$*

obtained in the algorithm, there exists $K < \infty$ (*depending on the sequence*) *such that for all* $k \geqslant K$ *it holds that*

$$\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)} \leqslant \tilde{C} 2^{-\alpha' k}, \tag{20}$$

for some constant $\tilde{C}$ *independent of* k (*in the case of algorithm 3.3 we estimate* $\nabla \mathbf{y}_{2^k-1} - \nabla \mathbf{y}_{2^{k+1}-1}$ *instead*). *In particular,* μ*-almost every generated sequence* $\{\mathbf{y}_k\}_{k\in\mathbb{N}}$ *is Cauchy in* $W^{1+s,p}(\Omega)$ *and has a limit* $\mathbf{y} \in W^{1+s,p}(\Omega)$. *The function* $\mathbf{y}$ *satisfies the differential inclusion problem* (5).

Proof of theorem 6. In theorem 4 we had shown that there exists $\alpha > 0$ such that

$$\mathbb{E}(\|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)}) \leqslant \tilde{C} 2^{-\alpha k}.$$

Let now $0 < \alpha' < \alpha$ and define $C > 1$ such that $2^{-\alpha' k} = C^k 2^{-\alpha k}$. Then by Chebychev's inequality it holds that

$$\mu(\{(x_1, d_1, \dots) : \|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)} \geqslant 2^{-\alpha' k}\}) \leqslant \tilde{C} C^{-k}.$$

In particular, given $K \in \mathbb{N}$ we may define the exceptional sets

$$U_K = \bigcup_{k \geqslant K} \{(x_1, d_1, \dots) : \|\nabla \mathbf{y}_k - \nabla \mathbf{y}_{k+1}\|_{W^{s,p}(\Omega)} \geqslant 2^{-\alpha' k}\}$$

and by subadditivity of the measure and the geometric series we obtain that

$$\mu(U_K) \leqslant \tilde{C} \sum_{k \geqslant K} C^{-k} = \tilde{C} C^{-K} \frac{1}{1 - C^{-1}}.$$

Thus for every $\epsilon > 0$ we may find $K \in \mathbb{N}$ sufficiently large such that $\mu(U_K) \leqslant \epsilon$ and therefore

$$W_\epsilon := (\Omega \times \{1, 2\})^{\mathbb{N}} \backslash U_K$$

has measure at least $1 - \epsilon$ and by construction of W_ϵ the estimate (20) holds for each sequence in W_ϵ for all $k \geqslant K$.

Let now ϵ_j be some sequence with $\epsilon_j \to 0$ and define

$$W^\star = \bigcup_j W_{\epsilon_j}.$$

This set has full measure since

$$\mu((\Omega \times \{1, 2\})^{\mathbb{N}} \backslash W^\star) \leqslant \inf_j \mu((\Omega \times \{1, 2\})^{\mathbb{N}} \backslash W_{\epsilon_j}) \leqslant \inf_j \epsilon_j = 0,$$

and therefore any sequence μ-almost surely is in $W^\star$. By construction, for any $(x_1, d_1, \dots) \in W^\star$ there exists ϵ_j and hence K such that (20) is valid for $k \geqslant K$.

As a consequence, $\nabla \mathbf{y}_k \to \nabla \mathbf{y}$ in $W^{s,p}(\Omega)$ for μ-almost every sequence constructed in algorithm 3.3 and $\mathbf{y} \in W^{1+s,p}(\Omega) \cap W^{1,\infty}(\Omega)$ μ-almost surely.

In order to observe that $\nabla \mathbf{y} \in K$ for μ-almost every sequence $(x_1, d_1, \dots)$, we argue analogously: by proposition 5.1 we have that $\mathbb{E}(|\mathcal{V}_k|) \leqslant c^k$ for some $c \in (0, 1)$. Then, as above, Chebychev's inequality again yields that for some $d > 1$

$$\mu(\{(x_1, d_1, x_2, d_2, \dots) : \ |\mathcal{V}_k| \geqslant d^k c^k\}) \leqslant d^{-k}.$$

Given $K \in \mathbb{N}$, we again define exceptional sets

$$\tilde{U}_K = \bigcup_{k \geqslant K} \{(x_1, d_1, \dots) : \ |\mathcal{V}_k| \geqslant (dc)^k\},$$

and obtain that

$$\mu(\tilde{U}_K) \leqslant d^{-k} \frac{1}{1 - d^{-1}}.$$

Defining sets $\tilde{W}_\epsilon$ and $\tilde{W}^\star$ as above, and noticing that the union of two null sets is again a null set, concludes the proof in the case of model A.

The result for algorithm 3.4 follows analogously from theorem 5. □

Proof of theorem 1. The proof of theorem 1 is an immediate consequence of theorem 6. □

8. Simulations

In this final section, we discuss the numerical implementation of our models from algorithms 3.3 and 3.4 and compare it to the results in [BCH15] and [TIVP17] on the one hand and to our simulations from [RTZ18] on the other hand.

8.1. Strategy

In order to run the simulations, we perform the two following simplifications to our models, which significantly reduce the computational cost of the algorithms.

Change 1. For model A we change the definition of the sets $\mathcal{B}_k^j$ whenever $\delta \ell^j_{d_k^j} \geqslant \ell^j_{(d_k^j)^\perp}$ (this is the degenerate case). In this case we define $\alpha_k^j := \frac{\ell^j_{(d_k^j)^\perp}}{\ell^j_{d_k^j}}$, $N_k^j = \left\lfloor \frac{\delta}{\alpha_k^j} \right\rfloor$ and

$$\mathcal{B}_k^j := \left\{ \mathbf{x} \in D_k^j : \mathbf{x} \cdot \mathbf{e}_{d_k^j} \in \mathbf{p}_k^j \cdot \mathbf{e}_{d_k^j} - \delta_k^j N_k^j \ell^j_{(d_k^j)^\perp}, \mathbf{p}_k^j \cdot \mathbf{e}_{d_k^j} + (1 - \delta_k^j) N_k^j \ell^j_{(d_k^j)^\perp} \right\},$$

where

$$\delta_k^j := \arg\min \left\{ \left| s - \frac{1}{2} \right| : s \in (0,1) \text{ and both } \mathbf{p}_k^j - s N_k^j \ell^j_{(d_k^j)^\perp} \mathbf{e}_{d_k^j}, \mathbf{p}_k^j + (1-s) N_k^j \ell^j_{(d_k^j)^\perp} \mathbf{e}_{d_k^j} \in D_k^j \right\}.$$

See figure 9 for an illustration. Similarly, for model B we set $\alpha_k := \frac{\ell_{(d_k)^\perp}}{\ell_{d_k}}$, $N_k = \left\lfloor \frac{\delta}{\alpha_k} \right\rfloor$ and

$$\mathcal{B}_k := \left\{ \mathbf{x} \in \mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k) : \mathbf{x} \cdot \mathbf{e}_{d_k} \in \mathbf{p}_k \cdot \mathbf{e}_{d_k} - \delta_k N_k \ell_{(d_k)^\perp}, \mathbf{p}_k \cdot \mathbf{e}_{d_k} + (1 - \delta_k) N_k \ell_{(d_k)^\perp} \right\},$$

Figure 9. If $\delta l^j_{d^j_k} \geqslant l^j_{(d^j_k)^\perp}$, that is, in the degenerate case, in the algorithms 3.3 and 3.4 we prescribed that we cover all of D^j_k. In order to simplify the numerical implementation we modify our construction to instead insert a maximal number of copies of the building blocks (in this example $N^j_k = 3$).

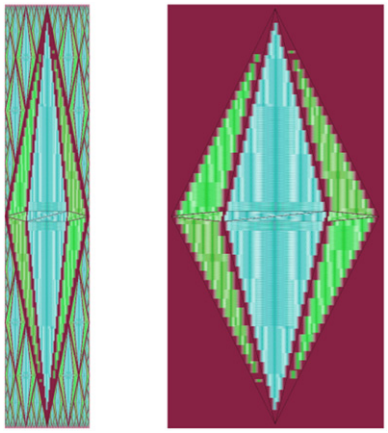


Figure 10. The vertical building blocks: the thin needle structures in their actually used size (left) and a blown-up version of this (right). The illustration shows the building block after several iterations of the convex integration scheme. We have used the parameters $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$.

whenever $\delta \ell_{d_k} \geqslant \ell_{(d_k)^\perp}$. Again, here

$$\delta_k := \arg\min \left\{ \left| s - \frac{1}{2} \right| : s \in (0,1) \text{ and both } \mathbf{p}_k - sN_k \ell_{(d_k)^\perp} \mathbf{e}_{d_k}, \mathbf{p}_k + (1-s)N_k \ell_{(d_k)^\perp} \mathbf{e}_{d_k} \in \mathcal{C}(\mathcal{V}_{k-1}, \mathbf{p}_k) \right\}.$$

We remark that, in both cases, one can repeat exactly the same proofs as in section 5, where the only difference here is that

$$\tilde{c}_A := \max\left\{ p + (1-p)\max\{2^{-1}, (1-\delta)\}, (1-p) + p\max\{2^{-1}, (1-\delta)\} \right\} \in (0,1).$$

That means, this change deteriorates the rate of convergence in the case that $\delta > \frac{1}{2}$.

Change 2. We define

$$\Omega_1 = (0,1) \times (0,\delta), \qquad \Omega_2 = (0,\delta) \times (0,1),$$

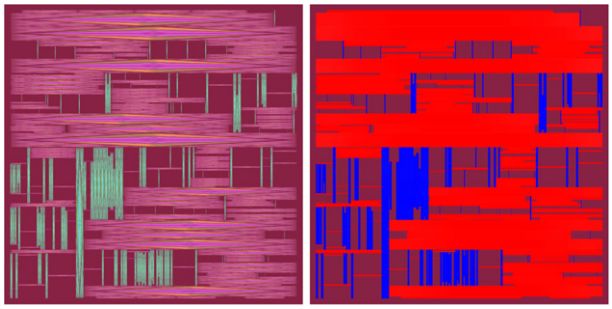


Figure 11. The random convex integration solution produced by algorithm 3.3 (left) and a random packing without interior structure generated by the same random covering arguments as in algorithm 3.3 (right) for $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$. By using the building blocks from theorem 2, the structures which are obtained in the limit $k \to \infty$ of the algorithm 3.3 become exactly stress-free solutions to the differential inclusion (5). The illustration shows the microstructure after 11 iterations of the covering procedure. Thus, we have inserted roughly 2000 building block structures according to the iteration rules of algorithm 3.3.

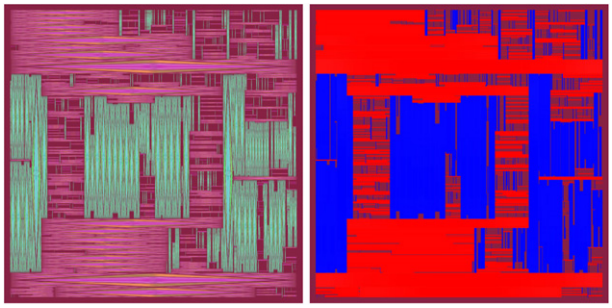
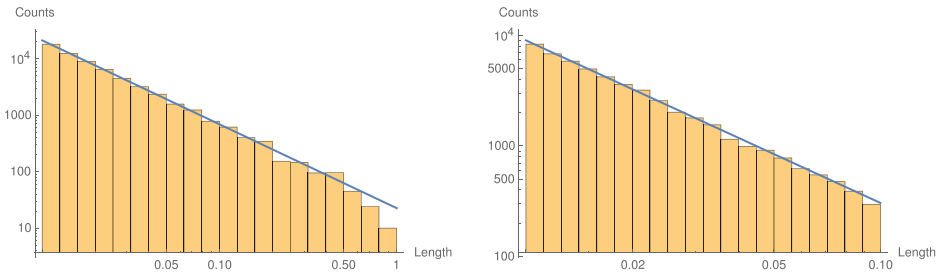


Figure 12. The random convex integration solution for the boundary data $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$ produced by algorithm 3.4 (left) and a random packing without interior structure generated by the same random covering arguments as in algorithm 3.4 (right). As in the setting of the algorithm 3.3 the fact that we rely on building blocks with convex integration structure implies that in the limit $k \to \infty$ the deformations are exactly stress-free solutions to the differential inclusion (5). In the illustration here we have iterated the algorithm roughly 1000 times and have thus introduced roughly 1000 covering rectangles. Due to the iteration scheme of algorithm 3.4 the covering boxes are distributed much more uniformly than in algorithm 3.3 and, on average, cover a larger volume fraction of the domain after the same number of boxes have been introduced.

and we construct, following [DPR20], two solutions $\mathbf{z}_1, \mathbf{z}_2 \in W^{1,\infty}(\Omega_i; \mathbb{R}^2) \cap W^{1+s,p}(\Omega_i; \mathbb{R}^2)$ to

$$\nabla \mathbf{u} \in K \text{ a.e. in } \Omega_i,$$
$$\mathbf{u} = \mathsf{M}\mathbf{x} \text{ on } \partial\Omega_i,$$

for $i = 1, 2$, and where K is given by (3) and $\mathsf{M} \in K^{\mathrm{qc}}$. Here, as in theorem 2, $\theta_0 \in (0, 1)$ and $(s, p) \in [0, 1) \times (1, \infty)$ are arbitrary and such that $sp < \theta_0$. Then, every time we have to make



(A) Histogram of length scales from Algorithm 3.3, with fitting $\frac{22.55}{x^{1.486}}$. An average volume fraction of approximately 0.914 was covered over these realisations.

