

Material Fracturing and Failure Simulation Datasets

*Original*

Material Fracturing and Failure Simulation Datasets / Hill, R.G., Gao, K., Pachalieva, A., Marcato, A., Wang, X., Grosset, P., Rougier, E., Lei, Z., Santos, J.E., Agrawal, V., Kang, Q., Hyman, J.D., Hunter, A., Sweeney, C.M., Debardeleben, N., Lawrence, E., Viswanathan, H., O'Malley, D.. - In: SCIENTIFIC DATA. - ISSN 2052-4463. - 13:1(2026), pp. 1-19.  
[10.1038/s41597-025-06412-8]

*Availability:*

This version is available at: 11583/3010682 since: 2026-05-08T11:14:30Z

*Publisher:*

Springer Nature

*Published*

DOI:10.1038/s41597-025-06412-8

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*



(Article begins on next page)



OPEN

# Material Fracturing and Failure Simulation Datasets

DATA DESCRIPTOR

Ryley G. Hill<sup>1,7</sup> , Kai Gao<sup>2,7</sup>, Aleksandra Pachalieva<sup>2</sup>, Agnese Marcato<sup>2</sup>, Xiaoyu Wang<sup>1</sup>, Pascal Grosset<sup>3</sup>, Esteban Rougier<sup>1</sup>, Zhou Lei<sup>1</sup>, Javier E. Santos<sup>2</sup>, Vinamra Agrawal<sup>4</sup>, Qinjun Kang<sup>2</sup>, Jeffrey D. Hyman<sup>2</sup>, Abigail Hunter<sup>4</sup>, Christine M. Sweeney<sup>5</sup>, Nathan DeBardeleben<sup>6</sup>, Earl Lawrence<sup>3</sup>, Hari Viswanathan<sup>2</sup> & Daniel O'Malley<sup>2</sup> 

Fracturing is a fundamental physics phenomena with broad relevance across multiple domains, ranging from infrastructure integrity, aerospace durability, reservoir production, and seismic events. We present a diverse dataset of simulated fracture evolution and material failure generated from two numerical solvers: the phase-field method and the combined finite-discrete element method (FDEM). These solvers differ in formulation, physical fidelity, and computational efficiency. The dataset includes five materials: PBX, anisotropic shale, tungsten, aluminum, and steel. For each, phase-field simulations span 400,000 cases: 200,000 under uniaxial tension and 200,000 under biaxial tension. The computationally expensive FDEM simulations include 90,000 split evenly among PBX, shale, and tungsten under uniaxial loading. All simulations begin with randomized initial fracture patterns. Each entry includes temporal data capturing fracture propagation dynamics. This comprehensive dataset is designed to support the development of foundational or surrogate machine learning approaches for predicting material failure. While no such models are introduced here, the dataset lays a robust foundation for advancing future research and innovation in these areas.

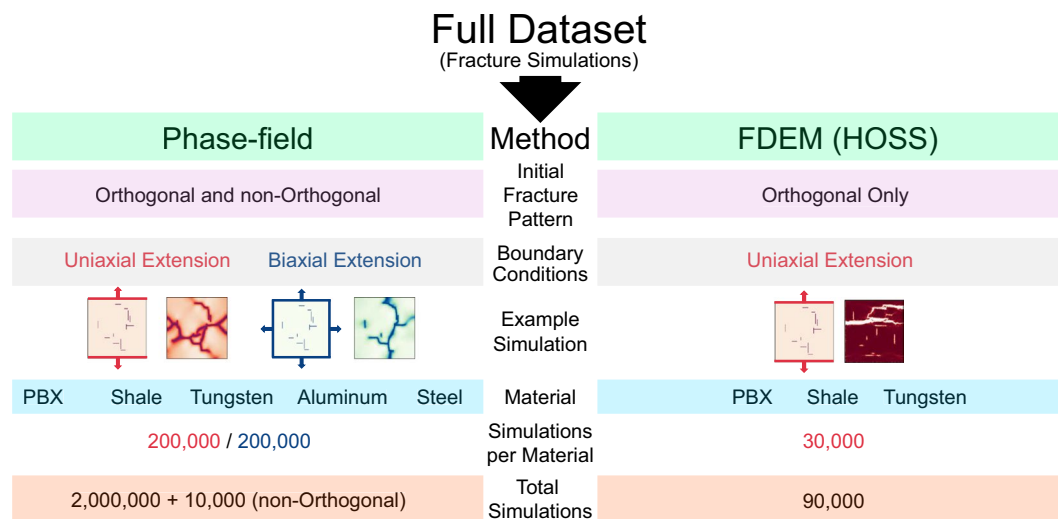
## Background & summary

Fracturing phenomena are ubiquitous across numerous scientific and engineering domains and play a critical role in various areas such as subsurface geology especially subsurface flow in fractured reservoir<sup>1,2</sup>, structural engineering<sup>3</sup>, earthquake rupture dynamics and anthropogenic seismic hazards<sup>4,5</sup>, infrastructure integrity<sup>6</sup>, response of human-made systems to impulsive loads<sup>7</sup>, material design<sup>8</sup>, high explosives performance<sup>9</sup>, biomaterials such as bones<sup>10</sup>. Accurately modeling how fractures initiate, propagate, and interact with materials under stress is critical for advancing knowledge and developing resilient, safe, and efficient materials and structures. Despite significant advancements in both theory and computational techniques, simulating fracturing evolution under internal stress or external loading remains computationally intensive and inherently challenging due to the complexity of physical interactions and material heterogeneity, sometimes involving multi-physical fields.

Recent breakthroughs in machine learning (ML), particularly within natural language processing, have demonstrated the transformative potential of large-scale, data-driven models<sup>11–13</sup>. Scientific disciplines, however, face unique challenges in applying such models due to the high cost of data generation, quality control complexities, and the multi-modal nature of scientific datasets. Nonetheless, fields such as protein folding<sup>14</sup>, drug discovery<sup>15</sup>, climate science<sup>16,17</sup>, and computational chemistry<sup>18,19</sup> have successfully leveraged large-scale ML, highlighting the potential for foundation models to overcome limitations of traditional simulation approaches.

In this work, we do not develop ML models, but rather introduce and release a comprehensive, high-quality dataset that captures the complexity of fracture evolution across materials and conditions. Our goal is to enable

<sup>1</sup>National Security Earth Science Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. <sup>2</sup>Energy and Natural Resources Security Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. <sup>3</sup>Statistical Sciences Group, Computer, Computational and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. <sup>4</sup>Materials and Physical Data Group, X Computational Physics Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. <sup>5</sup>Applied Computer Science Group, Computing and Artificial Intelligence Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. <sup>6</sup>High Performance Computing Design Group, High Performance Computing Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. <sup>7</sup>These authors contributed equally: Ryley G. Hill, Kai Gao. ✉e-mail: [rghill@lanl.gov](mailto:rghill@lanl.gov)



**Fig. 1** The data is divided into two distinct approaches – phase-field and FDEM (i.e. HOSS). Phase-field is composed of both uniaxial and biaxial extension boundary conditions while FDEM only includes uniaxial extension. The phase-field dataset includes five materials (PBX, shale, tungsten, aluminum and steel), while the FDEM dataset covers three materials (PBX, shale and tungsten). A small subset of data was made with randomly oriented initial fractures for the phase-field data (10,000) resulting from 1000 fractures  $\times$  the 2 boundary conditions  $\times$  the 5 materials. A full table of the material properties are presented in Table 1.

and accelerate research toward generalizable, large-scale ML models capable of inferring fracture patterns at the moment of material failure, and to support a broad range of scientific applications.

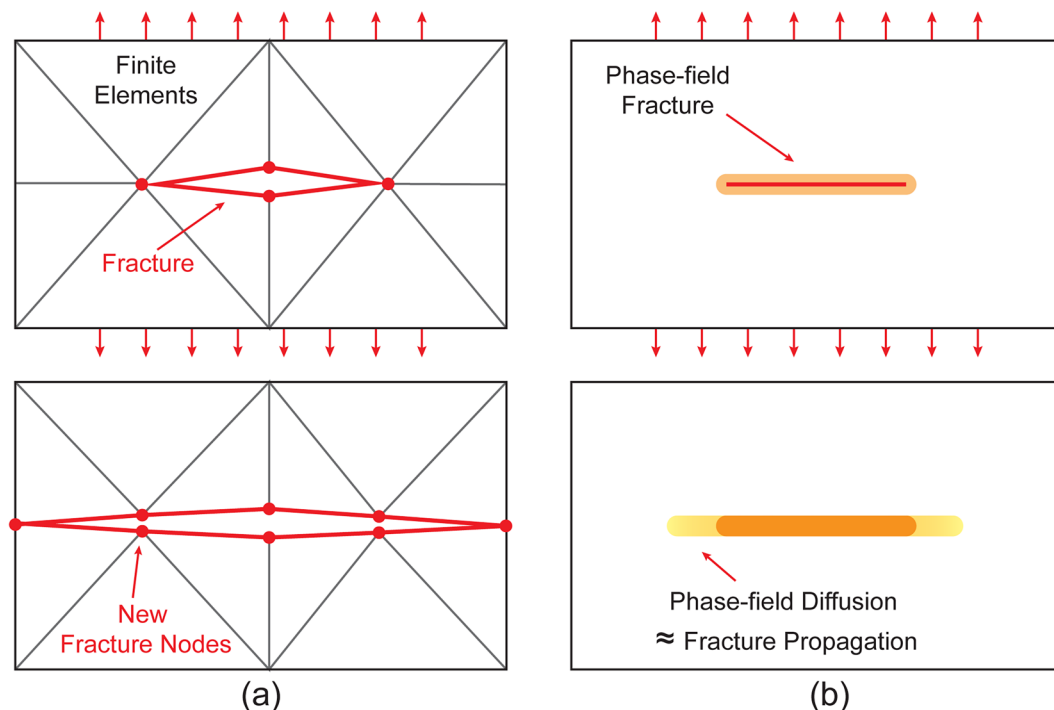
The dataset must include a sufficiently large number of diverse, random initial fracture patterns, as well as a variety of materials. Unfortunately, generating such a fracturing dataset through experiments is highly impractical. Preparing hundreds of thousands of material samples with different random initial fracture patterns is extremely difficult, if not impossible, and even if it were feasible, conducting that many experiments would be prohibitively time-consuming. To the best of our knowledge, no existing dataset in the domain of fracture and material failure provides images of both the random initial fracture patterns and their corresponding fracture patterns at the point of material failure under various boundary conditions.

We leverage computational fracture mechanics and fracturing simulation to present a dataset that will provide the basis to support the training of multi-modal foundational models for predicting material failure. The dataset comprises simulations from two distinct computational approaches that incrementally capture greater physical fidelity as shown in Fig. 1: (1) a fully dynamic phase-field fracture method that approximates fracture propagation as a nonlinear diffusion process of a scalar phase-field variable; and (2) a fully dynamic, combined finite-discrete element method (FDEM) that explicitly simulates fracture initiation, propagation, and arrest under realistic loading conditions. Elements separate along predefined interfaces once failure criteria are met.

Comparatively, the two methods differ in their treatment of discontinuities, computational efficiency, physical fidelity, and fracture resolution. Phase-field simulations are computationally more efficient and suited for large-scale representation. In contrast, the high-resolution modeling of failure mechanics from FDEM are offset by a higher computational cost. In unison, they create a rich multi-resolution dataset that encompasses a comprehensive and complementary acuity for material failure.

This dataset not only spans a wide spectrum of physical complexity but also encompasses diverse data representations, including Cartesian grids and unstructured meshes, thus enabling versatile multi-modal model training. It supports various predictive tasks essential to fracture analysis, such as forecasting fracture patterns and time-to-failure predictions for different materials under varied loading conditions. Additionally, the dataset contains temporal data capturing fracture propagation over time. The dataset provides a robust foundation for exploring how large-scale models can generalize fracture predictions across multiple scales, materials, and loading scenarios, thus opening pathways for broader scientific impact beyond traditional, scenario-specific modeling approaches. By making the curated, multi-modal dataset publicly available, we aim to support broader efforts in data-driven modeling of fracture processes and material degradation. A non-exhaustive list of specific use cases of the dataset:

1. Training ML models to predict fracture evolution: The dataset includes temporal data capturing fracture propagation over time across multiple materials, enabling supervised learning for fracture growth prediction under extensional loading conditions.
2. Studying material-specific fracture behavior: With simulations spanning both brittle (e.g., PBX, shale) and ductile (e.g., tungsten, steel, and aluminum) materials, the dataset allows for systematic comparisons of failure mechanisms across material classes.
3. Investigating the influence of pre-existing fractures: The randomized initial fracture configurations enable studies of how fracture orientation, density, and connectivity affect the resulting failure patterns.



**Fig. 2** A schematic plot comparing the two fracturing physics models used in this study. The FDEM method (a) shows the fracture as a discrete discontinuous displacement between elements that has the ability to propagate based on mode I (shown here) and mode II behavior. It captures cohesive contacts between elements with a much higher fidelity and requires explicit stepping of the fracture evolution. In the phase-field approach (b) fractures are represented as a smooth, diffuse damage zone, where the material gradually transitions from intact to fully dislocated constrained by characteristic width  $2w_0$ . The continuous formulation allows for implicit solutions to the fracture propagation.

4. Developing surrogate models for high-throughput materials screening: The dataset provides a foundation for reduced-order models that can approximate fracture outcomes rapidly, which is valuable in materials design and optimization contexts.
5. Evaluating time-to-failure prediction algorithms: Temporal simulation data can support development of models that estimate when material failure occurs based on evolving internal stress or fracture states.

The rest of the paper is organized as follows: In Section 2, we briefly describe the two numerical methods for generating the fracturing simulation datasets as well as the method for generating the random initial fracture patterns. In Section 3, we describe the mechanical and fracture properties of the used materials, as well as display example fracturing simulation results generated by the numerical solvers. In Section 4, we discuss the data records. Then in Section 5, details on the technical validation are provided. In Section 6, we give information on how to use the dataset, including instructions on how to download and manipulate the dataset records. Section 7 describes code availability.

## Methodology

We use two different fracture physics models to generate the fracturing simulation dataset: phase-field and FDEM methods. The fundamental physics assumptions, fidelity, as well as numerical flexibility of these two models differ from one another, but complement each other, enriching the fracturing simulation dataset.

Figure 2 displays a schematic comparing the fracture patterns in the simplest possible case for the two fracturing physics models used in this study (horizontal fracture propagation due to uniaxial extension). In the FDEM approach, fractures are modeled as discrete separations between elements that evolve explicitly under mode I and II loading. In contrast, the phase-field method treats fractures as diffuse damage zones that transition smoothly from intact to failed material over a characteristic width  $2w_0$ , enabling implicit fracture propagation. In the following, we briefly review the core methodologies of these two physics models.

**Phase-field method for fracturing simulation.** The phase-field method for fracture simulation approximates the evolution of material fractures with a nonlinear diffusion process of a continuous phase-field<sup>20–22</sup>. The most notable advantage of the phase-field method is its flexibility in simulating complex fracture propagation without the need of predefining or adaptively refining complex unstructured meshes along fracture propagation paths, which are essentially unknown beforehand.

In the phase-field fracture propagation theory, the elastic constitutive relation of the background material reads<sup>22,23</sup>:

$$\sigma(\mathbf{u}) = [(1 - \eta)(1 - \psi)^2 + \xi] \mathbf{C} : \varepsilon, \quad (1)$$

where  $\sigma$  is the stress tensor,  $\mathbf{u}$  is displacement,  $\mathbf{C} = \mathbf{C}(\mathbf{x}) = C_{ijkl}(\mathbf{x})$  is the fourth-order elasticity tensor of the matrix with up to 21 independent components;  $\varepsilon = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$  is the strain tensor;  $\xi$  is used to prevent stiffness from reaching 0 during full fracture,  $\eta$  is a small positive number to avoid instability at the fracture  $\psi$ . The Neumann boundary condition for  $\sigma$  is  $\sigma \cdot \mathbf{n} = \mathbf{t}_N$ , where  $\mathbf{n}$  is the normal vector of the boundary  $\partial\Omega_N$ ,  $\Omega$  is the computational domain, and  $\mathbf{t}$  is a traction vector.

In this work, we simulate dynamic rather than quasi-static fracture propagation to better capture the fracture evolution under dynamic boundary conditions (e.g., constant-velocity tension forces applied at boundaries). Thus, rather than solving the static elastic deformation equation in Eq. (1), we consider the dynamic elastic constitutive relation in a degraded medium:

$$\rho \partial_t^2 \mathbf{u} = \nabla \cdot \sigma + \mathbf{f}, \quad (2)$$

where  $\sigma$  follows Eq. (1),  $\mathbf{f} = \mathbf{f}(\mathbf{x}, t)$  is the dynamic external force term, and  $\rho$  is the mass density of the background medium. This is effectively the elastic-wave propagation equation in a degraded medium. The evolution of the fracture phase-field  $\psi$  is driven by the so-called positive elastic strain energy. Based on the principle of variation, the governing equation for the fracture evolution can be written as<sup>22,23</sup>:

$$G_c (w_0^{-1} \psi - w_0 \nabla^2 \psi) = 2(1 - \psi) H^+(\varepsilon, t), \quad (3)$$

where  $2w_0$  is the characteristic width of a fracture and  $G_c$  is the critical energy release rate;  $H^+(\varepsilon, t)$  is a history energy field defined as the maximum positive elastic energy at  $\mathbf{x}$  up to  $t$ :

$$H^+(\varepsilon, t) = \max_{\tau \in [0, t]} \Phi^+(\varepsilon, \tau). \quad (4)$$

While there are a number of choices to define the positive elastic strain energy function  $\Phi^+$ , here we choose one based on the elastic tensor  $\mathbf{C}$  in a general form for computing the following elastic energy:

$$\Phi^+(\varepsilon, t) = \frac{1}{2} (\varepsilon^+)^T \mathbf{C} \varepsilon^+, \quad (5)$$

where the tensile-positive elastic strain tensor reads

$$\varepsilon^+ = [\varepsilon_{xx}^+, \varepsilon_{yy}^+, \varepsilon_{zz}^+, 2\varepsilon_{yz}, 2\varepsilon_{xz}, 2\varepsilon_{xy}]^T, \quad (6)$$

and the tensile-positive strain tensor being

$$\varepsilon_{ij}^+ = \frac{1}{2} (\varepsilon_{ij} + |\varepsilon_{ij}|), \quad (7)$$

with  $\varepsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$  being the strain tensor. The above definition allows us to define the history energy function for both isotropic and anisotropic materials in a unified formulation. The resulting coupled dynamic fracture evolution system reads:

$$\rho \partial_t^2 \mathbf{u} - \nabla \cdot \sigma = \mathbf{f}, \quad (8)$$

$$\sigma(\mathbf{u}) = [(1 - \eta)(1 - \psi)^2 + \eta] \mathbf{C} \varepsilon, \quad (9)$$

$$\varepsilon = \frac{1}{2} (\nabla \mathbf{u} + \nabla \mathbf{u}^T), \quad (10)$$

$$G_c (w_0^{-1} \psi - w_0 \nabla^2 \psi) = 2(1 - \psi) H^+(\varepsilon, t). \quad (11)$$

This system of coupled partial differential equations requires solving two partial differential equations: one is the linear elastic wave equation in a degraded medium, and the other is the nonlinear phase-field diffusion equation. We solve them using the lowest-order spectral-element method and Newmark-beta implicit scheme. To capture fracture propagation in ductile, elastoplastic materials such as metals (e.g., tungsten, aluminum, steel, etc.), we incorporate the von Mises yield criterion<sup>24</sup> into the constitutive relation of the background medium. This involves modifying the effective elastic tensor to reflect the reduction in stiffness associated with plastic yielding and accumulated damage.

**Combined finite discrete-element method for fracturing simulation.** Researchers at Los Alamos National Laboratory have developed the Hybrid Optimization Software Suite (HOSS)<sup>25</sup>, based on the FDEM, to simulate complex mechanical behaviors of solid materials, including fracturing and fragmentation<sup>26</sup>. HOSS is widely used due to its robustness, advanced algorithms, massive parallelization features, and versatile capabilities.

HOSS results have been validated by numerous independent experiments<sup>25</sup>. The non-exhaustive list of HOSS applications include the following: experimental rock mechanics (e.g. triaxial uniaxial), well-bore and drilling stability, building integrity, seismic studies, high velocity impacts, thermal-hydraulic-mechanic behavior, high explosive performance, tissue and cell experiments, and weapons penetration<sup>25,27–32</sup>.

For material fracturing, the discrete system that solves the coupled elastic/ductile deformation and fracture propagation can be written as

$$\mathbf{M} \frac{d^2 \mathbf{x}}{dt^2} + \mathbf{C} \frac{d \mathbf{x}}{dt} = \mathbf{F}, \quad (12)$$

where  $\mathbf{M}$  is the mass matrix,  $\mathbf{C}$  is the damping matrix,  $\mathbf{F}$  is the equivalent nodal force, including all forces distributed to the node, such as body forces, forces due to material deformation, contact forces between solid bodies, and cohesion forces in damage<sup>26,27,33,34</sup>;  $\mathbf{x}$  is the vector enclosing the node displacement field<sup>35</sup>.

In this study the numerical domain was discretized using triangular finite elements, with their rotation, deformation, and stress evaluation through a multiplicative decomposition framework<sup>29,35</sup>. To identify contact interactions between elements, a contact detection algorithm with linear computational and memory complexity was employed<sup>36</sup>. A potential field-based penalty method was applied to handle the frictional contact interaction in the contact zones<sup>37</sup>. An explicit central difference time integration scheme was used to obtain the temporal evolution of the system<sup>38</sup>.