(B) Histogram of length scales from Algorithm 3.4, with fitting $\frac{10.34}{x^{1.470}}$. An average volume fraction of approximately 0.926 was covered over these realisations.

Figure 13. Histograms of length scales of inclusions from the algorithms 3.3 and 3.4 with parameters $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$. The particular implementation is exact, in the sense that the algorithm terminates only when it is impossible to generate an inclusion of length greater than 10^{-2}, which is the range shown. In both cases, the histograms are generated over 10 realisations. Fitting curves are found by least squares regression of the histograms, using only the data range of 10^{-1} to 10^{-2} to avoid noise during the burn-in of the algorithm.

a replacement construction in the rectangle $\mathcal{B}$ (that is in one of the rectangles $\mathcal{B}_k^j$ in model A, or in one of the rectangles $\mathcal{B}_k$ for model B) we argue as follow:

(a) If $\mathcal{B} = \mathbf{c}_0 + \lambda\Omega_i$ for some $\mathbf{c}_0 \in \mathbb{R}^2$, $\lambda \in (0, 1]$, then we set $\mathbf{y}_k(\mathbf{x}) = \lambda \mathbf{z}_i\left(\frac{\mathbf{x}-\mathbf{c}_0}{\lambda}\right)$ in $\mathcal{B}$. Indeed, we remark that, in $\mathcal{B}$, $\mathbf{z}_{\mathcal{B}}(\mathbf{x}) := \lambda \mathbf{z}_i\left(\frac{\mathbf{x}-\mathbf{c}_0}{\lambda}\right)$ satisfies $\nabla \mathbf{z}_{\mathcal{B}} \in K$ a.e., $\mathbf{z}_{\mathcal{B}}(\mathbf{x}) = \mathsf{M}\mathbf{x}$ on $\partial\Omega$. Therefore, the convergence of the model to the desired limiting stress free deformation is not affected by this change.

(b) If $\mathcal{B} \neq \mathbf{c}_0 + \lambda\Omega_i$ for any $\mathbf{c}_0 \in \mathbb{R}^2, \lambda \in (0, 1]$, according to the changes to the model in the above paragraph, we have the existence of $\mathbf{c}_0 \in \mathbb{R}^2, \lambda \in (0, 1], N \in \mathbb{N}$ such that $\mathcal{B} = \bigcup_{n=0}^{N-1} (\mathbf{c}_0 + n\lambda\mathbf{e}_i + \lambda\Omega_i)$. In this case, as in the above one, we set $\mathbf{y}_k(\mathbf{x}) = \lambda \mathbf{z}_i\left(\frac{\mathbf{x}-\mathbf{c}_0-n\lambda\mathbf{e}_i}{\lambda}\right)$ for any $\mathbf{c}_0 + n\lambda\mathbf{e}_i + \lambda\Omega_i$ and $n = 0, \ldots, N-1$. Again the convergence of the algorithm to the desired limiting stress free deformation is not affected by this change.

The outlined modifications of the algorithms thus have the *computational advantage* that in the degenerate case it suffices to have the two 'standard' convex integration solutions $\mathbf{z}_1, \mathbf{z}_2$ which can be inserted into the covering instead of having to produce new convex integration solutions for each aspect ratio. This is a substantial numerical improvement, since the production of the convex integration building blocks is the computationally most expensive part in our simulations. As explained in the appendix D this change does *not* effect changes in our theoretical regularity estimates for the solutions: the presented numerics illustrate our theoretically findings that while probabilistically the limiting configurations are equivalent, the convergence behaviour of the algorithms is very different.

8.2. Output of the simulations and comparison

In this section, we present some of the output of our simulations of the described algorithms 3.3 and 3.4. As explained above, this combines a random covering by horizontal or vertical rectangles of a fixed width-to-length ratio $\delta > 0$ and the filling of these by the convex integration

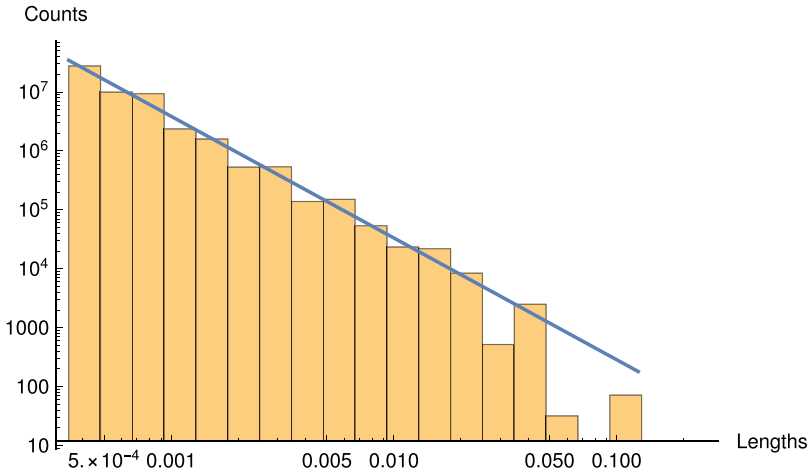


Figure 14. Histogram of length scales for the convex integration algorithm using the parameters $M \approx \mathrm{diag}(0.939, 1.064)$, $\delta = 0.05$, $\gamma = 0.5$. This corresponds to the positive definite square root of C_1 for the parameters λ, $\mu = 0.1$ in the construction of [DPR20]. The histogram is shown with a fit of the form $\frac{2.47}{x^{2.107}}$, obtained by linear regression on the log–log histograms, using only data between 3.5×10^{-4} to 10^{-2} to avoid noise from burn-in. The data is exact on the data shown, in the sense that the implementation only terminates when all possible inclusions of length scale greater than 3.5×10^{-4} have been generated.

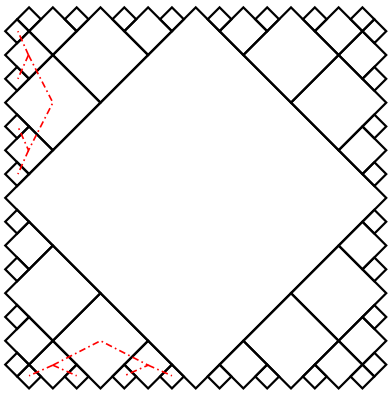


Figure 15. Example of the dyadic packing, showing two lineages.

building blocks from theorem 2. For given boundary data $\mathsf{M} \approx \begin{pmatrix} 0.939 & 0 \\ 0 & 1.064 \end{pmatrix}$, $\delta = 0.05$ and $\gamma = 0.5$, the horizontal and vertical convex integration building blocks are illustrated in figures 5 and 10, respectively.

For the algorithm 3.3 this yields structures as depicted in figure 11 left. Here fractal structures emerge similarly as in [BCH15, CH18] and [TIVP17] with the main difference that the use of convex integration building blocks results in a *stress-free* solution in the limit $k \to \infty$ (which thus involves further fine length scales in the building blocks from theorem 2).

For the algorithm 3.4 we also obtain highly fractal structures (see figure 12). As already observed in [CH18], after inserting the same number of rectangles, these however are more homogeneous than the ones from algorithm 3.4. Compared to the illustrations in [CH18] this is still strongly observable but possibly slightly less pronounced in our illustrations than in [CH18] due to the presence of a finite width.

8.3. Length scale statistics obtained in the algorithms 3.3 and 3.4

In order to eventually compare our results to the experimental data (see for instance [VOM94, CMO98], where universal exponents are obtained for each phase transformation), below we present length scale statistics of our solutions after a finite number of iterations of our algorithms. Here as a measure of the lengths we consider the long side of the rhombi-constructions as a respective measure. The distribution of the lengths involved in the random covering is analogous to the ones obtained in [CH18] or [TIVP17]. Further, finer length scales are however involved in the individual rhombi-building blocks (which themselves are obtained through iterative algorithms, see for instance [DPR20] or [RZZ18], which here are illustrated up to a third order iteration). Due to the use of the 'infinite iteration building blocks' in the rhombi-constructions (see [RTZ18] for this notation), we however remark that while these statistics may eventually serve as a comparison to the experimental data, they are not directly linked to the regularity exponents of the convex integration solutions as in the case of the 'finite convex iteration building blocks' which had been discussed in [RTZ18].

8.4. Combined length scale distributions

A quantity that is of considerable experimental interest is the number and strength of acoustic emissions during the nucleation process [PMV13, VOM94]. It is believed that this is related to the length scale distribution of the microstructure which emerges upon nucleation [CH18]. In order to eventually allow for comparisons of our theoretic findings with experimental results, we also analyse this quantity and present some numerical experiments on its computation.

In our algorithms essentially two length scale distributions enter in the computation of the overall length scale distribution: on the one hand, we consider the lengths scales of the outer random packing (this essentially has the same distribution as the length scale distribution from [TIVP17]). On the other hand these also have an internal length scale distribution, since there is structure also within our building blocks. The inner structure in turn is again determined by two lengths scale distributions which consist of a covering of the given domain by model rectangles and the covering of the model rectangles by the rhombi-constructions.

A heuristic computation shows that in general the length scale distribution will be given by a competition of the involved length scales. We next discuss this for the case of two nested scales: let us assume that for $y \in (0, 1)$ the length scale distribution of the outer blocks is described by a function $f(y)$ which counts the number of blocks (of ratio δ) of size y (in this back-of-the-envelope calculation we exclude the degenerate setting). Let us further suppose that each block of the size $(0, 1) \times (0, \delta)$ has an interior length scale distribution modelled by a function $g(x)$ which assigns to every structure of length scale x the number $g(x)$ of such scales (again we assume that there is a certain non-degeneracy of our structures here). Thus, for every fixed outer structure of the size $(0, y) \times (0, \delta y)$ (up to translation), by scaling, the inner structure has a length scale distribution given by $g(\frac{x}{y})$.

As a consequence, the overall number of structures of lengths x are roughly given by $\int_0^1 f(y)g(\frac{x}{y})\mathrm{d}y$. If both distributions f, g are power laws, e.g. $f(y) = y^\alpha$, $g(x) = x^\beta$, we thus infer

that

$$\int_0^1 f(y)g\left(\frac{x}{y}\right)\mathrm{d}y = c(\alpha,\beta)x^{\beta}(1-x^{\alpha+1-\beta}) \sim \max\{x^{\alpha+1}, x^{\beta}\}. \tag{21}$$

Numerically, we observe that, indeed, at least at our finite numerical resolution of the random convex integration scheme, the length scales can be well-approximated by power laws: considering the boundary data

$$\mathsf{M} \approx \begin{pmatrix} 0.939 & 0 \\ 0 & 1.064 \end{pmatrix},$$

$\delta = 0.05$ and $\gamma = 0.5$ we obtain the following length scale distributions:

(a) The histogram of the length scale distribution inside a diamond building block (as in figure 10) is shown in figure 14. A least square fit gives $g(x) \sim Cx^{-2.107}$.
(b) Using this diamond we dyadically fill a rectangle as in figure 15, which yields a length scale distribution $f(y) \sim Cy^{-1}$ (see lemma A.1 below for a more detailed argument).
(c) These rectangles then serve as building blocks in our stochastic packing, which itself exhibits length scale distributions $h_A(z) \sim Cz^{-1.49}$ and $h_B(z) \sim Cz^{-1.48}$ (see figure 13) for the algorithms 3.3 and 3.4, respectively.

The combination of the first two items (a) and (b) provides the overall length scale distribution of the building block constructions from theorem 2.

Inserting the described distributions into (an iterated version of) (21), we obtain an overall power law with exponent $\alpha = -2.107$. It is thus the length scale of the convex integration building blocks which dominate the overall length scales in our model.