In this work, a generalized traction-separation model is used to simulate the damage in the material. Specifically, the traction is defined from the gradient of the interfacial potential which is a function of the components of the separation vector<sup>27</sup>. In a simplified form, the traction is defined as:

$$\mathbf{p} = (f_t \mathbf{n} + f_s \mathbf{t})z(d) \quad (13)$$

where  $f_t$  and  $f_s$  are the tensile and shear strengths, while  $\mathbf{n}$  and  $\mathbf{t}$  are the normal and tangential directions. The function  $z(d)$  characterizes the shape of the traction-separation curve and is expressed as a function of the damage variable  $d$ . Damage variable  $d$  takes a value between 0 and 1.0 and is given by:

$$d = \sqrt{d_n^2 + d_t^2} \quad (14)$$

where

$$d_n = \frac{\delta_n}{\delta_{nc}}; d_t = \frac{\delta_t}{\delta_{tc}} \quad (15)$$

are the normal and tangential damage components. The  $\delta_n$  and  $\delta_t$  are the normal and tangential components of the separation while  $\delta_{nc}$  and  $\delta_{tc}$  are the maximum tensile and shear separations.

The fracture release energies in mode I and mode II are given by:

$$G_I = f_t \delta_{nc} \int_0^1 z(d) \quad (16)$$

$$G_{II} = f_s \delta_{tc} \int_0^1 z(d) \quad (17)$$

Note that damage  $d = \frac{\delta}{\delta_c} \leq 1.0$  will evolve from  $\delta \rightarrow \delta_c$  where  $\delta_c$  is a critical value at which  $f(\delta_c) = 0$  whereby the cohesive resistance vanishes leading to complete decohesion and possible fragmentation, capturing crack propagation and the transition from continuous to fully discrete material behavior.

In addition, we employ in-house state-of-the-art discretized contact solutions to handle contact detection and contact interaction<sup>25,28</sup>. The contact detection follows the MRCK scheme<sup>39,40</sup> where the physical space (two-dimensional here) is uniformly partitioned into square cells of size  $h$ . Each target sample (points along the element boundaries) is mapped to a cell  $(i, j)$  with  $i = \frac{x-x_0}{h}$  and  $j = \frac{y-y_0}{h}$  and stored in a single MR-linear sort reduced to determine whether the contactor and the target share at least one square cell. In the two-dimensional implementation here, a candidate pair is accepted if a target point lies within the contactor triangle. The triangle-to-point contact interaction approach is used such that target triangles are discretized into boundary points that act as Gauss integration points for evaluation distributed contact forces. The MRCK contact search algorithm identifies contact pairs between contactor triangles and target points, with interaction computed only when a target point lies within a contactor triangle.

Material failure analysis requires careful consideration of key indicators to accurately determine the point of failure. In addition, the data must be processed in a way that captures critical trends while maintaining consistency between the initial setup and the final fracture domains. Therefore, two important data processing steps must take place to ensure that there is (1) a systematic choice of the time-to-failure, and (2) a consistent data structure between the initial and final time steps. To resolve these difficulties, simulation results are post-processed by incorporating some of the unique variables tracked by the HOSS simulations which are not necessarily important to the final fracture results. HOSS tracks the entire solid kinetic energy of the system and also the stress tensor for each element. To determine the time-to-failure we use an automated picking routine that follows the following steps: (1) evaluate the total simulation stress and total kinetic energy for each time

step, (2) identify all peaks in the stress and kinetic energy throughout the simulation, (3) find the “key” kinetic energy peak with the steepest drop, which corresponds to failure onset, (4) find the stress peak that first occurs after the “key” kinetic energy peak, which marks the transition where the kinetic energy spikes may indicate failure events, (5) determine the final failure time step by computing the gradient of the stress data and picking where the stress stops decreasing after previously being negative, which represents the moment when the material stops significantly responding to stress likely indicating failure. In this way, each simulation incorporates a systematic choice for the time-to-failure. Through visual inspection we found this method to produce the most consistent times where fractures had nearly or fully propagated across the samples. Material failure can take various forms, but these results provide a conservative estimate of the time-to-failure. In reality, failure may occur earlier if other factors, such as sudden stress drops, are considered.

Since the simulations produce displacement between fractures the initialized domain increases in size vertically as shown in Fig. 9. These results produce an unstructured grid of information that may also be long past the first time-to-failure step. To resolve these difficulties results are post-processed using the PyVista package<sup>41</sup>. HOSS retains unique global nodal identification numbers (IDs) which for our triangular mesh means we have unique edges for the entire simulation. Each edge retains information called “damage” that represents the loss of cohesion between its edge and neighboring triangle edge<sup>27</sup>. It is important to note that within each simulation of the dataset each edge will retain a value between 0.0 and 1.0, but only values equal to 1.0 should be classified as a true fracture. After selecting the time step at which failure occurs in the simulation, we map the damage values from the final and intermediate steps back onto the initialized mesh. This process preserves both the edges of elements that fractured at the earliest failure time and the original grid structure of the domain. We evaluate three of the most unique material end members similarly explored in the phase-field model (i.e. PBX, shale, and tungsten).

**Generation of random initial fracture patterns.** Each fracture simulation begins with a randomly generated initial fracture pattern, which serves as the starting point for subsequent fracture propagation. Our first distribution of data consists of orthogonally aligned initial fracture patterns randomly arranged in a 2D computational domain with a size of 0.25 m in each dimension. In 2D, the key properties of the initial fracture patterns are: location, length, and orientation. In order to generate a sufficiently large dataset that ensures random fracture properties, we adopt a procedure as described below.

First, we draw three sets of values from three distinct uniform distributions:

$$(S_x, S_z, L) \in [\mathcal{U}_\xi(\delta_x, H_x - \delta_x), \mathcal{U}_\zeta(\delta_z, H_z - \delta_z), \mathcal{U}_\eta(L_{\min}, L_{\max})], \quad (18)$$

where  $S_x$  and  $S_z$  represent the starting coordinates of a fracture in  $x$ - and  $y$ -dimensions,  $L$  is the length of a fracture,  $\mathcal{U}(a, b)$  is a uniform random distribution with a lower bound of  $a$  and an upper bound of  $b$ , with  $\xi$ ,  $\zeta$  and  $\eta$  being three different random realization seeds. In addition,  $L_{\min} = 0.01$  m and  $L_{\max} = 0.05$  m represent the minimum and maximum lengths of a fracture, respectively,  $H_x = H_z = 0.25$  m is the domain size, and  $\delta_x = \delta_z = 0.02$  m represent the “buffer” zone of the fractures to avoid having fractures on the boundaries. The coordinates of the termination point for each fracture are calculated relative to the starting coordinates and orientation:

$$(E_x, E_z) = (S_x, S_z) + (L \cos(\theta), L \sin(\theta)), \quad (19)$$

where  $\theta$  is the polar orientation of a fracture. By setting  $\theta = 0$  or  $\pi/2$ , we can get a horizontal or vertical fracture. Furthermore, in each initial fracture pattern the number of horizontal/vertical fractures is also random:

$$(N_h, N_v) \in [\mathcal{U}_h(N_{\min}, N_{\max}), \mathcal{U}_v(N_{\min}, N_{\max})], \quad (20)$$

with  $N_{\min} = 3$  and  $N_{\max} = 15$  representing the minimum and maximum numbers of fractures in a model, respectively. The total number of fractures for different fracture patterns therefore satisfies

$$2N_{\min} \leq N_f \leq 2N_{\max}. \quad (21)$$

In this way, not only the fracture locations vary, but also the number of horizontal/vertical fractures. Using this method, we generated a total of 200,000 orthogonal initial fracture patterns. Figure 3 displays 36 examples, where in each fracture pattern, fractures are colored with different colors for better visual distinction. These 200,000 initial fracture patterns are included in our open-access dataset.

To support future validation and generalization of ML models on unseen fracture samples, we generate 1,000 random non-orthogonal fracture patterns where the fractures are not necessarily orthogonal with one another. In this case, each model contains a single fracture set, and fracture orientations are randomly assigned by drawing from a uniform distribution:

$$\theta \in \mathcal{U}_\theta\left(0, \frac{\pi}{2}\right). \quad (22)$$

We display 36 examples of the non-orthogonal fracture patterns in Fig. 4.



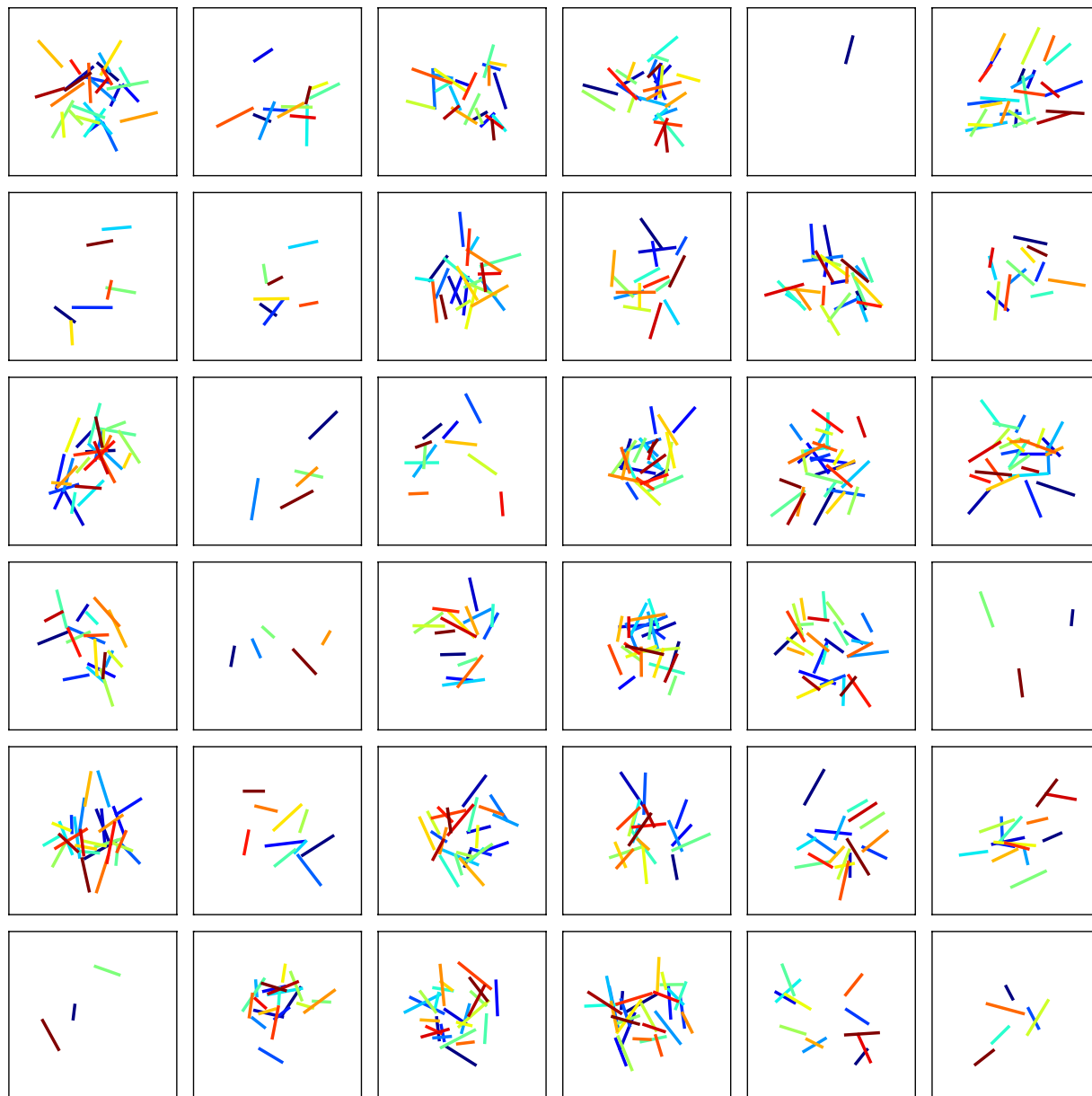
**Fig. 3** Examples of randomly generated initial fracture patterns with orthogonal orientations used in fracturing simulations. Colors in the plots represent distinct fractures. The size of the simulated domain is  $H_x = H_z = 0.25$  m (omitted for brevity).

**Fracture Simulation Results.** In the following subsections, we describe the properties of the materials that we used to create the datasets and highlight the material dependent differences in fracture propagation. We also show a handful of examples of the simulated fracture patterns using the phase-field and HOSS simulators.

**Material properties.** We perform fracturing simulations for five materials: PBX, shale rock, tungsten, steel, and aluminum. PBX and shale are simulated as elastic materials, while the other three as ductile elastoplastic materials.

The PBX is representative of a brittle elastic fracture behavior common to many geo-materials like rocks and concrete. This material exhibits brittle behavior, with some localized damage before a full fracture which, at the strain rates applied by our boundary conditions, implies dynamic failure.

The shale is representative of a brittle elastic fracture behavior that incorporates anisotropy – a key characteristic in geomechanics and reservoir engineering. It is the only material which anisotropy is accounted for. The mechanically anisotropic bedding planes produced by the clay-rich shale causes resistance to fracture perpendicular to the bedding and lower resistance to fracture parallel to the bedding. This anisotropy allows for fractures to propagate slowly, and in greater abundance, over time before sudden failure. The elasticity of the shale is specified with elasticity constants in the Voigt notation (i.e.,  $C_{ij}$ ) rather than its original form  $c_{ijkl}$ . Since,



**Fig. 4** Examples of randomly generated initial fracture patterns with non-orthogonal orientations used in fracturing simulations. Colors in the plots represent distinct fractures. The size of the simulated domain is  $H_x = H_z = 0.25$  m (omitted for brevity).

we only simulate fracturing in a 2D plane ( $x - z$  plane or equivalently  $x_1 - x_3$  plane), we only display the relevant elasticity parameters on this plane (i.e.,  $C_{11}$ ,  $C_{13}$ ,  $C_{33}$ , and  $C_{55}$ ).

Tungsten is a high density metal that is representative of plastic deformation behavior important to many industrial alloys and other metals. For the phase-field-based fracturing simulations, we set the Griffith critical energy release rate ( $G_c$ ) of tungsten to  $500 \text{ J/m}^2$ , which is a slightly higher than usual value for tungsten at room temperature. For FDEM we use both modes: mode I fracture energy ( $G_I$ ), and mode II ( $G_{II}$ ) fracture energy. We simulated fracturing for tungsten at elevated temperatures.

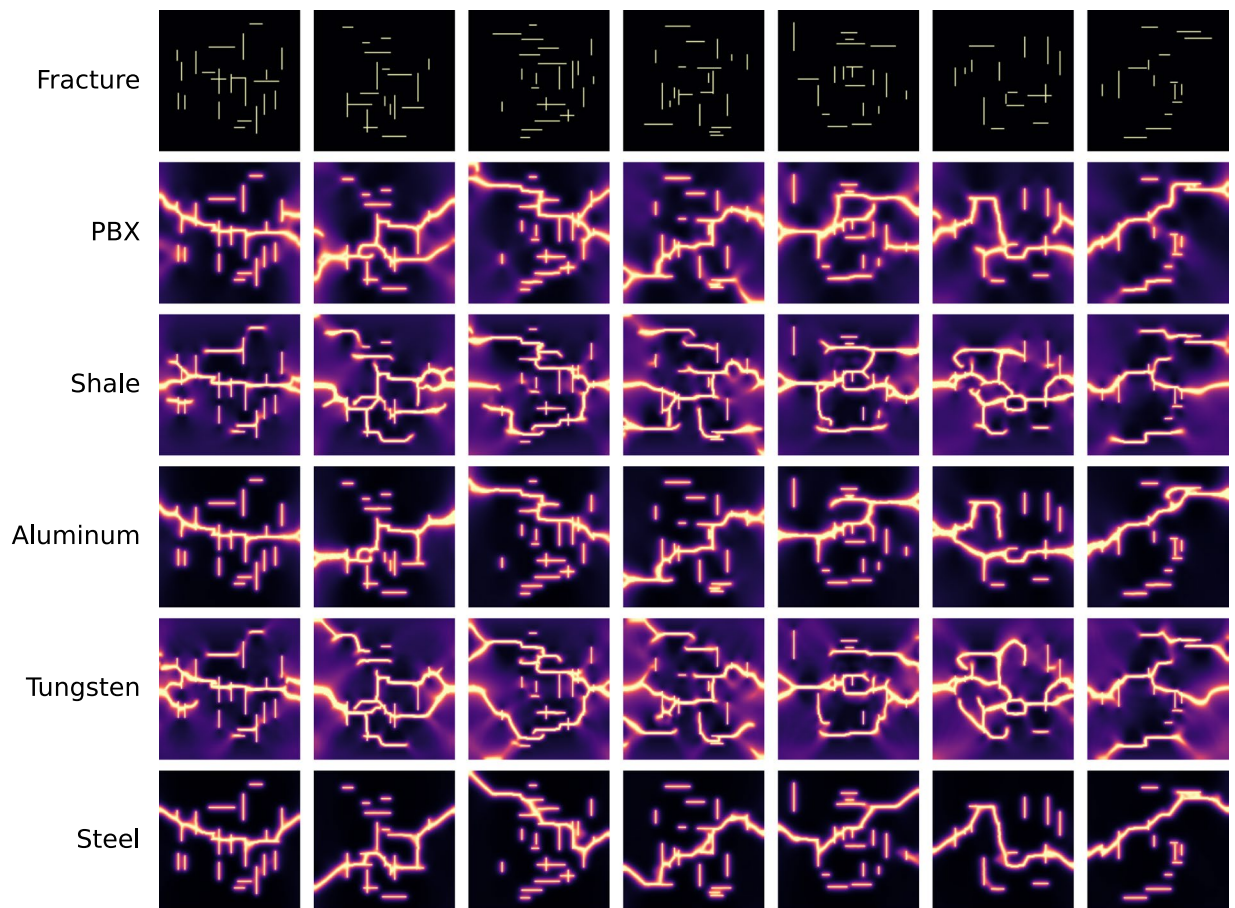
The aluminum we use here is a type of hardened, high-strength aluminum alloy rather than pure aluminum. Therefore, its Poisson's ratio (0.25) is notably smaller than that of the pure aluminum (typically 0.31 to 0.34). Henceforth, we continue to use the term "aluminum" to represent this aluminum alloy for brevity.

We summarize the elasticity, ductility, and fracture-related parameters of the five materials in Table 1.

*Phase-field-based fracturing simulation results.* With the generated random fracture patterns and the mechanical properties of the five materials showed above, we performed 200,000 fracturing simulations using our in-house solver. We discretize the random fracture patterns using  $127 \times 127$  regular first-order spectral

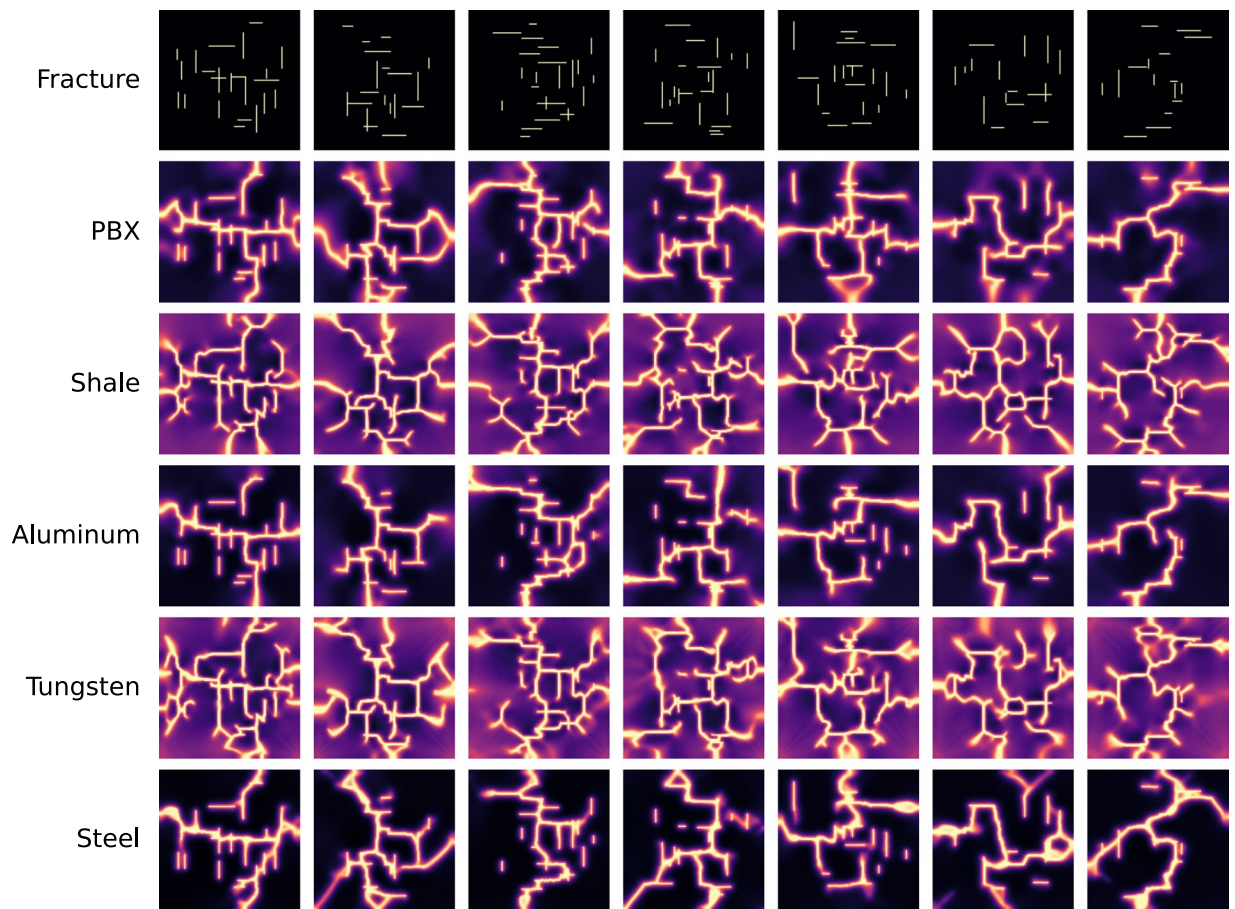
Property	PBX	Shale	Tungsten	Aluminum	Steel
Density (kg/m <sup>3</sup> )	$1.82 \times 10^3$	$2.075 \times 10^3$	$19.25 \times 10^3$	$2.7 \times 10^3$	$7.85 \times 10^3$
Young's Modulus (Pa)	$1.0 \times 10^{10}$	—	$4.0 \times 10^{11}$	$6.49 \times 10^{10}$	$2 \times 10^{11}$
Poisson's Ratio	0.36	—	0.28	0.25	0.3
$C_{11}$ (Pa)	—	$3.125738 \times 10^{10}$	—	—	—
$C_{13}$ (Pa)	—	$3.399087 \times 10^9$	—	—	—
$C_{33}$ (Pa)	—	$2.248732 \times 10^{10}$	—	—	—
$C_{55}$ (Pa)	—	$6.486048 \times 10^9$	—	—	—
Yield Strength (Pa)	—	—	$7.5 \times 10^8$	$2.5 \times 10^8$	$6 \times 10^8$
Hardening Modulus (Pa)	—	—	$5.0 \times 10^9$	$2.5 \times 10^8$	$2.5 \times 10^9$
<sup>†</sup> Critical Energy Release Rate (J/m <sup>2</sup> )	640.885	50	500	$1 \times 10^4$	$2.5 \times 10^5$
<sup>†</sup> Characteristic Width (m)	0.004	0.002	0.002	0.004	0.004
*Coulomb Friction	0.6	0.6	0.6	—	—
*Tensile Strength (Pa)	$0.5 \times 10^6$	$1.0 \times 10^7$	$8.5 \times 10^8$	—	—
*Shear Strength (Pa)	$2.0 \times 10^6$	$4.0 \times 10^7$	$4.905 \times 10^8$	—	—
* $G_I$ (J/m <sup>2</sup> )	125	100	12750	—	—
* $G_{II}$ (J/m <sup>2</sup> )	500	400	7358	—	—
*Characteristic Width (m)	0.003	0.003	0.003	—	—