Although, in general, it is not yet proven that the functions $f(y)$ and $g(x)$ must be power laws, we conclude from the back-of-the-envelope computation from above, that the details of the interior structure and thus of the *compatibility* requirement has an interesting, measurable impact on the experimentally measured length scale distributions in the described covering algorithms. In particular including compatibility thus provides important new and experimentally measurable information on the models of [BCH15, CH18, TIVP17].

Acknowledgments

Francesco Della Porta and Angkana Rüland would like to thank the MPI MIS where part of this work was carried out. Angkana Rüland gratefully acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project ID 441068247 within the SPP 2256 'Variational Methods for Predicting Complex Phenomena in Engineering Structures and Materials'. She is a member of the Heidelberg STRUCTURES Cluster of Excellence (which is part of Germany's Excellence Strategy, EXC-2181/1-390900948). Jamie M Taylor has been partially supported by the Basque Government through the BERC 2018–2021 program; and by Spanish Ministry of Economy and Competitiveness MINECO through BCAM Severo Ochoa excellence accreditation SEV-2017-0718 and through Project MTM2017-82184-R funded by (AEI/FEDER, UE) and acronym 'DESFLU'. The work was partially written while Christian Zillinger was at BCAM. During that time his research was supported by the ERCEA under the Grant 014 669689-HADE

and also by the Basque Government through the BERC 2014–2017 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323. Christian Zillinger's research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 2573447—CRC 1173.

Appendix A. A covering result

For completeness, we here discuss the covering result used in section 8.4(b).

Lemma A.1. *Consider an axis-parallel rectangle R of lengths $1:\delta$ and its greedy covering by dyadically rescaled copies of our diamond domain (illustrated in figure 15). Then for any $n \geqslant 1$ there are $8\cdot 2^n$ diamonds with length scale $4^{-1}2^{-n+1}$.*

Proof. We note that the largest diamond in the center has length scale 1 and covers $\frac{1}{2}$ of the total volume. In each of the four remaining regions we then insert two copies of diamond rescaled by a factor $\frac{1}{4}$, which corresponds to the case $n=1$. As illustrated in figure 15 starting from each of these 8 diamonds we obtain a tree-like structure of 2^n diamonds rescaled by a factor $\left(\frac{1}{2}\right)^n$, which concludes the proof.

We remark that by the same argument at step N we have covered a total volume

$$\frac{1}{2}+\frac{1}{2}\sum_{n=1}^{N}2^n4^{-n}=1-2^{-N-2},$$

while the size of the boundary grows proportionally to N. □

Appendix B. On the equivalence of the limits produced in algorithms 3.3 and 3.4

In this section we show that algorithm 3.4, where we randomly pick points according to volume, and algorithm 3.3, where pick points in each connected component, converge to the same limit and have 'equivalent' probability distributions after reordering.

However, we stress that while their limit is the same, their convergence properties, speed of convergence, behaviour after finitely many steps and other qualitative properties are very different. Thus, similarly to numerical schemes which both recover a solution, when viewed as a process or approximation scheme these algorithms are very distinct (as illustrated in our numerical implementations from section 8).

Remark B.1 (Necessity of reordering). We remark that while algorithm 3.3 produces the same distribution of rectangles in the limit, it does not recover the ordering of insertions of algorithm 3.4. Indeed, consider a sequence of points (and corresponding rectangles) generated by algorithm 3.4 as pictured in figure 16. Then the point x_1 and its associated rectangle $\mathcal{B}_1$ bisect the domain Ω into two connected components of (in most cases) different volume. Hence the point x_2 is more likely to be placed in the *larger* connected component. In this case, only at a later point, almost surely after sufficiently many steps a point x_n will be inserted in the other connected component (lemma B.3). In contrast to this, in algorithm 3.3 we first insert a point x_1 by the same equidistributed measure as in algorithm 3.4 and then insert the points x_2^1, x_2^2 in the generated connected exponents at the *same* time. We thus *cannot* recover the asymmetry of the ordering due to the different volumes present in algorithm 3.4 in any way. Therefore, it is *not* possible to construct a bijection between the two algorithm without reordering.

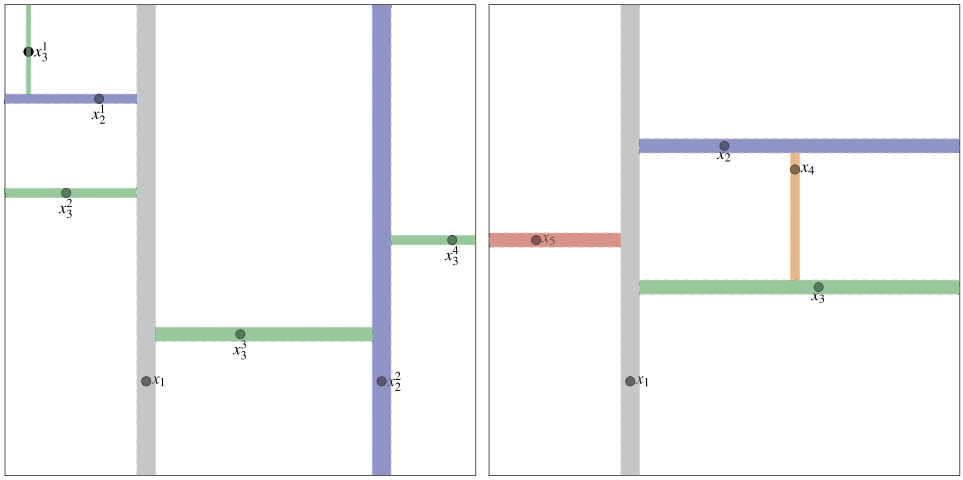


Figure 16. Samples of finite steps of the algorithms. In algorithm 3.3 (left) at each step points are picked each connected component. In algorithm 3.4 (right) we instead randomly pick points equidistributed by volume. Hence, it can happen that many steps are required before a point is inserted in a given connected component.

In order to rigorously state our bijection which is compatible with the probability distributions of the algorithms we will make use of stopping times, construct suitable random variables and establish a correspondence of the probability distributions. This is necessary, since the algorithms 3.3 and 3.4 are defined on different probability spaces. However, before discussing this rigorously let us informally outline the strategy of constructing the bijection. This is illustrated in figures 16–18:

(a). We consider the infinite sequence of points x_j and associated rectangles $\mathcal{B}_j$ generated by algorithm 3.4 as pictured in figure 16. In particular, this keeps track of the order of insertion of the individual rectangles $\mathcal{B}_j$.

(b). To each finite sequence $x_1, \ldots, x_N$ we then associate a tree structure as in figure 17. That is, points inserted in a connected component $\mathcal{C}(\mathcal{V}_j)$ generated by x_j are descendants of x_j.

(c). Given any point x_j and the connected components $\mathcal{C}(\mathcal{V}_j)$ it (newly) generated, we note that after a possibly large number of steps, a point will be picked in this component with positive probability bounded below. Thus, almost surely after finitely many steps a point will have been inserted in each connected component. In terms of the tree this means that almost surely at each level n of the tree all possible nodes will be populated.

(d). In the notation of figure 17 our mapping then associates all points (and associated rectangles) at level n of the tree associated with algorithm 3.4 to the points $x_n^1, x_n^2, \ldots, x_n^{2^n}$ generated by algorithm 3.3.

(e). In both cases, the points coupled through the bijection are picked uniformly according to volume in each connected component and thus obey the same (conditional) probability distribution.

It remains to make this construction rigorous. To this end, we first introduce the following definition of a tree generated by a sequence of points generated by algorithm 3.4. In addition to illustrating the placement of rectangles, here empty nodes keep track of the number of

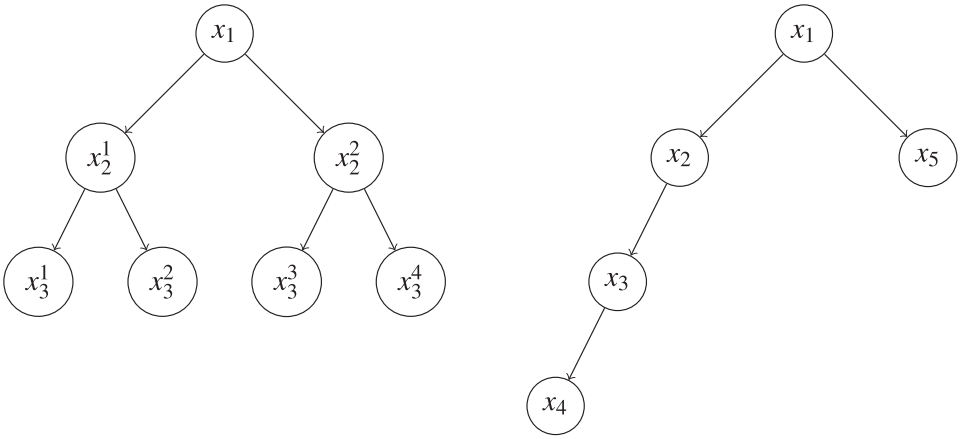


Figure 17. Trees associated to the sequences of figure 16. The root of the tree is given by x_1 and a node x_j is inserted as a child of x_i, $i < j$ if x_j is placed in a connected component newly generated by x_i. Here the edges connecting to empty nodes as in definition B.2 are not pictured.

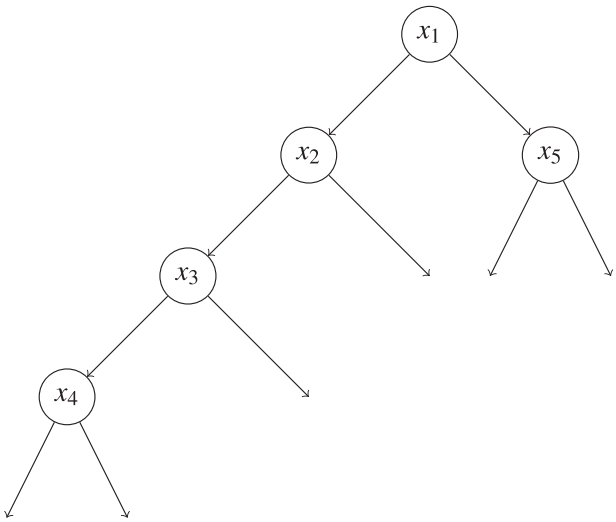


Figure 18. Tree with empty edges. This tree corresponds to the sequences of rectangles given in figure 16 (right) and includes empty edges according to definition B.2. We note that it is *filled* up to level 2 but not further. Thus, this sequence has been stopped before the third stopping time τ_3.

connected components generated by inserting a given rectangle. A version of the tree of figure 17 (right) with these empty nodes is depicted in figure 18.

Definition B.2. Given a finite sequence of points and directions (x_j, d_j) and associated rectangles $\mathcal{B}_j$ we associate a tree to it as follows:

- The root of the tree is given by x_1 (which we identify with its generated rectangle $\mathcal{B}_1$).

- If the insertion of the rectangle $\mathcal{B}_1$ generated two new connected components, we add to the tree two child edges with, at this time, empty nodes. This is the most common case since the rectangle $\mathcal{B}_1$ tends to bisect the region it is inserted in.
- If the insertion of $\mathcal{B}_1$ generated only one connected component (because it touches the boundary of the region it is inserted in with three sides) we only introduce one edge (and one empty node).
- If $\mathcal{B}_1$ fully covers the region it is inserted in, we introduce no edges (and no empty nodes).
- We next insert a rectangle $\mathcal{B}_2$ according to algorithm 3.4 in one of the connected components and place x_2 as a child of the node x_1 (filling up from the left in the graphical representation), thus replacing one of the previously placed empty nodes. Inserting the rectangle $\mathcal{B}_2$ then again generates 2, 1 or 0 new connected components and we insert edges starting at x_2 and empty nodes, accordingly.
- For any later point x_{j+1} and associated rectangle $\mathcal{B}_{j+1}$ we insert it as child of the rectangle which generated the connected component it was placed in, thus replacing a previously empty node. We also add edges and empty nodes according to the number of connected components generated.

While each node may thus have 2, 1 or 0 children, with slight abuse of notation we refer to this construction as a *binary tree*. We say that this tree is *filled at level n* if none of the nodes placed at level n is empty.

The following lemma defines stopping times τ_n such that the process stopped at that time generates a tree which is fully populated until level n. In particular, we show that this almost surely is the case after finitely many steps.

Lemma B.3. *Consider the random sequence of points and directions* $(x_j, d_j)_{j\in\mathbb{N}}$ *and its associated rectangles* $(R_j)_{j\in\mathbb{N}}$ *generated by algorithm* 3.4*. Formally define the stopping times* $\tau_n, n \geqslant 1$ *as the minimal step number N such that the tree generated by the sequence up to step N is filled at all levels less or equal than n. Then* τ_n *are well-defined stopping times and are almost surely finite. Moreover, if we denote rectangles, points and directions picked at level n of the tree by* $R_n^1, \dots R_n^{J(n)} \subset \Omega$, $y_n^1, \dots, y_n^{J(n)} \in \Omega$ *and* $\tilde{d}_n^1, \dots, \tilde{d}_n^J \in \{1, 2\}$*, then these are random variables with respect to the process stopped at the time* τ_n*.*

Proof of lemma B.3. We show by induction that τ_j indeed defines a stopping time and is finite almost surely.