**Table 1.** Properties for the five materials for fracturing simulation. Properties marked with <sup>†</sup> are exclusive to phase-field simulations and those marked with \* are exclusive to HOSS.



**Fig. 5** Examples of phase-field simulations for five materials with randomly generated orthogonal fractures subjected to uniaxial boundary conditions.

elements<sup>42</sup>, resulting in a total of  $128 \times 128$  degrees of freedom for the phase-field equation, Eq. (3), and a total of  $2 \times 128 \times 128$  degrees of freedom for the elastic constitutive equation, Eq. (2) in the 2D case.



**Fig. 6** Examples of phase-field simulations for five materials with randomly generated orthogonal fractures subjected to biaxial boundary conditions.

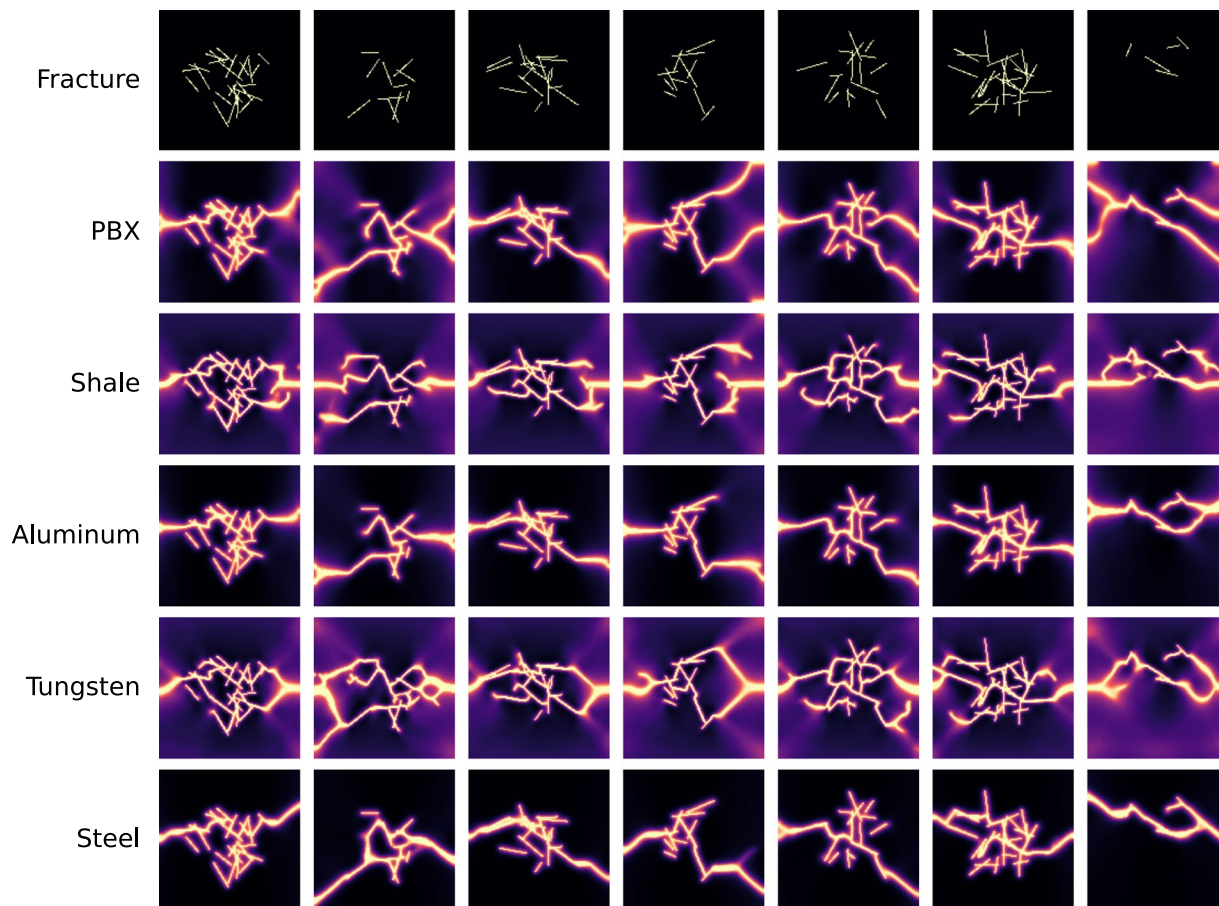
For each initial fracture pattern and material, we perform two simulations under different boundary conditions: In the first one we apply uniaxial extension (vertical tension force) with a constant velocity of 1 m/s at the top and bottom boundaries (i.e.,  $v_{\text{top}} = -1$  m/s,  $v_{\text{bottom}} = 1$  m/s, assuming that the  $z$ -axis points from the bottom to the top); in the second one we apply biaxial extension on all the boundaries of a sample (i.e.,  $v_{\text{top}} = v_{\text{left}} = -1$  m/s,  $v_{\text{bottom}} = v_{\text{right}} = 1$  m/s, assuming that the  $x$ -axis points from the left to the right).

Figure 5 displays seven examples of fracturing simulations under uniaxial extension for each of the five materials. The top row shows the randomly generated initial fracture patterns, while each subsequent row illustrates the final fracture failure patterns for a specific material based on those initial conditions. The rows show how the mechanical properties of different materials influence their failure behavior. The columns emphasize that, for a given material, the initial fracture geometry strongly dictates the final fracture outcome at the point of failure.

By comparison, Fig. 6 shows seven examples of fracturing simulation under biaxial extension for the five materials. In this case, we observe fractures growing in both directions (vertical and horizontal), resulting in more complex fracture patterns than the ones observed for the uniaxial boundary conditions. Under vertical tension (uniaxial), fractures mostly grow horizontally or diagonally, while fractures growing vertically are rare. Since, fracturing is a complex mechanical process that involves multiple closely coupled factors, it is challenging to isolate the sole factor that leads to some of the failure pattern similarities between the different materials.

We also perform fracturing simulations for random non-orthogonal fracture patterns using uniaxial and biaxial boundary conditions. We display examples for non-orthogonal fractures in Fig. 7 with uniaxial extension and in Fig. 8 with biaxial extension. Because of the notable differences in the initial fracture patterns compared with those in Figs. 5,6, the final material failure fracture patterns also differ significantly. For pre-existing fracture patterns with low connectivity (e.g., the fourth and the seventh columns in Fig. 7), different materials exhibit notably different failure patterns.

**HOSS fracturing simulation results.** For the HOSS fracturing simulations, we use the same set of random orthogonal initial fracture patterns generated for the phase-field method simulations, as shown in Fig. 3. Since, these simulations are much more expensive than the phase-field ones, we consider only three materials (PBX, shale and tungsten) and one type of boundary conditions (uniaxial extension). These three materials were chosen to cover the unique range of material types discussed in section 6- quasi-brittle (PBX), anisotropic brittle (shale), and plastic deformation (tungsten). To accurately impose the boundary conditions (tension), we confine



**Fig. 7** Examples of phase-field simulations for five materials with randomly generated non-orthogonal fractures under uniaxial boundary conditions.

the 0.25 m by 0.25 m domain by 5-mm plates the top and bottom, as indicated by Fig. 9. We assume free boundary conditions for the sides of the model and use a linear ramping (0 to  $10^{-4}$  s) nodal velocity boundary conditions for the plates that are  $\pm 1$  m/s (tension). The ramping allows the boundary conditions to slowly approach the  $\pm 1$  m/s to avoid erroneous shearing at the boundaries of the material.

Different from the phase-field solver used in this work, HOSS is a solver that uses an explicit time stepping scheme during simulation limited by the critical time step  $t_c$  which is a function of the smallest element size  $h$  and wave speed  $c$  ( $t_c \propto h/c$ ). This condition ensures that waves do not propagate across more than one element per time step, maintaining numerical stability. To ensure numerical stability, we use a fixed time increment of  $10^{-8}$  s. The mesh and element sizes are relatively fixed and do not vary during the simulation, such that our chosen critical time represents the largest stable time step that can be used throughout the analysis. As we cannot anticipate the exact time-to-failure of the material during the simulation, we allow each simulation to run for a specified wall-clock time of eight hours using eight MPI domains and CPU cores, which is the typical duration needed to achieve full failure.