Since the first inserted rectangle generates the root of the tree (which is level 1), $\tau_1 = 1$ is deterministic and thus trivially well-defined and finite.

Thus suppose that τ_j is a stopping time and finite almost surely. Then by assumption, up to null-sets, we may restrict ourselves to considering only sequences on which τ_j is finite. We now consider τ_{j+1} on such a sequence. Then for any finite $n > \tau_j$ we can construct a tree associated to the sequence as in definition B.2 (see also figure 17). In particular, we can easily check whether this tree satisfies our condition that the level $j + 1$ has been filled (all lower levels are filled by the definition of the previous stopping time τ_j). Moreover, this property only depends on the finite sequence. Therefore $\tau_{j+1} \in \mathbb{N} \cup \{\infty\}$ is a well-defined hitting time, which is a standard example of a stopping time.

It thus only remains to show that τ_{j+1} is almost surely finite. For this purpose we consider conditional probabilities. That is, given any finite stopping time $\tau_j = n$ (which are all cases except for a set of probability zero) and a sequence $(x_k, d_k)_{k\leqslant n}$ with this stopping time, we consider the generated connected components $\mathcal{C}(\mathcal{V}_j)$. Let $x_{j_1}, \dots, x_{j_r}$ denote the points at level j of the associated tree. By assumption this level of the tree is fully populated. If also all descendants at level $j + 1$ are fully populated, we have already hit the stopping time τ_{j+1}.

Thus suppose, that this is not yet the case and let $\tilde{\Omega}$ correspond to an empty node at level $j+1$ (i.e. $\tilde{\Omega}$ as not yet been hit by the algorithm. Then the probability that it is chosen in the next step i is given by $\frac{|\tilde{\Omega}|}{|\mathcal{V}_i|}$. Since $\mathcal{V}_i$ is decreasing in $i \in \mathbb{N}$, after N further steps the probability that $\tilde{\Omega}$ is still not hit is bounded above by

$$\left(1 - \frac{|\tilde{\Omega}|}{|\mathcal{V}_i|}\right)^N,$$

which converges to zero as $N \to \infty$. Conversely this means that a point will be picked in $\tilde{\Omega}$ for some finite N almost surely and, since $\tilde{\Omega}$ was arbitrary, for all connected components almost surely at least one point will be picked after finitely many steps. Using the structure of the (finite step) probability measure in terms of the conditional probability measures, it follows that also after integrating with respect to $(x_j, d_j)_{j\leqslant n}$ (with stopping time $\tau_j = n$) the (unconditional) probability that we do not hit the stopping time τ_{j+1} after finitely many steps is zero. Thus τ_{j+1} is finite almost surely, which completes the induction step.

Finally we note that for the stopped processes $(x_j, d_j)_j^{\tau_n}$, the nodes of the tree at level n are exactly the points generating the rectangles $(R_n^i)_i$ which is our desired random variable. □

The stopping times and rectangles of lemma B.3 thus provides a candidate for the desired bijection (up to ordering) between the algorithms 3.3 and 3.4. In the following proposition we show that this mapping indeed obeys the provides a coupling between the algorithms compatible with the probability distributions of both algorithms.

Proposition B.4. *Consider the random variables generated by the stopping times τ_n with $n \in \mathbb{N}$. Then the corresponding distributions of the set of points at level n obey the same law as the points generated by algorithm 3.3. More precisely, let $y_1^1, \ldots, y_1^{J(1)}, \ldots, y_n^1, \ldots, y_n^{J(n)}$ and $d_1^1, \ldots, d_1^{J(1)}, \ldots, d_n^1, \ldots, d_n^{J(n)}$ be the points and orientations from lemma B.3. Then these have the same distribution as the points $x_1, x_2^1, x_2^{J(2)}, \ldots, x_n^1, \ldots, x_n^{J(n)}$ from algorithm 3.3 (in the notation of figure 17).*

Proof. The first point x_1 is picked uniformly at random from the full domain Ω in both algorithms 3.3 and 3.4 and thus this statement is true for $n = 1$.

We next show by induction that the probabilities of the newly picked points at level $n+1$ agree. More precisely, by the induction assumption we know that the points up to level n are picked according to the same law. We will then show that the conditional probabilities of the placement of points at level $n+1$ given the placement at level n agree in law, from which we can the deduce equality of the unconditional laws.

Let thus $x_1, \ldots x_n^i$ be one such choice up to level n and consider the conditional probability of the points picked at level $n+1$. By construction of the stopping time a point will be picked in each connected component generated by $x_1, \ldots x_n^i$. Furthermore, since these are the first points inserted in each component they are equidistributed inside each such component. However, this is exactly the (conditional) law imposed in the definition algorithm 3.3 for placing the points x_{n+1}^j given a prior choice of points $x_1, \ldots x_n^i$. Thus, the conditional probability is indeed given by the same law.

Finally, we note that the full law is given by integrating the conditional probabilities against the distribution of the points up to level n. By the induction assumption the distributions of these points up to level n agree with the distributions generated by algorithm 3.3. Hence by integration also the unconditional distributions agree, which concludes our proof. □

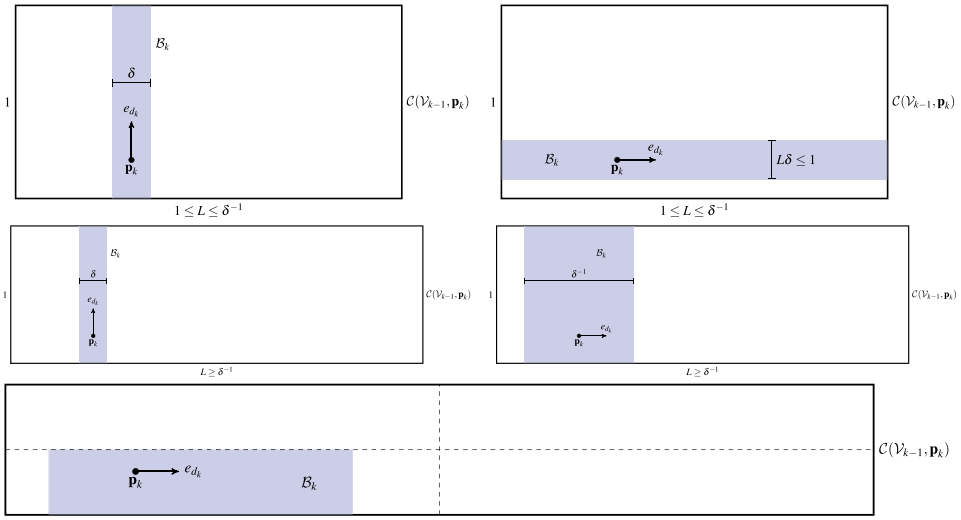


Figure 19. Top: if the rectangle D_k^j has aspect ratio $1 : L$ with $1 \leqslant L \leqslant \delta^{-1}$ the inserted rectangles are translates of $(0, \delta) \times (0, 1)$ and $(0, L) \times (0, L\delta)$. Center: if the aspect ratio is too long, that is $L > \delta^{-1}$, a rectangle $(0, L) \times (0, L\delta)$ is too tall to fit. We thus instead insert a translate of $(0, \delta^{-1}) \times (0, 1)$. Bottom: if the aspect ratio satisfies $\frac{1}{2}\delta^{-1} \leqslant L \leqslant 2\delta^{-1}$ we only modify the quadrant $Q \subset D_k^j$ which contains the picked point $\mathbf{p}_k^j$. More precisely, we use the constructions on the top and center with D_k^j replaced by Q.

Appendix C. On large aspect ratios and tail estimates

As remarked in section 3 in our algorithms 3.3 and 3.4 we opted to completely cover the rectangle D_k^j, if

$$\ell^j_{d_k^j} \geqslant \delta^{-1} \ell^j_{(d_k^j)^\perp},$$

that is, if the length of the rectangle in the direction $\mathbf{e}_{d_k^j}$ we picked is too long. This cut-off simplifies the covering arguments and allows us to more easily deduce uniform bounds on volume fractions which are iteratively covered (as proved for instance in propositions 5.1 and 5.4).

In the following we show that for a slightly modified version of our algorithm such a cut-off is *not* required. These modifications are made precise in algorithm C.1 below and are illustrated in figure 19. All other, not explicitly defined quantities are defined in the same way as in algorithm 3.3.

Algorithm C.1. We consider algorithm 3.3 but make the following two modifications with respect to its dynamics:

(a) For each $D_k^j \in \mathcal{C}(\mathcal{V}_{k-1})$ (which is a rectangle of side lengths ℓ_1^j, ℓ_2^j) we define $\mathcal{B}_k^j$ as follows. If $\ell^j_{d_k^j} < \frac{1}{2}\delta^{-1}\ell^j_{(d_k^j)^\perp}$, we keep the previous definition:

$$\mathcal{B}_k^j := \left\{ \mathbf{x} \in D_k^j : \mathbf{x} \cdot \mathbf{e}^\perp_{d_k^j} \in (\mathbf{p}_k^j \cdot \mathbf{e}^\perp_{d_k^j} - \delta\delta_k^j \ell^j_{d_k^j}, \mathbf{p}_k^j \cdot \mathbf{e}^\perp_{d_k^j} + \delta(1 - \delta_k^j)\ell^j_{d_k^j}) \right\}, \tag{22}$$

where δ_k^j is chosen such that $\mathcal{B}_k^j$ is contained in D_k^j:

$$\delta_k^j := \arg\min\left\{\left|s-\frac{1}{2}\right| : s\in(0,1) \text{ and both } \mathbf{p}_k^j - s\ell^j_{d_k^j}\mathbf{e}^\perp_{d_k^j}, \mathbf{p}_k^j + (1-s)\ell^j_{d_k^j}\mathbf{e}^\perp_{d_k^j} \in D_k^j\right\}.$$

See figure 19 (top) for an illustration.

If $\delta\ell^j_{d_k^j} > 2\ell^j_{(d_k^j)^\perp}$, we cannot insert a translate of $(0,\ell^j_{d_k^j})\times(0,\delta\ell^j_{d_k^j})$, since $(0,\delta\ell^j_{d_k^j}) \not\subset (0,\ell^j_{(d_k^j)^\perp})$. We thus instead insert a translate of $(0,\delta^{-1}\ell^j_{(d_k^j)^\perp})\times(0,\ell^j_{(d_k^j)^\perp})$ according to the following definition:

$$\mathcal{B}_k^j := \left\{\mathbf{x}\in D_k^j : \mathbf{x}\cdot\mathbf{e}_{d_k^j} \in (\mathbf{p}_k^j\cdot\mathbf{e}_{d_k^j} - \lambda_k^j\delta^{-1}\ell_{(d_k^j)^\perp}, \mathbf{p}_k^j\cdot\mathbf{e}_{d_k^j} + (1-\lambda_k^j)\delta^{-1}\ell_{(d_k^j)^\perp})\right\}, \tag{23}$$

where

$$\lambda_k^j := \arg\min\left\{\left|s-\frac{1}{2}\right| : s\in(0,1) \text{ and both } \mathbf{p}_k^j - s\delta^{-1}\ell_{(d_k^j)^\perp}\mathbf{e}_{d_k^j}, \mathbf{p}_k^j + (1-s)\delta^{-1}\ell_{(d_k^j)^\perp}\mathbf{e}_{d_k^j} \in D_k^j\right\}.$$

See figure 19 (center) for an illustration.

(b) Suppose that $D_k^j \in \mathcal{C}(\mathcal{V}_{k-1})$ has lengths ℓ_1,ℓ_2 with $\delta^{-1}/2 \leqslant \frac{\ell_1}{\ell_2} \leqslant 2\delta^{-1}$ and we picked $d_k^j = 1$. Then we divide D_k^j into four quadrants and consider only the quadrant Q which contains the point $\mathbf{p}_k^j$ which had been picked. All other three quadrants remain unchanged and are added to the remainder set $\mathcal{V}^{k+1}$. In the picked quadrant Q, we define $\mathcal{B}_k^j$

1. By the formula (22) (with D_k^j replaced by Q) if $\delta^{-1} \geqslant \frac{\ell_1}{\ell_2}$,
2. And by the formula (23) (with D_k^j replaced by Q) if $\frac{\ell_1}{\ell_2} \geqslant \delta^{-1}$.