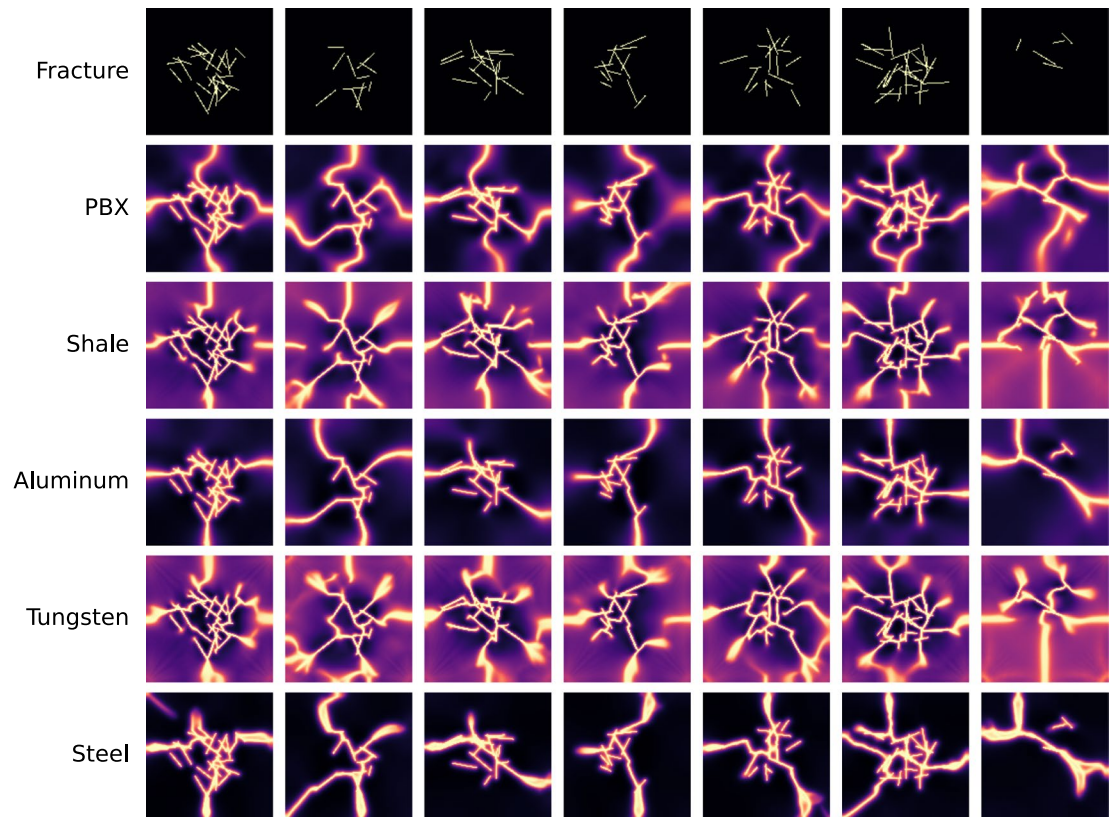
Figure 9 displays an example of fracturing in PBX produced by our HOSS solver over six snapshots. It could be seen that the propagation of pre-existing fractures initiates from the lower-left part where two fractures intersect, and propagates horizontally over the sample until the sample fails. Although there are other fractures that propagate before failure, this initially propagated fracture in the lower half of the model dominates the fracturing process.

An example of the initialized fracture patterns and final results of the three materials PBX, shale, and tungsten is shown in Fig. 10.

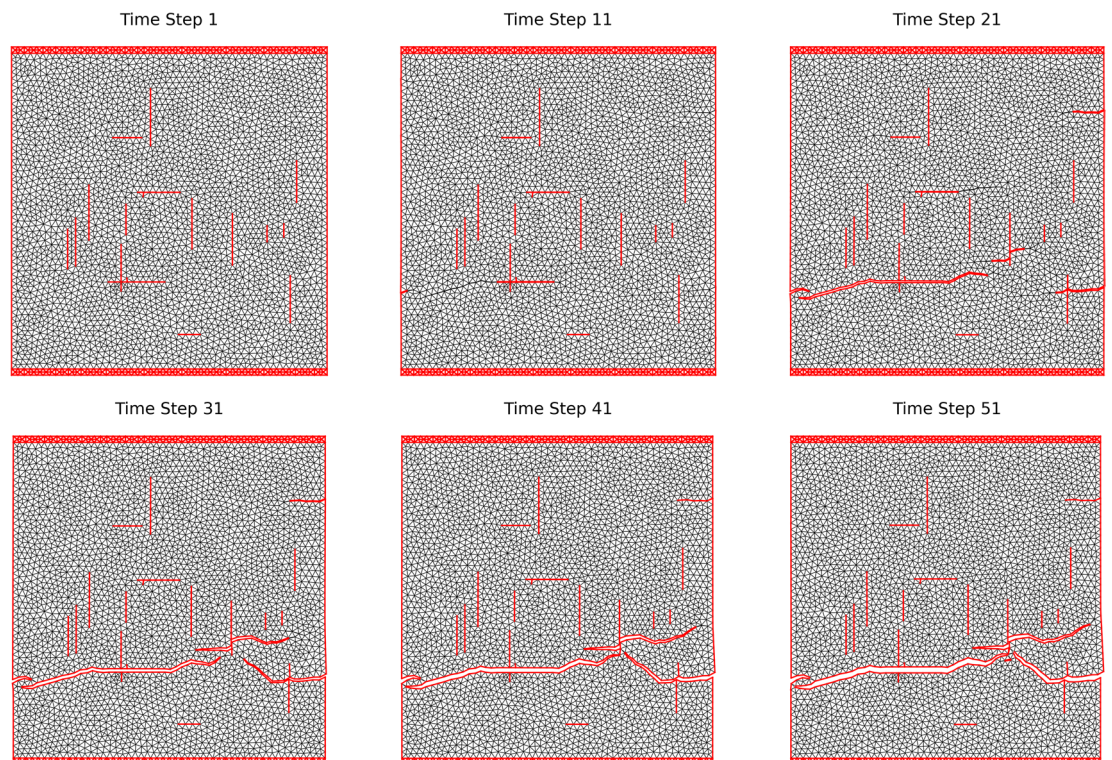
### Data Records

The dataset supporting this study has been made available on Hugging Face Datasets [https://huggingface.co/datasets/smartFRACs/material\\_fracturing](https://huggingface.co/datasets/smartFRACs/material_fracturing)<sup>43</sup>.

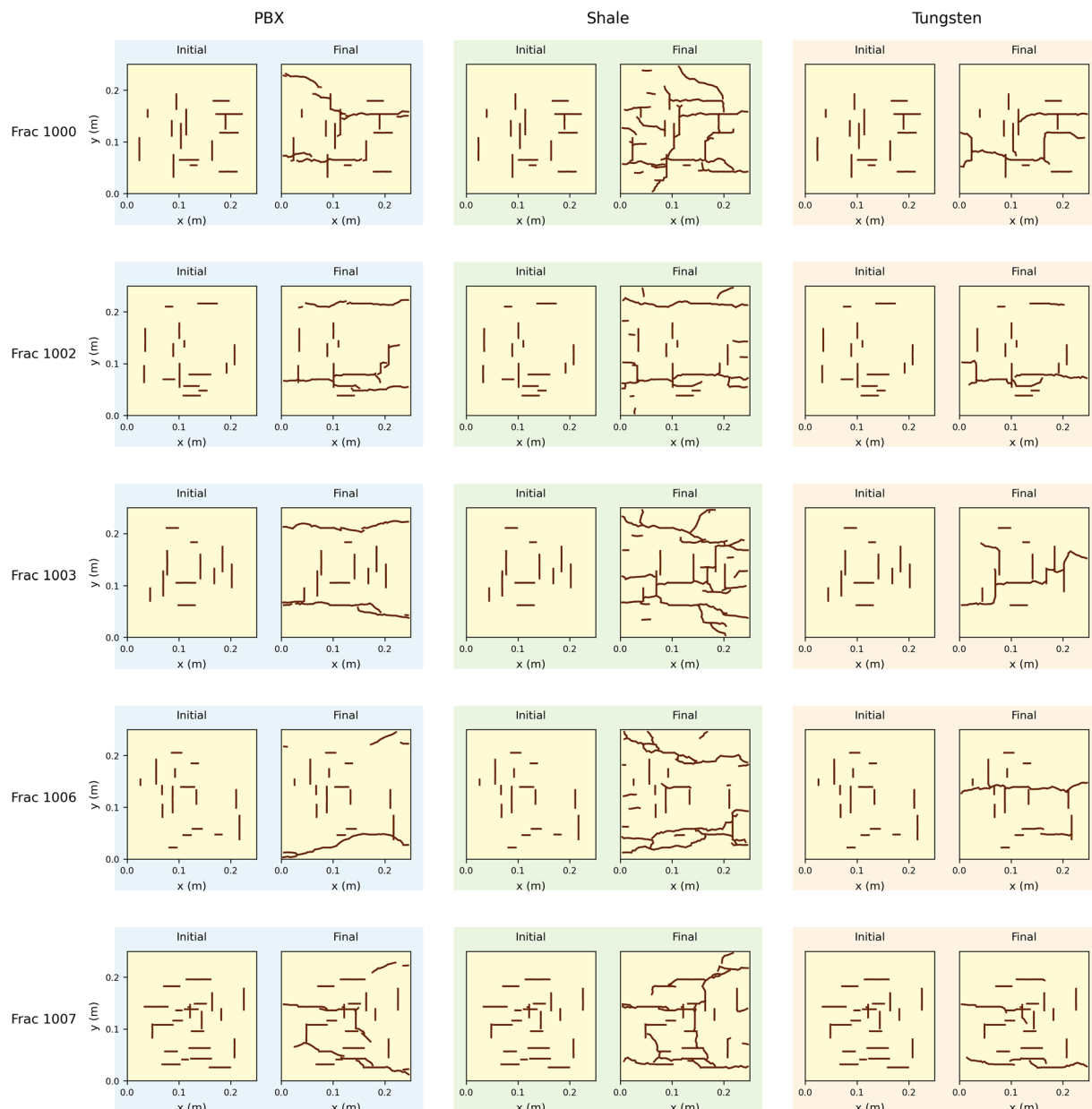
The full dataset is split into two primary folders: HOSS and PHASE-FIELD, corresponding to the two numerical solvers. Within each of these folders, data is further categorized by material type and boundary condition (phase-field only). The simulation outputs are stored as compressed .tar.gz archives, each containing approximately 1000 .h5 files representing individual fracture simulations. Additional metadata, file descriptions, and usage instructions are provided in the accompanying README file at the dataset URL.



**Fig. 8** Examples of phase-field simulations for five materials with randomly generated non-orthogonal fractures under biaxial boundary conditions.



**Fig. 9** Single HOSS fracture example at different time steps for PBX. Note, time-to-failure could be earlier in the simulation but requires post-processing from the final fracture pattern.