See figure 19 (bottom) for an illustration.

For simplicity of presentation in the following we further restrict to the case where the horizontal and vertical directions are chosen with equal probabilities $p = 1-p = \frac{1}{2}$.

Remark C.2. The first point (a) in algorithm C.1 is a relaxation of algorithm 3.3 where we allow for rectangles with large aspect ratio also in the case of alignment with the replacement. The condition (b) is a technical assumption which we do not expect to be necessary. It ensures that we do not cover 'too much' volume, see lemma C.5 and remark C.8.

We show that for the modified algorithm C.1 in expectation only a (uniformly bounded) fraction of the total volume is covered by very long rectangles while most of the volume is covered by non-degenerate rectangles. In order to make this more precise we sort rectangles into *buckets* according to their aspect ratio.

Definition C.3. Let $\mathcal{V}^k$ be a collection of rectangles (defined in algorithm C.1), let $0<\delta<1$ and $1<\lambda$ be given and for simplicity of notation assume that $\delta = \lambda^{-J+1/2}$ for some positive integer $J\in\mathbb{N}$. We then say that a rectangle $R\in\mathcal{V}^k$ is in the *class* C_j, $j\in\{J, J-1,\dots,0,-1,\dots\}\subset\mathbb{Z}$, if its aspect ratios $1:L$, $L\geqslant 1$ satisfies

$$\lambda^{-j-1/2}\delta^{-1} \leqslant L < \lambda^{-j+1/2}\delta^{-1}.$$

We note that C_l with $l > 0$ corresponds to aspect ratios $1 : L$ with $1 \leqslant L < \delta^{-1}\lambda^{-1/2}$ and C_l with $l < 0$ corresponds to 'long' rectangles with aspect ratio $L > \delta^{-1}\lambda^{1/2}$.

Furthermore, we introduce the corresponding volumes

$$V_j^k = \sum_{R\in\mathcal{V}^k : R\in C_j} |R|.$$

We note that the total volume $|\mathcal{V}^k|$ satisfies

$$|\mathcal{V}^k| = \sum_{-\infty\leqslant j\leqslant J} V_j^k = \sum_{R\in\mathcal{V}^k} |R|.$$

Our objective in the following is to show that if $\mathcal{V}^k$ is the random variable given by our algorithm C.1, then there exists $J_1 < 0$ such that for all k it holds that

$$\sum_{j\leqslant J_1} \mathbb{E}(V_j^k) \leqslant 0.1\ \mathbb{E}(|\mathcal{V}^k|). \tag{24}$$

We call this a *tail estimate* since it shows that the contribution of $j \leqslant J_1$ (which corresponds to long, thin rectangles) to the sum of the sequence $(\mathbb{E}(V_j^k))_{j\leqslant J}$ is small.

Supposing for the moment that this estimate holds, we deduce that the expectation of the remaining volume decreases at an exponential rate.

Theorem 7. *Suppose that for some $J_1 < 0$ the random process generated by algorithm C.1 satisfies the estimate (24) for all $k \in \mathbb{N}$. Let $\delta \in (0, 0.1)$ be the threshold from algorithms 3.3 and C.1. Then there exists $c = c(J_1, \delta) \in (0, 1)$ such that for all k*

$$\mathbb{E}(|\mathcal{V}^{k+1}|) \leqslant c\ \mathbb{E}(|\mathcal{V}^k|)$$

and, as a consequence,

$$\mathbb{E}(|\mathcal{V}^k|) \leqslant c^k|\Omega|.$$

We thus obtain similar results as in section 5.1 even without completely covering long rectangles, however possibly with worse rates.

Proof of theorem 7. We recall that in algorithm C.1 we *independently* insert a building block into each rectangle. For any given rectangle R $(= D_k^j \in \mathcal{V}^k)$ we may thus compute the expected volume fraction (of R) covered by the inserted building block and subsequently sum over all rectangles. We note that, by scaling, this volume fraction only depends on the aspect ratio of R which is comparable to $\lambda^{-l}\delta^{-1}$ if $R \in C_l$.

We claim that there exists a sequence of coefficients $c_j \in (0, 1)$ (which is independent of k) such that

$$\mathbb{E}(|\mathcal{V}^k|) - \mathbb{E}(|\mathcal{V}^{k+1}|) \geqslant \sum_{J_1<j\leqslant J} c_j\mathbb{E}(V_j^k). \tag{25}$$

We remark that the quantity on the left equals the total volume covered by building blocks when passing from step k to $k + 1$ (since only building blocks are removed). In the following we will thus have to estimate the expected volume fraction covered by building blocks for any given rectangle $R \in \mathcal{V}^k$. Before proving (25) let us discuss how it allows us to conclude our proof. To this end, we may further estimate the right-hand side of (25) by invoking (24)

$$\min(c_j) \sum_{J_1<j\leqslant J} \mathbb{E}(V_j^k) \geqslant 0.9\ \min(c_j)\mathbb{E}(|\mathcal{V}^k|).$$

Inserting this estimate back into (25) we deduce that

$$\mathbb{E}(|\mathcal{V}^{k+1}|) \leqslant (1 - 0.9\ \min(c_j))\mathbb{E}(|\mathcal{V}^k|),$$

which yields the result of theorem 7.

It hence remains to prove the claimed inequality (25). Let thus R be a given rectangle of lengths ℓ_1, ℓ_2 and for simplicity of notation denote $L = \frac{\max(\ell_1,\ell_2)}{\min(\ell_1,\ell_2)} \geqslant 1$. Then after rescaling, translation and possibly rotating by $\frac{\pi}{2}$ we may assume that

$$R = (0, L) \times (0, 1).$$

Let $j \in \{\dots, -1, 0, 1, \dots, J\}$ such that $R \in C_j$ and hence

$$\lambda^{-j-1/2}\delta^{-1} \leqslant L \leqslant \lambda^{-j+1/2}\delta^{-1}. \tag{26}$$

We then estimate the expected volume covered by the inserted building block as follows:

If the direction $\mathbf{e}_{d_k^j}$ picked is vertical, $\mathbf{e}_{d_k^j} = \mathbf{e}_2$, and $L \notin (\frac{1}{2}\delta^{-1}, 2\delta^{-1})$ then we insert a translate of $(0, \delta) \times (0, 1)$ into a rectangle $(0, L) \times (0, 1)$ (see figure 19 on the left side of the top and center rows). If instead $L \in (\frac{1}{2}\delta^{-1}, 2\delta^{-1})$, then by point (b) of algorithm C.1 we only modify a quadrant and hence insert a translate of $(0, \delta/2) \times (0, 1/2)$ (see figure 19 bottom). In both cases we cover at most a volume fraction

$$\frac{\delta}{L} \leqslant \lambda^{j-1/2}\delta^2 \tag{27}$$

and recall that we picked the direction $e_{d_k^j} = e_2$ with a probability $\frac{1}{2}$.

Suppose the direction picked is horizontal, that is $\mathbf{e}_{d_k^j} = \mathbf{e}_1$, and let again without loss of generality $R = (0, L) \times (0, 1) \in C_j$. Then we distinguish three cases:

- If $L \leqslant \frac{1}{2}\delta^{-1}$ (which implies that $j \geqslant 0$ in (26)), we insert a translate of $(0, L) \times (0, L\delta)$ and hence cover a volume fraction $L\delta \leqslant \lambda^{-j-1/2} = \lambda^{-|j|-1/2}$.
- If $L \geqslant 2\delta^{-1}$ (which implies that $j \leqslant 0$ in (26)), we insert a translate of $(0, \delta^{-1}) \times (0, 1)$ and hence cover a volume fraction $\frac{\delta^{-1}}{L} \leqslant \lambda^{j-1/2} = \lambda^{-|j|-1/2}$.
- Finally, if $\frac{1}{2}\delta^{-1} \leqslant L \leqslant 2\delta^{-1}$ (which implies that $|j|$ is small in (26)), we only modify R inside a quadrant and hence may bound the volume fraction covered from above by $\frac{1}{4}$.

We thus cover at most a volume fraction $\max(\frac{1}{4}, \lambda^{-|j|-1/2})$ and recall that we picked $\mathbf{e}_{d_k^j} = \mathbf{e}_1$ with probability $\frac{1}{2}$. Combining this estimate and (27), we may thus choose

$$c_j = \frac{1}{2}\delta^2\lambda^{j-1/2} + \frac{1}{2}\ \max\left(\frac{1}{4}, \lambda^{-|j|-1/2}\right) \in (0, 1). \tag{28}$$

This establishes the claimed inequality (25) and hence concludes the proof. □

Remark C.4. We remark that our computations of $\mathbb{E}(|\mathcal{V}^k|) - \mathbb{E}(|\mathcal{V}^{k+1}|)$ in (25) are close to being sharp. More precisely, given the aspect ratio of a rectangle R we can precisely compute the expected volume fraction (of R) which is covered by the building block. Since in our buckets C_j we group ratios which differ by at most a factor $\lambda^{\pm 1}$, we may bound these volume fractions from above and below by constants c_j^* and c_j which differ from each other by a factor at most $\lambda^{\pm 1}$ (see lemma C.6 for a calculation of lower bounds).

The remainder of this section is concerned with establishing the claimed estimate (24). More precisely, we make the stronger claim that there exists a constant $C > 1$ (for our purpose the constant can, for instance, be chosen to be $C = 100$) such that for all $k \geqslant 0$ and all $j \leqslant J_1$ it holds that

$$\mathbb{E}(V_j^k) \leqslant C\lambda^j \mathbb{E}(|\mathcal{V}^k|). \tag{29}$$

That is, rectangles with a very large aspect ratio comparable to $\lambda^{|j|}\delta^{-1}$ cover an exponentially decreasing amount of the total volume $\mathbb{E}(|\mathcal{V}^k|)$. As $j \leqslant J_1 < 0$ is negative, we may relate this to the geometric series in $\frac{1}{\lambda} < 1$ (starting at $|J_1|$) and after possibly choosing J_1 even more negative it holds that

$$C\sum_{j\leqslant J_1} \lambda^j = C\frac{\lambda^{J_1}}{1-\lambda^{-1}} \leqslant 0.1,$$

which implies the desired result (24). In order to prove (29) we proceed by induction using an upper and a lower bound given by the following two lemmas.

Lemma C.5. *Let $\mathcal{V}^k$ be as above. Suppose that (29) holds for a given k and C large ($C = 100$) and $\lambda = 1.1$. Then for all $j \leqslant J_1 < 0$ it holds that*

$$\mathbb{E}(V_j^{k+1}) \leqslant 0.7C\lambda^j \mathbb{E}(|\mathcal{V}^k|). \tag{30}$$

Lemma C.6. *Let $\mathcal{V}^k$ be as above. Then it holds that*

$$\mathbb{E}(|\mathcal{V}^{k+1}|) \geqslant 0.7\mathbb{E}(|\mathcal{V}^k|). \tag{31}$$

Remark C.7. We remark that a failure of the lower bound (31) corresponds to covering a large volume fraction in a single iteration step of the algorithm, which at first sight seems very desirable. However, by covering this large volume fraction we might possibly lose control of relative volume fractions (e.g. it might be that the tail is not anymore relatively small). We believe that (31) remains true also for algorithm 3.3, but our current method of proof for that case only allows to derive a suboptimal lower bound by $0.6\mathbb{E}(|\mathcal{V}^k|)$, which is not sufficient to close the argument. For simplicity of presentation we hence opted to modify the algorithm to cover a lower fraction in the 'best case' (leading to the condition (b) in algorithm C.1). We comment on some partial results for the unmodified case at the end of this section in remark C.8.

We emphasize that in contrast to lemma C.6 the result of lemma C.5 is valid for both algorithms 3.3 and C.1 and, in particular, does not need the modifications from algorithm C.1.

The combination of lemmas C.5 and C.6 allows us to prove (29).

Proof of the claim (29) *using lemmas C.5 and C.6.* We note that initially, that is for $k = 0$, $V_j^k = 0$ for all $j \leqslant J_1$ and thus (29) is trivially satisfied. We then aim to proceed by induction. Suppose that (29) holds for a given k and with $\lambda = 1.1$. Then by lemmas C.5 and C.6 it holds that

$$\mathbb{E}(V_j^{k+1}) \overset{(30)}{\leqslant} C0.7\lambda^j \mathbb{E}(|\mathcal{V}^k|) \overset{(31)}{\leqslant} C\lambda^j \frac{0.7}{0.7}\mathbb{E}(|\mathcal{V}^{k+1}|) = C\lambda^j \mathbb{E}(|\mathcal{V}^{k+1}|),$$

and the estimate (29) therefore also holds for $k + 1$. We thus conclude by induction. □

It remains to prove lemmas C.5 and C.6.

Proof of lemma C.5. We argue similarly as in equation (25) in the proof of theorem 7 and estimate $\mathbb{E}(V_j^{k+1})$ in terms of $\mathbb{E}(V_l^k)$, $l \in \mathbb{Z} \cap \{m \leqslant J\}$. Here we use that every rectangle in $\mathcal{V}^{k+1}$ is obtained as one of the connected components of a rectangle $R \in \mathcal{V}^k$ generated by inserting a building block (see figure 19). More precisely, let $j \leqslant J_1$ be arbitrary but fixed and let $R \in \mathcal{V}^k$ be a given rectangle. As in the proof of theorem 7 after rescaling and rotation we may assume that

$$R = (0, L) \times (0, 1), L \geqslant 1.$$

Then given the random point $\mathbf{p} \in R$ and direction $\mathbf{e}_d \in \{\mathbf{e}_1, \mathbf{e}_2\}$ we insert a building block $\mathcal{B} = \mathcal{B}(\mathbf{p}, \mathbf{e}_d) \subset R$, which divides

$$R \backslash \mathcal{B} =: R_1 \cup R_2,$$

into two connected components $R_1 = R_1(\mathbf{p}, \mathbf{e}_d)$ and $R_2 = R_2(\mathbf{p}, \mathbf{e}_d)$ (if $\mathcal{B}$ touches the boundary of R some of these components might be trivial). We then compute the contribution of $R \in \mathcal{V}^k$ to $\mathbb{E}(V_j^{k+1})$ by determining for which $\mathbf{p}$ and $\mathbf{e}_d$ it holds that $R_1 \in C_j$ or $R_2 \in C_j$ (and integrating $|R_1|$ and $|R_2|$ with respect to the probability density) and finally sum over all R.

More precisely, we claim that for any $j \leqslant J_1$ it holds that

$$\begin{aligned}\mathbb{E}(V_j^{k+1}) \leqslant\ & (0.5\lambda^{j+1/2}\delta + 0.5\lambda^{j+1/2})\mathbb{E}(V_{\geqslant 0}^k) \\ & + \delta^2\lambda^{j+1/2}\mathbb{E}(V_{0\geqslant l>j}^k) \\ & + (1-\lambda^{-2})\mathbb{E}(V_j^k) \\ & + \sum_{l<j}\lambda^{2l-2j}(1-\lambda^{-2})\mathbb{E}(V_l^k).\end{aligned} \tag{32}$$

Here we used the short-hand notation

$$\mathbb{E}(V_{\geqslant 0}^k) := \sum_{l\geqslant 0}\mathbb{E}(V_l^k), \qquad \mathbb{E}(V_{0\geqslant l>j}^k) := \sum_{0\geqslant l>j}\mathbb{E}(V_l^k).$$

Using (29) and the fact that C is large, we will argue that the main contribution on the right-hand side of (32) is given by the last two terms. More precisely, inserting the estimate (29), the last two contributions are controlled by

$$(1-\lambda^{-2})\sum_{l\geqslant j} C\lambda^{2j-3l}\mathbb{E}(|\mathcal{V}^k|) = C\lambda^{-j}\frac{1-\lambda^{-2}}{1-\lambda^{-3}}\mathbb{E}(|\mathcal{V}^k|).$$

In particular, we observe that

$$\beta(\lambda) := \frac{1-\lambda^{-2}}{1-\lambda^{-3}} = \frac{\lambda^{-1}+1}{\lambda^{-2}+\lambda^{-1}+1}$$

approaches $\frac{2}{3}$ as λ approaches 1. Inserting these estimates into (32) and choosing $\lambda = 1.1$, we may thus deduce that

$$\mathbb{E}(V_j^{k+1}) \leqslant \lambda^j\left(\delta\lambda^{\frac{1}{2}} + \delta^2\lambda^{\frac{1}{2}} + 0.5\lambda^{\frac{1}{2}} + 0.68C\right)\mathbb{E}(|\mathcal{V}^k|) < 0.7C\lambda^j\,\mathbb{E}(|\mathcal{V}^k|),$$

provided C is sufficiently large compared to 0.5.

It thus remains to prove the estimate (32). Let thus $j \leqslant J_1$ be arbitrary but fixed. As discussed above, for any $R \in \mathcal{V}^k$ we determine with which probability R_1 and R_2 are in C_j by estimating the probability of the associated sets of $(\mathbf{p}, \mathbf{e}_d)$. Using this, we compute the expectations of $|R_1|1_{R_1 \in C_j} + |R_2|1_{R_2 \in C_j}$ and then compare these to the volume $|R|$ of R.

We remark that if the aspect ratio $1 : L$, $L \geqslant 1$ of R satisfies $\frac{1}{2}\delta^{-1} \leqslant L \leqslant 2\delta^{-1}$ and we are thus in case (b) of algorithm C.1, then the rectangles R_1, R_2 are further rescaled by a factor $\frac{1}{2}$ and hence cover $\frac{1}{4}$ of the volume which they would else have occupied without this modification. Since we only require upper bounds on $|R_1|1_{R_1 \in C_j} + |R_2|1_{R_2 \in C_j}$, this gain of a factor $\frac{1}{4}$ only improves the estimates. Thus, for simplicity of notation in the following we establish the stronger estimate for the algorithm without this second modification.

In the following let always $R \in \mathcal{V}^k$ and without loss of generality, after rescaling, rotating and translating let

$$R = (0, L) \times (0, 1)$$

with $L \geqslant 1$.

The contribution by $V^k_{\geqslant 0}$: suppose that R is such that $1 \leqslant L \leqslant \delta^{-1}$ (and hence $R \in C_l$ for some $l \geqslant 0$). We then want to estimate the volume of the generated rectangles R_1, R_2 if they are in C_j. Here we say that R_1 (or R_2) is *vertical* if it is a translate of $(0, a) \times (0, 1)$ for some $a \in (0, 1)$ (that is the $\mathbf{e}_2$ direction is the longest) and otherwise call it *horizontal*.

Let us first consider the case when R_1 (or R_2) is in C_j and vertical. Then R_1 is a translate of $(0, \lambda^j\delta\gamma) \times (0, 1)$ (see figure 19 left) with $\gamma \in (\lambda^{-1/2}, \lambda^{+1/2})$ (since the class C_j was defined in this way). We may thus roughly bound its volume fraction by

$$\frac{\lambda^{j+1/2}\delta}{L} \leqslant \lambda^{j+1/2}\delta.$$

Next suppose that R_1 (or R_2) is in C_j and horizontal. Then (by the definition of $V^k_{\geqslant 0}$ and C_j as well as the replacements explained in algorithm C.1) R_1 is a translate of $(0, L) \times (0, \alpha)$ (see figure 19 top right) with $\frac{\alpha}{L} \in (\lambda^{j-1/2}\delta^{-1}, \lambda^{j+1/2}\delta^{-1})$ and thus in particular covers a volume fraction less than

$$\alpha \leqslant \lambda^{j+1/2}.$$

The contribution by V^k_l with $0 \geqslant l > j$: let again $R \in \mathcal{V}^k$ and suppose that $\delta^{-1} \leqslant L < \delta^{-1}\lambda^{-j-1/2}$ (and thus $R \in C_l$, $0 \geqslant l > j$). Since $L \geqslant \delta^{-1}$, the generated rectangles R_1, R_2 have height 1 and are thus translates of $(0, \alpha) \times (0, 1)$ for some $\alpha \in (0, L)$ (see figure 19 center). As $\alpha < L < \delta^{-1}\lambda^{-j-1/2}$, it is not possible for R_1 or R_2 to be horizontal rectangles in C_j (since α is too small for the rectangles to be in C_j). If R_1 (or R_2) is a vertical rectangle and in C_j, it is a translate of $(0, \lambda^j\delta\gamma) \times (0, 1)$ with $\gamma \in (\lambda^{-1/2}, \lambda^{1/2})$ and hence covers a volume fraction at most

$$\frac{\lambda^j\delta\gamma}{L} \leqslant \lambda^{j+1/2}\delta^2.$$

The contribution by V^k_l with $l \leqslant j$: finally, let again $R \in \mathcal{V}^k$ and suppose that $R \in C_l$ and thus

$$\lambda^{-l-1/2}\delta^{-1} \leqslant L < \lambda^{-l+1/2}\delta^{-1}.$$

If the generated rectangle R_1 (or R_2) is vertical, it will only cover a volume fraction

$$\frac{\delta\lambda^j\gamma}{L} \leqslant \delta^2\lambda^{1+j+l}, \tag{33}$$

which is negligible.

In the following we thus focus on estimating the expected volume fraction covered by R_1 in C_j being a horizontal rectangle (which by symmetry is the same volume fraction as covered by R_2). Here for concreteness we again fix

$$R = (0, L) \times (0, 1)$$

and let $R_1 = R_1(\mathbf{p}, \mathbf{e}_d)$ be the rectangle generated on the left of the building block $\mathcal{B}$ which had been removed from R (see figure 19).

The case $l = j$: by construction the building block $\mathcal{B}$ inserted depends on $\mathbf{p}$ only in terms of its $\mathbf{e}_1$ component p_1. We thus ask for which p_1 (for given $\mathbf{e}_d$) it holds that $R_1 \in C_j$ and require an estimate of $|R_1(p_1, \mathbf{e}_d)|$ in that case. Since p_1 was chosen according to the Lebesgue measure, we then can compute the volume fraction for a given $\mathbf{e}_{d'}$ by

$$\begin{aligned}\mathbb{E}\left(\frac{|R_1|}{|R|} :\ R_1 \in C_j \text{ is horizontal},\ d = d'\right) &= \frac{1}{|R|}\int_0^L |R_1(p_1, \mathbf{e}_{d'})| 1_{R_1 \in C_j} \frac{\mathrm{d}p_1}{L} \\ &= \frac{1}{L^2}\int_0^L |R_1(p_1, \mathbf{e}_{d'})| 1_{R_1 \in C_j}\, \mathrm{d}p_1.\end{aligned}$$

We first discuss the case when p_1 is not close to 0 or L and the building block $\mathcal{B}$ is thus centered at $\mathbf{p}$ (see the definitions of δ_k^j and λ_k^j in algorithm C.1). Then if $\mathbf{e}_d = \mathbf{e}_1$, $\mathcal{B} = (p_1 - \delta^{-1}/2, p_1 + \delta^{-1}/2) \times (0, 1)$ and $R_1 = (0, p_1 - \delta^{-1}/2) \times (0, 1)$. Similarly, if $\mathbf{e}_d = \mathbf{e}_2$, $\mathcal{B} = (p_1 - \delta/2, p_1 + \delta/2) \times (0, 1)$ and $R_1 = (0, p_1 - \delta/2) \times (0, 1)$. Thus, for R_1 to be in C_j, we need that either $p_1 - \delta^{-1}/2 \in (\lambda^{-j-1/2}\delta^{-1}, \lambda^{-j+1/2}\delta^{-1})$ or $p_1 - \delta/2 \in (\lambda^{-j-1/2}\delta^{-1}, \lambda^{-j+1/2}\delta^{-1})$, respectively.

We remark that if $\mathcal{B}$ is not centered in p_1, it touches the right-boundary. Hence, the generated rectangle R_1 will only be shorter than it would be otherwise and we may hence bound from above by the previously derived formula.