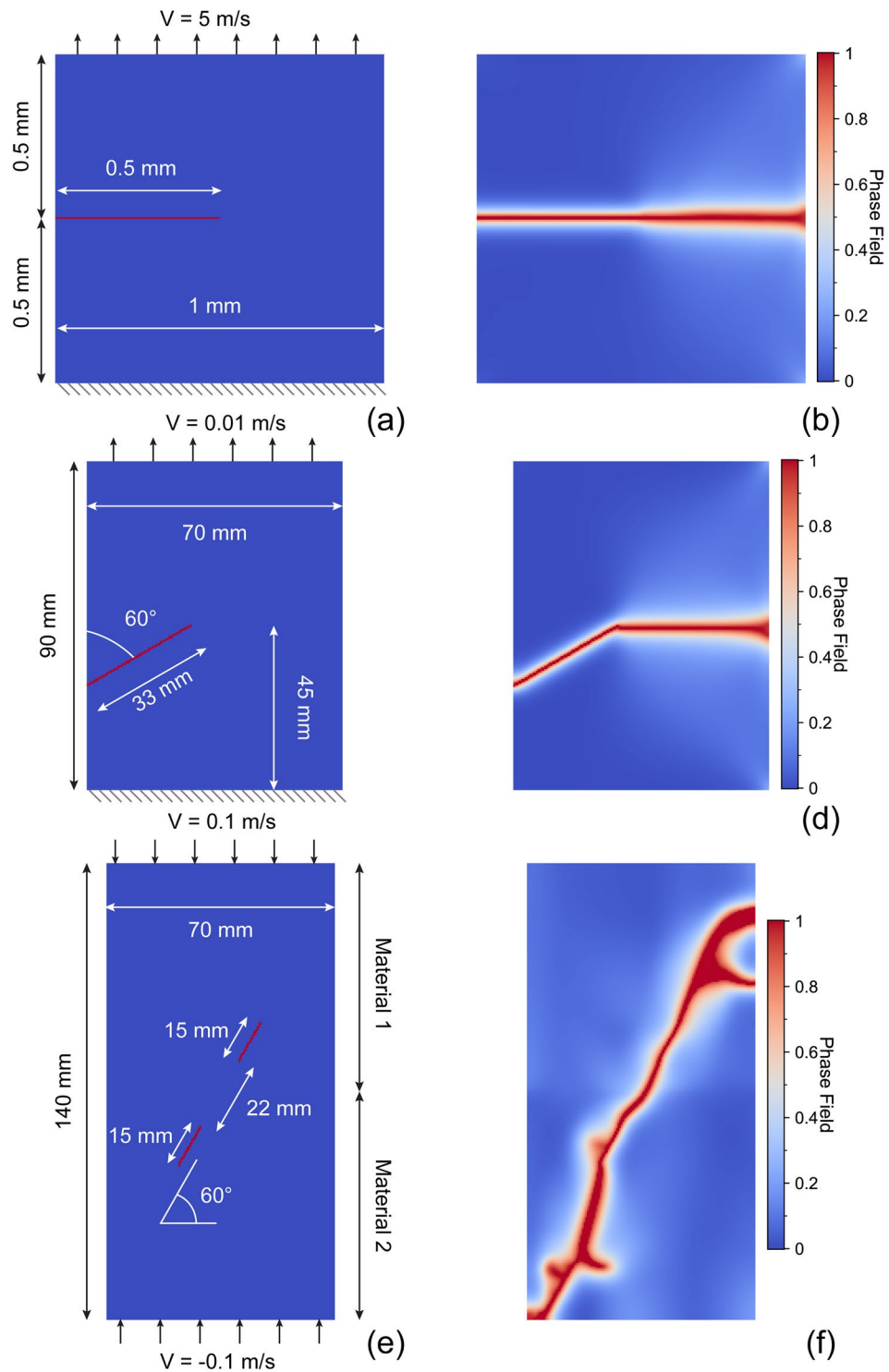
**Fig. 10** Multiple examples of fracturing simulations in HOSS for the three materials tested in that portion of the dataset with initial orthogonal fracture patterns and final fracture pattern at time of failure.

### Technical Validation

To establish confidence in the numerical methods implemented here, we outline in the following subsections how each simulation approach was validated. To validate our implementation of the phase-field method – developed specifically for generating this dataset – we compare it against a set of benchmark crack-propagation problems. We ensure consistency and correctness of fracture evolution to common examples in the literature. For the FDEM simulations, which are resolved from the more mature and similarly validated HOSS software, we highlight common issues encountered during large-scale runs and the corrective measures applied to ensure result integrity.

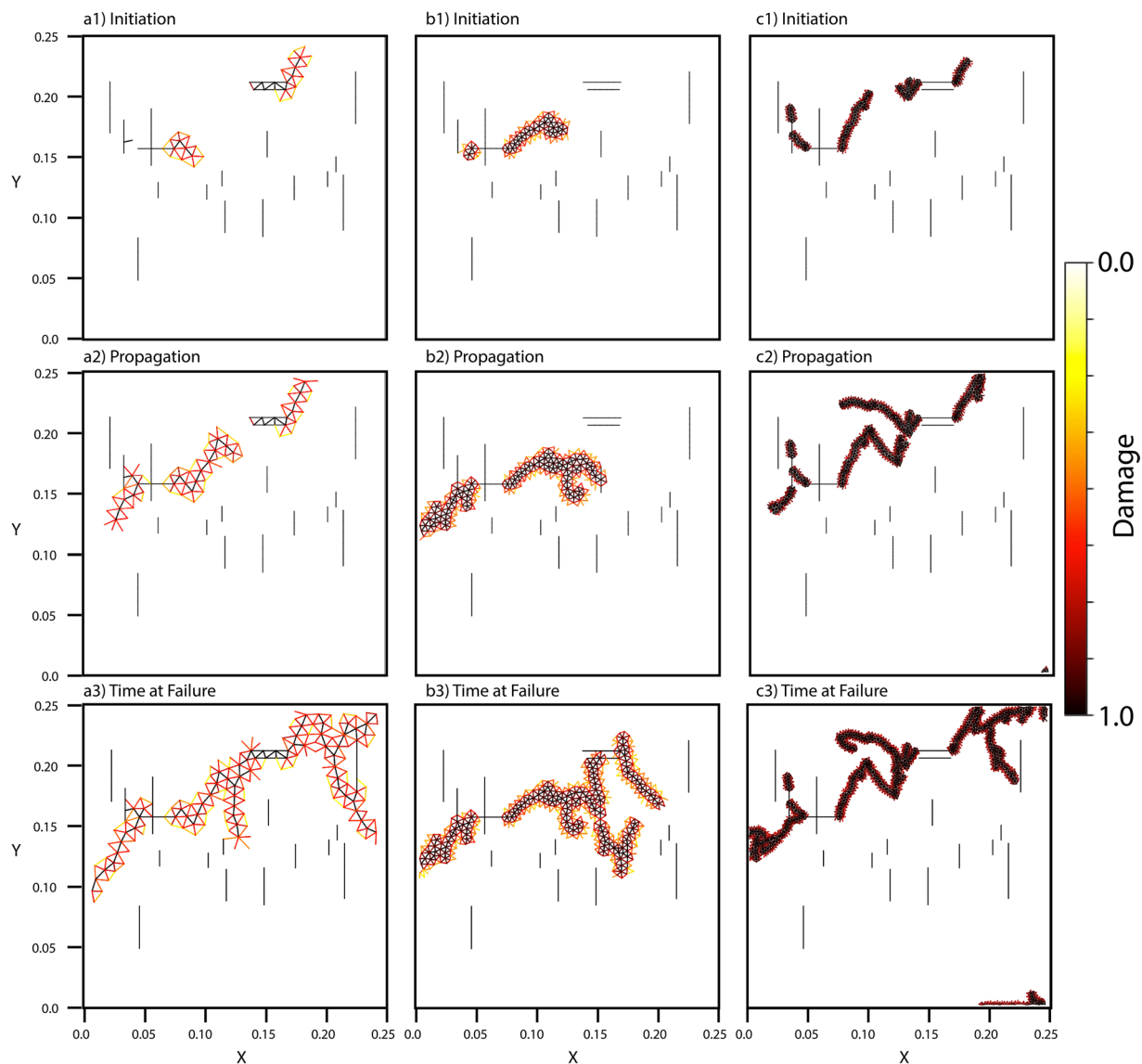
**Phase-field method.** Phase-field method has been widely used in the last decade for simulating crack and fracture propagation due to its flexibility and accuracy. Here, we perform three simulations in three samples of different sizes and material properties to verify the efficacy and accuracy of our phase-field solver.

In the first example, we simulate the evolution of a boundary crack in a tiny, square sample under tensile boundary conditions. This test example aims to replicate the numerical experiment in<sup>44</sup>. Figure 11a displays the configuration of the experiment. The size of the sample is 1 mm in both dimensions. The sample is discretized with  $200 \times 200$  regular rectangle spectral elements with a grid size of  $5 \times 10^{-6}$  m. The characteristic fracture width is  $l_0 = 2 \times 10^{-5}$  m. The sample has a Young's modulus  $E$  of 210 GPa, a Poisson's ratio  $\nu$  of 0.3, a density  $\rho$



**Fig. 11** Three examples for verifying the efficacy and accuracy of our in-house phase-field solver. (a–b) Sample with a horizontal boundary crack. (c–d) Sample with a tilt boundary crack. (e–f) Sample with two tilt interior cracks. Panels (a, c, e) display the sample dimensions, crack positions, and boundary conditions while (b, d, f) display the simulation results and final fracture evolution using the phase-field solver. Result in panel (b) show a good match with Fig. 1c of<sup>44</sup>. Result in panel (d) show a good match with Fig. 4b of<sup>22</sup>. Result in (f) show a good match with Fig. 7j and k of<sup>46</sup>.

of  $4,000 \text{ kg/m}^3$ , and a critical energy release rate  $G_c$  of  $2,700 \text{ J/m}^2$ . A 0.5-mm long crack, extending from the left boundary to the horizontal center, is placed at the vertical center position. The bottom boundary of the sample is fixed. A velocity of  $5 \text{ m/s}$  is applied at the top boundary, generating an effective tensile boundary condition. Figure 11b shows the resulting fracture pattern at the failure time, which shows that the fracture horizontally



**Fig. 12** HOSS fracture example for same initial fracture configuration and material at time to failure with different mesh sizes. Each damaged cell edge is superficially extended in a Gaussian distribution about itself such that  $2\sigma \approx 1\ell$  where  $\ell$  is the characteristic element length. The distribution shows areas of most intense fracturing where  $d = 1.0$  while simultaneously visualizing the mesh sizing. **(a1–a3)** is a mesh size  $\times 2$  the mesh size used for our simulations. **(b1–b3)** is the mesh size used for our simulations ( $\approx 3 \cdot 10^{-3}$  m). **(c1–c3)** is  $\frac{1}{2}$  the mesh size used for our simulations. The bulk behavior across all domains is relatively similar, but the sensitivity on fracture propagation is dependent on the mesh once strain localization initiates. A detailed mesh-objective analysis remains beyond the scope of current dataset.

propagates to the right boundary from the center along a straight line. The result is consistent with the simulation result displayed in Fig. 1c of<sup>44</sup>.

In the second example, we simulate fracture propagation using the phase-field solver in a larger, rectangle sample under tensile boundary conditions. This test example aims to replicate the numerical example in<sup>22</sup>. The sample, as displayed in Fig. 11c, is 90 mm in height and 70 mm in width. The sample is discretized with  $N_{\text{horizontal}} \times N_{\text{vertical}} = 156 \times 200$  regular rectangle spectral elements with a grid size of  $4.5 \times 10^{-4}$  m. The characteristic fracture width is  $l_0 = 15 \times 10^{-4}$  m. The properties of the medium are  $E = 280$  MPa,  $\nu = 0.45$ ,  $\rho = 1,250$  km/m<sup>3</sup>, and  $G_c = 27$  J/m<sup>2</sup>, resembling CO poly (ethylene carbon monoxide) copolymer. A tilt crack, forming a  $60^\circ$  angle with the vertical boundary, extends from the left boundary to the interior of the sample. The length of the crack is 33 mm, and the end of the fracture inside the sample is at 45 mm in the vertical direction. The bottom boundary of the sample is fixed, while a small velocity of 0.01 m/s is applied at the top boundary. Figure 11d shows the fracturing simulation result at the moment of material failure. The crack, even though tilted at the initial state, horizontally propagates to the right boundary of the sample. The result is consistent with the simulation result in<sup>22</sup> (Fig. 4b within<sup>22</sup>). Furthermore, our simulation result is in good agreement with the physics experiment results in<sup>45</sup>, where the real

fracture propagation direction is  $-28 \pm 1.5^\circ$  (note that this is a value based on their experiment configuration; our simulation result is equivalently  $-30^\circ$  in their configuration).