Introducing the new variables of integration $x_1 = p_1 - \delta^{-1}/2 \leqslant L$ or $x_1 = p_1 - \delta/2 \leqslant L$, we may thus bound

$$\begin{aligned}\frac{1}{L^2}\int_0^L |R_1(p_1, e_d)| 1_{R_1 \in C_j}\, \mathrm{d}p_1 &\leqslant \frac{1}{L^2}\int_{\lambda^{-j-1/2}\delta^{-1}}^L x_1\, \mathrm{d}x_1 \\ &= \frac{x_1^2}{2L^2}\Big|_{\lambda^{-j-1/2}\delta^{-1}}^L.\end{aligned} \tag{34}$$

We recall that by symmetry the volume fraction due to R_2 on the right-hand side is of the same size and we can hence estimate the full volume fraction by:

$$\begin{aligned}&\mathbb{E}\left(\frac{|R_1|}{|R|} :\ R_1 \in C_j \text{ is horizontal}\right) + \mathbb{E}\left(\frac{|R_2|}{|R|} :\ R_2 \in C_j \text{ is horizontal}\right) \\ &\leqslant \frac{x_1^2}{L^2}\Big|_{\lambda^{-j-1/2}\delta^{-1}}^L = 1 - \left(\frac{\lambda^{-j-1/2}\delta^{-1}}{L}\right)^2.\end{aligned}$$

We recall that $L \in (\lambda^{-j-1/2}\delta^{-1}, \lambda^{-j+1/2}\delta^{-1})$ and thus this is zero if L is on the smaller end of the range and bounded above by

$$1 - \lambda^{-2}.$$

The case $l < j$: we argue analogously as in the case $l = j$ except that the upper limit of the interval of integration in the analogue of (34) is given by $\lambda^{-j+1/2}\delta^{-1} < L$ instead: thus, in this case,

$$\mathbb{E}\left(\frac{|R_1|}{|R|} : \ R_1 \in C_j \text{ is horizontal}\right) + \mathbb{E}\left(\frac{|R_2|}{|R|} : \ R_2 \in C_j \text{ is horizontal}\right)$$
$$\leqslant \frac{x_1^2}{L^2}\Big|_{\lambda^{-j-1/2}\delta^{-1}}^{\lambda^{-j+1/2}\delta^{-1}} = \frac{\lambda^{-2j+1}\delta^{-2} - \lambda^{-2j-1}\delta^{-2}}{L^2}.$$

Since $L^2 \geqslant \lambda^{-2l-1}\delta^{-2}$ and $l \leqslant j-1$ this can be estimated by

$$\frac{\lambda^{-2j+1} - \lambda^{-2j-1}}{\lambda^{-2l-1}} \leqslant \lambda^{-2j+2l}(1 - \lambda^{-2}).$$

This concludes the proof of the claim (32) and thus of the lemma. □

It remains to prove lemma C.6. Due to the modification (b) in the definition of algorithm C.1 we here obtain a very short, straightforward proof. Subsequently we discuss how to obtain similar results for algorithm 3.3 using more sophisticated methods.

Proof of lemma C.6. We argue similarly as in the derivation of equation (25) and claim that there exist constants such that we obtain the following *upper* bound on the volume covered:

$$\mathbb{E}(|\mathcal{V}^k|) - \mathbb{E}(|\mathcal{V}^{k+1}|) \leqslant \sum_j c_j \mathbb{E}(V_j^k), \tag{35}$$

for constants $c_j > 0$ which are independent of k. Since $\sum_j \mathbb{E}(V_j^k) = \mathbb{E}(|\mathcal{V}^k|)$, a (possibly highly suboptimal) upper bound of the right-hand side is given by

$$\max(c_j)\mathbb{E}(|\mathcal{V}^k|).$$

It hence follows that

$$\mathbb{E}(|\mathcal{V}^{k+1}|) \geqslant (1 - \max(c_j))\mathbb{E}(|\mathcal{V}^k|).$$

We now claim that due to the second modification in the definition of algorithm C.1 it holds that $\max(c_j) \leqslant 0.3$ and the result hence follows.

In order to compute the constants c_j, we again individually consider each rectangle $R \in \mathcal{V}^k$ and after rescaling and possibly rotating by $\frac{\pi}{2}$ may assume that

$$R = (0, L) \times (0, 1)$$

with $L \geqslant 1$. If we pick the vertical direction, $\mathbf{e}_d = \mathbf{e}_2$, (which occurs with probability $\frac{1}{2}$), we insert a translate of $(0, \delta) \times (0, 1)$ and hence cover a very small fraction

$$\frac{\delta}{L} \leqslant \delta.$$

If we instead pick the horizontal direction, $\mathbf{e}_d = \mathbf{e}_1$, (which also occurs with probability $\frac{1}{2}$), we expect to cover a larger volume fraction of R. We distinguish three cases:

- If $L < \delta^{-1}/2$, the inserted rectangle is a translate of $(0, L) \times (0, L\delta)$ and hence covers a volume fraction
$$L\delta \leqslant \frac{1}{2}.$$
See figure 19 top right.
- Similarly, if $L \geqslant 2\delta^{-1}$, we insert a translate of $(0, \delta^{-1}) \times (0, 1)$ and thus cover a volume fraction
$$\frac{\delta^{-1}}{L} \leqslant \frac{1}{2}.$$
See figure 19 center right.
- Finally, if $\frac{\delta^{-1}}{2} \leqslant L < 2\delta^{-1}$, we are in the case (b) of algorithm C.1 (see figure 19 bottom). As we only modify R inside one quadrant we cover at most
$$\frac{1}{4}$$
of the volume.

Thus, in all theses cases for $\mathbf{e}_d = \mathbf{e}_1$ we cover at most $\frac{1}{2}$ of the volume.

Combining the estimates for $\mathbf{e}_d = \mathbf{e}_2$ and $\mathbf{e}_d = \mathbf{e}_1$ (each with probability $1/2$), then yields the bound

$$c_j \leqslant \frac{1}{2}\delta + \frac{1}{2}\frac{1}{2} = \frac{1}{2}\delta + \frac{1}{4} \leqslant 0.3,$$

provided $\delta \leqslant 0.1$. □

Remark C.8. Finally, let us briefly comment on some additional challenges in carrying out the tail estimates (29) *without* the quadrant modification (b) in algorithm C.1. Consider a rectangle $R \in \mathcal{V}^k$, $R = (0, L) \times (0, 1)$ with $\frac{1}{2}\delta^{-1} \leqslant L \leqslant 2\delta^{-1}$ and suppose we picked $\mathbf{e}_d = \mathbf{e}_1$. Then, if $\frac{1}{2}\delta^{-1} \leqslant L \leqslant \delta^{-1}$, we insert a translate of $(0, L) \times (0, L\delta)$ and cover a fraction $L\delta \in [\frac{1}{2}, 1]$. Similarly, if $\delta^{-1} \leqslant L \leqslant 2\delta^{-1}$, we insert a translate of $(0, \delta^{-1}) \times (0, 1)$ and cover a fraction $\frac{\delta^{-1}}{L} \in [\frac{1}{2}, 1]$. Therefore, the best naive upper bound for $\max(c_j)$ as in the proof of lemma C.5 we can achieve is given by

$$\frac{1}{2}\delta + \frac{1}{2}1,$$

and hence

$$\mathbb{E}(|\mathcal{V}^{k+1}|) \geqslant \left(1 - \frac{1}{2}\delta - \frac{1}{2}1\right)\mathbb{E}(|\mathcal{V}^k|) = \left(0.5 - \frac{1}{2}\delta\right)\mathbb{E}(|\mathcal{V}^k|).$$

Unlike the factor 0.7 obtained in lemma C.6 this estimate is not sufficient to close the inductive argument for (29).

In order to improve this bound, we thus need to exploit that the estimate in terms of $\max(c_j)$ is very rough and not actually attained. Indeed, we may employ an approach similar to the one of lemma C.5 to show that

$$\mathbb{E}(V_0^k) \leqslant \theta\mathbb{E}(|\mathcal{V}^k|),$$

for an explicit constant $\theta \in (0,1)$. That is, only some part of the total volume is covered by rectangles $R \in C_0$. Then instead of bounding by $\max(c_j)$ we may use

$$\theta\left(\frac{1}{2}\delta + \frac{1}{2}1\right) + (1-\theta)\max_{j\neq 0} c_j.$$

Unfortunately, while these and further improvements allow us to deduce that $\mathbb{E}(|\mathcal{V}^{k+1}|) \geqslant 0.6\mathbb{E}(|\mathcal{V}^k|)$, this still is not sufficient to close the inductive estimate (29). We thus opted to simplify discussions by considering the modified algorithm C.1.

Appendix D. Regularity of the algorithm of the numerical implementation

Last but not least, we prove that the slightly modified algorithms from section 8 satisfy the same regularity properties as their original counterparts. The simplifications are thus only technical and allow for a more efficient numerical implementation.

Indeed, in order to prove that the modified algorithms satisfy analogous regularity properties we argue as in the proof of theorem 4. With the notation as in section 8.1, we thus have to ensure that the modification of our construction still satisfies an estimate of the form

$$\begin{aligned}\int_\Omega\int_\Omega \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \nabla \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} &\leqslant \sum_j \int_{\mathcal{B}_k^j}\int_{\mathcal{B}_k^j} \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \nabla \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{dxd}\mathbf{y}\\ &\quad + 2\sum_j \int_{\mathcal{B}_k^j}\int_{(\mathcal{B}_k^j)^c} \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \nabla \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\\ &\leqslant c\sum_j \mathrm{per}(\mathcal{B}_k^j)|\mathcal{B}_k^j|^{1-sp},\end{aligned}$$

where, in our modified construction, we have to replace the old building blocks $\mathcal{B}_k^j$ by the blocks $\mathcal{B}$ described above. Since the second contribution is estimated 'generically', not using properties of v_k (see the proof of theorem 4), it suffices to discuss contributions of the form

$$\int_{\mathcal{B}}\int_{\mathcal{B}} \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \nabla \mathsf{v}_k(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}.$$

To this end, we consider the two cases (a) and (b) described above: first, by stacking N blocks of the microstructures on top of each other (that means in case $\mathcal{B} = \bigcup_{n=0}^{N-1}(\mathbf{c}_0 + n\lambda\mathbf{e}_i + \lambda\Omega_i) =: \bigcup_{n=1}^{N-1}\Omega_i^n$), we have that for $\mathsf{v}_k := \mathbf{y}_{k+1} - \mathbf{y}_k$ (cf proof of theorem 4)

$$\begin{aligned}&\int_{\mathcal{B}}\int_{\mathcal{B}} \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \nabla \mathsf{v}_k(\hat{\mathbf{x}})|^p}{|\mathbf{x}-\hat{\mathbf{x}}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\hat{\mathbf{x}}\\ &= \sum_{n=0}^{N-1}\int_{\Omega_i^n}\int_{\Omega_i^n} \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \mathsf{v}_k(\hat{\mathbf{x}})|^p}{|\mathbf{x}-\hat{\mathbf{x}}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\hat{\mathbf{x}}\\ &\quad + \sum_{n=0}^{N-1}\int_{\mathcal{B}\setminus\Omega_i^n}\int_{\Omega_i^n} \frac{|\nabla \mathsf{v}_k(\mathbf{x}) - \nabla \mathsf{v}_k(\hat{\mathbf{x}})|^p}{|\mathbf{x}-\hat{\mathbf{x}}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\hat{\mathbf{x}}\end{aligned}$$

$$\leqslant N\left|\nabla \mathbf{z}_i\left(\frac{\mathbf{x}-\mathbf{c}_0}{\lambda}\right)\right|^p_{\dot{W}^{s,p}(\mathbf{c}_0+\lambda\Omega_i)} + cN\,\mathrm{per}(\Omega_i^n)^{sp}|\Omega_i^n|^{1-sp}$$
$$\leqslant N\left|\nabla \mathbf{z}_i\left(\frac{\mathbf{x}-\mathbf{c}_0}{\lambda}\right)\right|^p_{\dot{W}^{s,p}(\mathbf{c}_0+\lambda\Omega_i)} + C(N\alpha\ell_d)^{sp}(|\mathcal{B}|)^{1-sp}$$
$$\leqslant N\left|\nabla \mathbf{z}_i\left(\frac{\mathbf{x}-\mathbf{c}_0}{\lambda}\right)\right|^p_{\dot{W}^{s,p}(\mathbf{c}_0+\lambda\Omega_i)} + C(\mathrm{per}(\mathcal{B}))^{sp}(|\mathcal{B}|)^{1-sp}.$$

But since

$$\begin{aligned}&\int_{\lambda\Omega_i}\int_{\lambda\Omega_i}\frac{|\nabla\mathbf{z}_i(\lambda^{-1}\mathbf{x})-\nabla\mathbf{z}_i(\lambda^{-1}\hat{\mathbf{x}})|^p}{|\mathbf{x}-\hat{\mathbf{x}}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\hat{\mathbf{x}}\\ &=\lambda^{2-sp}\int_{\Omega_i}\int_{\Omega_i}\frac{|\nabla\mathbf{z}_i(\mathbf{x})-\nabla\mathbf{z}_i(\hat{\mathbf{x}})|^p}{|\mathbf{x}-\hat{\mathbf{x}}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\hat{\mathbf{x}}\leqslant\lambda^{2-sp}|\nabla\mathbf{z}_i|^p_{W^{s,p}(\Omega_i)},\end{aligned}\tag{36}$$

and since in this case

$$N\lambda^{2-sp}=(N\lambda)^{sp}(N\lambda^2)^{1-sp}\leqslant c(\mathrm{per}(\mathcal{B}))^{sp}|\mathcal{B}|^{1-sp},$$

we obtain

$$\begin{aligned}&\int_{\mathcal{B}}\int_{\mathcal{B}}\frac{|\nabla\mathbf{v}_k(\mathbf{x})-\nabla\mathbf{v}_k(\hat{\mathbf{x}})|^p}{|\mathbf{x}-\hat{\mathbf{x}}|^{2+sp}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\hat{\mathbf{x}}\\ &\leqslant c(\mathrm{per}(\mathcal{B}))^{sp}|\mathcal{B}|^{1-sp}|\nabla\mathbf{z}_i|^p_{W^{s,p}(\Omega_i)}+(\mathrm{per}(\mathcal{B}))^{sp}|\mathcal{B}|^{1-sp}.\end{aligned}$$

We now notice that, by (36), also when $\mathcal{B}=\mathbf{c}_0+\lambda\Omega_i$ we have

$$|\nabla\mathbf{v}_k|^p_{W^{s,p}(\mathcal{B})}\leqslant\lambda^{2-sp}|\nabla\mathbf{z}_i|^p_{W^{s,p}(\Omega_i)}\leqslant c|\mathcal{B}|^{1-sp}(\mathrm{per}(\mathcal{B}))^{sp}|\nabla\mathbf{z}_i|^p_{W^{s,p}(\Omega_i)}.$$

Therefore, we still infer (19) in the proof of theorem 4, and hence we deduce the same regularity result as in section 6 above also under this implementation of the model.

ORCID iDs

Francesco Della Porta https://orcid.org/0000-0002-3090-5770
Angkana Rüland https://orcid.org/0000-0003-1169-344X
Jamie M Taylor https://orcid.org/0000-0002-5423-828X
Christian Zillinger https://orcid.org/0000-0002-0480-2719

References

[AA01] Ahluwalia R and Ananthakrishna G 2001 Power-law statistics for avalanches in a martensitic transformation *Phys. Rev. Lett.* **86** 4076–9
[BTW87] Bak P, Tang C and Wiesenfeld K 1987 Self-organized criticality: an explanation of the 1/*f* noise *Phys. Rev. Lett.* **59** 381
[BBB15] Balandraud X, Barrera N, Biscari P, Grédiac M and Zanzotto G 2015 Strain intermittency in shape-memory alloys *Phys. Rev.* B **91** 174111
[Bal02] Ball J M 2002 Some open problems in elasticity *Geometry, Mechanics, and Dynamics* (Berlin: Springer) pp 3–59

[Bal04] Ball J M 2004 Mathematical models of martensitic microstructure *Mater. Sci. Eng.* A **378** 61–9
[BCH15] Ball J M, Cesana P and Hambly B 2015 A probabilistic model for martensitic avalanches *MATEC Web of Conf.* vol 33 (EDP Sciences) p 02008
[BJ89] Ball J M and James R D 1989 Fine phase mixtures as minimizers of energy *Analysis and Continuum Mechanics* (Berlin: Springer) pp 647–86
[BJ92] Ball J M and James R D 1992 Proposed experimental tests of a theory of fine microstructure and the two-well problem *Phil. Trans. R. Soc.* A **338** 389–450
[Ber06] Bertoin J 2006 *Random Fragmentation and Coagulation Processes* vol 102 (Cambridge: Cambridge University Press)
[Bha03] Bhattacharya K 2003 *Microstructure of Martensite: Why it Forms and How it Gives Rise to the Shape-Memory Effect* (*Oxford Series on Materials Modeling*) (Oxford: Oxford University Press)
[BUZZ16] Biscari P, Urbano M F, Zanzottera A and Zanzotto G 2016 Intermittency in crystal plasticity informed by lattice symmetry *J. Elast.* **123** 85–96
[BBG20] Blaysat B, Balandraud X, Grédiac M, Vives E, Barrera N and Zanzotto G 2020 Concurrent tracking of strain and noise bursts at ferroelastic phase fronts *Commun. Mater.* **1** 3
[CO09] Capella A and Otto F 2009 A rigidity result for a perturbation of the geometrically linear three-well problem *Commun. Pure Appl. Math.* **62** 1632–69
[CO12] Capella A and Otto F 2012 A quantitative rigidity result for the cubic-to-tetragonal phase transition in the geometrically linear theory with interfacial energy *Proc. R. Soc. Edinburgh* A **142** 273–327
[CMO98] Carrillo L, Mañosa L, Ortín J, Planes A and Vives E 1998 Experimental evidence for universality of acoustic emission avalanche distributions during structural transitions *Phys. Rev. Lett.* **81** 1889
[CDPR20] Cesana P, Della Porta F, Rüland A, Zillinger C and Zwicknagl B 2020 Exact constructions in the (nonlinear) planar theory of elasticity: from elastic crystals to nematic elastomers *Arch. Ration. Mech. Anal.* **237** 383–445
[CH18] Cesana P and Hambly B 2018 A probabilistic model for interfaces in a martensitic phase transition (arXiv:1810.04380)
[CZ01] Chung K L and Zhong K 2001 *A Course in Probability Theory* (New York: Academic)
[Con00] Conti S 2000 Branched microstructures: scaling and asymptotic self-similarity *Commun. Pure Appl. Math.* **53** 1448–74
[Con08] Conti S 2008 Quasiconvex functions incorporating volumetric constraints are rank-one convex *J. Math. Pure Appl.* **90** 15–30
[CDMZ20] Conti S, Diermeier J, Melching D and Zwicknagl B 2020 Energy scaling laws for geometrically linear elasticity models for microstructures in shape memory alloys *ESAIM Control, Optim. Calc. Var.* **26** 115
[CDK07] Conti S, Dolzmann G and Kirchheim B 2007 Existence of Lipschitz minimizers for the three-well problem in solid-solid phase transitions *Ann. Inst. Henri Poincare* C **24** 953–62
[CKZ17] Conti S, Klar M and Zwicknagl B 2017 Piecewise affine stress-free martensitic inclusions in planar nonlinear elasticity *Proc. R. Soc.* A **473** 20170235
[CT05] Conti S and Theil F 2005 Single-slip elastoplastic microstructures *Arch. Ration. Mech. Anal.* **178** 125–48
[DP19b] Della Porta F 2019 A model for the evolution of highly reversible martensitic transformations *Math. Models Methods Appl. Sci* **29** 493
[DP19a] Della Porta F 2019 Analysis of a moving mask hypothesis for martensitic transformations *J. Nonlinear Sci.* **29** 2341–84
[DPR20] Della Porta F and Rüland A 2020 Convex integration solutions for the geometrically nonlinear two-well problem with higher Sobolev regularity (arXiv:1905.12521)
[DM95] Dolzmann G and Müller S 1995 The influence of surface energy on stress-free microstructures in shape memory alloys *Meccanica* **30** 527–39
[FGRV95] Frontera C, Goicoechea J, Ràfols I and Vives E 1995 Sequential partitioning: an alternative to understanding size distributions of avalanches in first-order phase transitions *Phys. Rev.* E **52** 5671
[Ia] Inamura T in preparation.

[IHM13] Inamura T, Hosoda H and Miyazaki S 2013 Incompatibility and preferred morphology in the self-accommodation microstructure of β-titanium shape memory alloy *Phil. Mag.* **93** 618–34

[Kir98] Kirchheim B 1998 Lipschitz minimizers of the 3-well problem having gradients of bounded variation https://mis.mpg.de/preprints/1998/preprint1998_12.pdf

[Kir03] Kirchheim B 2003 Rigidity and geometry of microstructures *MPI-MIS Lecture Notes* https://mis.mpg.de/publications/other-series/ln/lecturenote-1603.html

[KK11] Knüpfer H and Kohn R V 2011 Minimal energy for elastic inclusions *Proc. R. Soc.* A **467** 695–717

[KKO13] Knüpfer H, Kohn R V and Otto F 2013 Nucleation barriers for the cubic-to-tetragonal phase transformation *Commun. Pure Appl. Math.* **66** 867–904

[KM94] Kohn R V and Müller S 1994 Surface energy and microstructure in coherent phase transitions *Commun. Pure Appl. Math.* **47** 405–35

[MB05] Mayergoyz I D and Bertotti G 2005 *The Science of Hysteresis* (New York: Academic)

[MRSS95] Mohan Rao M, Sengupta S and Sahu H K 1995 Emergence of scale invariance in martensite growth *Solid State Phenomena* vol 42–43 (Switzerland: Trans Tech Publications Ltd) pp 133–40

[Mül99] Müller S 1999 Variational models for microstructure and phase transitions *Calculus of Variations and Geometric Evolution Problems* (Berlin: Springer) pp 85–210

[MŠ98] Müller S and Šverák V 1998 Unexpected solutions of first and second order partial differential equations *International Congress* p. 691

[MŠ99] Müller S and Šverák V 1999 Convex integration with constraints and applications to phase transitions and partial differential equations *J. Eur. Math. Soc.* **1** 393–422

[MS01] Müller S and Sychev M A 2001 Optimal existence theorems for nonhomogeneous differential inclusions *J. Funct. Anal.* **181** 447–75

[PLKK97] Pasko A Y, Likhachev A A, Koval Y N and Kolomytsev V I 1997 2D Fourier analysis and its application to study of scaling properties and fractal dimensions of ε-martensite distribution in γ-matrix of Fe–Mn–Si alloy *J. Phys. IV* **7** C5-435

[PRTZT16] Perez-Reche F J, Triguero C, Zanzotto G and Truskinovsky L 2016 Origin of scale-free intermittency in structural first-order phase transitions *Phys. Rev.* B **94** 144102

[PRTZ09] Pérez-Reche F J, Truskinovsky L and Zanzotto G 2009 Martensitic transformations: from continuum mechanics to spin models and automata *Contin. Mech. Thermodyn.* **21** 17–26

[PRV04] Pérez-Reche F J and Vives E 2004 Spanning avalanches in the three-dimensional Gaussian random-field Ising model with metastable dynamics: field dependence and geometrical properties *Phys. Rev.* B **70** 214422

[PRTZ07] Pérez-Reche F-J, Truskinovsky L and Zanzotto G 2007 Training-induced criticality in martensites *Phys. Rev. Lett.* **99** 075501

[PRTZ08] Pérez-Reche F-J, Truskinovsky L and Zanzotto G 2008 Driving-induced crossover: from classical criticality to self-organized criticality *Phys. Rev. Lett.* **101** 230601

[PMV13] Planes A, Mañosa L and Vives E 2013 Acoustic emission in martensitic transformations *J. Alloys Compd.* **577** S699–704

[Pom] Pompe W 2007 Convex integration based on the martingale convergence theorem http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.501.9762 & rep=rep1 & type=pdf

[RSS95] Rao M, Sengupta S and Sahu H K 1995 Kinematic scaling and crossover to scale invariance in martensite growth *Phys. Rev. Lett.* **75** 2164

[Rül16b] Rüland A 2016 A rigidity result for a reduced model of a cubic-to-orthorhombic phase transition in the geometrically linear theory of elasticity *J. Elast.* **123** 137–77

[Rül16a] Rüland A 2016 The cubic-to-orthorhombic phase transition: rigidity and non-rigidity properties in the linear theory of elasticity *Arch. Ration. Mech. Anal.* **221** 23–106

[RTZ18] Rüland A, Taylor J M and Zillinger C 2018 Convex integration arising in the modelling of shape-memory alloys: some remarks on rigidity, flexibility and some numerical implementations *J. Nonlinear Sci.* **29** 2137–84

[RZZ18] Rüland A, Zillinger C and Zwicknagl B 2018 Higher Sobolev regularity of convex integration solutions in elasticity: the Dirichlet problem with affine data in int(K^{lc}) *SIAM J. Math. Anal.* **50** 3791–841

[RZZ19] Rüland A, Zillinger C and Zwicknagl B 2019 Higher Sobolev regularity of convex integration solutions in elasticity: the planar geometrically linearized hexagonal-to-rhombic phase transformation *J. Elast.* **138** 1–76

[SKR09] Salje E K H, Koppensteiner J, Reinecker M, Schranz W and Planes A 2009 Jerky elasticity: avalanches and the martensitic transition in $Cu_{74.08}Al_{23.13}Be_{2.79}$ shape-memory alloy *Appl. Phys. Lett.* **95** 231908

[Sim17] Simon T 2017 Rigidity of branching microstructures in shape memory alloys (arXiv:1705.03664)

[SCD13] Song Y, Chen X, Dabade V, Shield T W and James R D 2013 Enhanced reversibility and unusual microstructure of a phase-transforming material *Nature* **502** 85

[TIVP17] Torrents G, Illa X, Vives E and Planes A 2017 Geometrical model for martensitic phase transitions: understanding criticality and weak universality during microstructure growth *Phys. Rev.* E **95** 013001

[VOM94] Vives E, Ortín J, Mañosa L, Ràfols I, Pérez-Magrané R and Planes A 1994 Distributions of avalanches in martensitic transformations *Phys. Rev. Lett.* **72** 1694