In the third experiment, we simulate compression-shear fracturing with our in-house phase-field solver in a horizontally slim sample under vertical uniaxial pressures. This test example aims to mimic the experiment conducted by<sup>46</sup>. More specifically, we aim to mimic their sample named S-60-60 as listed in their Table 2. The sample, as shown in Fig. 11e, is 140 mm in height and 70 mm in width. The sample is discretized with  $N_{\text{horizontal}} \times N_{\text{vertical}} = 140 \times 280$  regular rectangle spectral elements with a grid size of  $5 \times 10^{-4}$  m. The characteristic fracture width is  $l_0 = 20 \times 10^{-4}$  m. According to<sup>46</sup>, the upper half of the sample has a Young's modulus  $E = 4.35$  GPa, a Poisson's ratio  $\nu = 0.21$ , and a density  $\rho = 2,182$  kg/m<sup>3</sup>, while the lower half of the sample has  $E = 5.35$  GPa,  $\nu = 0.17$ , and  $\rho = 2,355$  kg/m<sup>3</sup>. In addition, we set  $G_c = 200$  J/m<sup>2</sup>, mimicking common granite rock. Two tilt cracks, both with a length of 15 mm and a tilt angle of  $60^\circ$ , are at the center of the sample, with a separation of 22 mm. The boundary conditions applied at the top and bottom boundaries are pressures in the form of opposite velocity of 0.1 m/s along the vertical direction. The compression boundary conditions are expected to induce shear fracturing within the sample<sup>46</sup>. Figure 11f displays the simulation result generated with our phase-field solver for this sample at the moment of failure, where the two separate cracks connect with each other and extend to sample boundaries due to shearing. It is clear that our simulation result is in close correlation with the real experiment photo of rock sample failure displayed in Fig. 7j and k of<sup>46</sup>, especially regarding the dominant cracking path photographed in the experiment. It is important to notice that our phase-field solver does not generate the wing cracks and other secondary fractures that are observed by<sup>46</sup>. However, these types of fractures are not dominant cracks in the real physics experiment. Improvement on the formulation of the history energy function to more accurately capture compression-shear strain energy concentration and evolution may improve the fidelity and accuracy of our solver in such cases.

The above three simulation results verify the efficacy and accuracy of our phase-field solver for simulating fracture nucleation and propagation from very small scale (1 mm) to large scale (100 m) under different boundary conditions (tensile and compression-shearing) and within different types of materials. This indicates that our phase-field solver provides reliable fracturing simulation results for the materials mentioned in Section 6.

**Combined finite discrete-element method.** The simulation software used in this work, HOSS, includes a robust suite of benchmarks and validation studies, which we do not detail here as they are thoroughly documented elsewhere<sup>25</sup>. However, we do expand upon the common instances and corrective measures we account for in our analysis. These common instances are directly related to the numerical instability of the problem that is important to consider for materials undergoing fracturing, fragmentation, and contact interactions.

With any meshing method that incorporates an explicit time step there must be careful consideration of the critical time step based on the Courant condition. If the time step is too large numerical instability and divergence will occur. We appropriately account for this with our small time stepping  $1 \cdot 10^{-8}$  s. The critical time step is dependent on the mesh size with an increasing mesh allowing for a larger critical time step, but a loss of fidelity in the fracture patterns. Fracture-based finite element formulations are known to exhibit sensitivity to the underlying mesh once strain localization initiates<sup>47,48</sup>. For example, HOSS accurately assesses fracture coalescence by capturing distinct mixed-mode fracture processes when compared with experiment<sup>48</sup>, but will not match the observed fracture pattern exactly due to the mesh sensitivity. This behavior is intrinsic to damage-based representations of fracture and arising from the fracture being modeled as separation along pre-defined element edges, governed by cohesive laws. In practice, the bulk material response, rather than the exact fracture path, is often used to assess mesh convergence<sup>47</sup>. Accordingly, we include a qualitative figure demonstrating that the bulk damage evolution remains spatially consistent across different mesh sizes, indicating that the chosen discretization provides an effective convergence for the purposes of this study (Fig. 12). As the focus of the present work is on the comparative fracture behavior and material responses rather than numerical regularization, a detailed mesh-objective analysis remains beyond the scope of current dataset.

Another important consideration for FDEM problems is high-frequency oscillations. These oscillations are a direct result of small elements experiencing excessive numerical noise. During the meshing process there are sometimes combinations of fracture patterns that produce excessively small elements. These instances create numerical instability and divergence as previously explained. We apply a check of the mesh prior to our batch submissions which ignores these spurious simulations prior to running. Furthermore, there is a damping term introduced for discrete elements called Munjiza viscosity that targets numerical artifacts while leaving the real wave (energy) propagation largely unaffected<sup>26,35</sup>.

Due to the large number of files running in batch across multiple HPC systems there were a small handful of instances where simulations resulted in corrupted files. Since we had to post-process each HOSS simulation using the procedure detailed in Section 6, we have already assessed every simulation's accessibility by removing those that would error when attempting to open. By post-processing the data we have ensured that there are no corrupt files or missing fractures and have also opened every file individually to assess the reliability of its contents.

## Usage Notes

The repository contains simulation output data (phase-field-based and HOSS-based simulations) organized into compressed .tar.gz archives of HDF5 (.h5) files. The data are grouped under five different material directories for the phase-field data (aluminum, PBX, shale, steel, and tungsten), and three material directories for the HOSS data (PBX, shale, and tungsten). We have provided a Python script named "extract\_h5.py" in the original data repository (see [https://huggingface.co/datasets/smartFRACs/material\\_fracturing](https://huggingface.co/datasets/smartFRACs/material_fracturing)<sup>43</sup>), that shows how to download and manipulate sample records from the datasets.

```

import h5py
from urllib.request import urlretrieve
import tarfile
import os
import glob

# URLs
url_drp = 'https://huggingface.co/datasets/smartFRACs/material_fracturing/resolve/main/'
# example link to tree structure
# https://huggingface.co/datasets/smartFRACs/material_fracturing/tree/main
url_proj = url_drp + 'HOSS/pbx/' # HOSS project folder
batch_name = 'batch_000.tar.gz' # example batch file
url_batch = url_proj + batch_name

# output file name
output_filename = batch_name

# download the files from the portal
urlretrieve(f'{url_batch}', output_filename)

# Extract to this folder (change if desired)
extract_to = "batch_000_contents"

# Open and extract
with tarfile.open(output_filename, "r:gz") as tar:
    tar.extractall(path=extract_to)

# get list of filenames and extract choose to open just the first
filenames = h5_files = glob.glob(os.path.join(extract_to, '**', '*.h5'), recursive=True)
filename = filenames[0] # single example fracture to extract from full filename list

# Save data to dictionary
data_dict = {}
# Open and Load the HDF5 file
with h5py.File(filename, 'r') as f:
    def collect_datasets(name, obj):
        if isinstance(obj, h5py.Dataset):
            print(f>Loading dataset: {name}, shape: {obj.shape}, dtype: {obj.dtype}")
            data_dict[name] = obj[()] # Load entire dataset into memory

    f.visititems(collect_datasets) # walk through h5 and collect in data_dict

```

**Listing 1** Python code to download and load the data into a dictionary and print its constituent objects.

## Data availability

The code and data for this project is available online<sup>43</sup>.

Received: 5 August 2025; Accepted: 2 December 2025;

Published online: 13 December 2025

## References

1. Viswanathan, H. S. *et al.* From fluid flow to coupled processes in fractured rock: Recent advances and new frontiers. *Reviews of Geophysics* **60**, e2021RG000744, <https://doi.org/10.1029/2021RG000744> (2022).
2. Hyman, J. *et al.* Understanding hydraulic fracturing: a multi-scale problem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374**, 20150426 (2016).
3. Golewski, G. L. The phenomenon of cracking in cement concretes and reinforced concrete structures: the mechanism of cracks formation, causes of their initiation, types and places of occurrence, and methods of detection—a review. *Buildings* **13**, 765 (2023).
4. Husseini, M. L., Jovanovich, D. B., Randall, M. & Freund, L. The fracture energy of earthquakes. *Geophysical Journal International* **43**, 367–385 (1975).
5. Ellsworth, W. L. Injection-induced earthquakes. *Science* **341**, 1225942, <https://doi.org/10.1126/science.1225942> (2013).
6. Mohtadi-Bonab, M. Effects of different parameters on initiation and propagation of stress corrosion cracks in pipeline steels: a review. *Metals* **9**, 590 (2019).
7. Kalthoff, J. Fracture behavior under high rates of loading. *Engineering fracture mechanics* **23**, 289–298 (1986).
8. Launey, M. E. & Ritchie, R. O. On the fracture toughness of advanced materials. *Advanced Materials* **21**, 2103–2110 (2009).
9. Hobbs, M. & Kaneshige, M. Ignition experiments and models of a plastic bonded explosive (pbx 9502). *The Journal of Chemical Physics* **140** (2014).
10. Gupta, H. & Zioupos, P. Fracture of bone tissue: the ‘hows’ and the ‘whys’. *Medical engineering & physics* **30**, 1209–1226 (2008).
11. Brown, T. B. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).
12. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Burstein, J., Doran, C. & Solorio, T. (eds) *BERT: Pre-training of deep bidirectional transformers for language understanding*. (eds Burstein, J., Doran, C. & Solorio, T.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186 (Association for Computational Linguistics, Minneapolis, Minnesota, 2019). <https://aclanthology.org/N19-1423/>.
13. Chowdhery, A. *et al.* Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research* **24**, 1–113 (2023).
14. Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *nature* **596**, 583–589 (2021).
15. Stokes, J. M. *et al.* A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702 (2020).
16. Palmer, T., Doblas-Reyes, F., Weisheimer, A. & Rodwell, M. Toward seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bulletin of the American Meteorological Society* **89**, 459–470 (2008).
17. Bodnar, C. *et al.* Aurora: A foundation model of the atmosphere. *arXiv preprint arXiv:2405.13063* (2024).
18. Rupp, M., Tkatchenko, A., Müller, K.-R. & Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters* **108**, 058301 (2012).
19. Ahmad, W., Simon, E., Chithrananda, S., Grand, G. & Ramsundar, B. Chemberta-2: Towards chemical foundation models. *arXiv preprint arXiv:2209.01712* (2022).

20. Mikelić, A., Wheeler, M. F. & Wick, T. Phase-field modeling of a fluid-driven fracture in a poroelastic medium. *Computational Geosciences* **19**, 1171–1195, <https://doi.org/10.1007/s10596-015-9532-5> (2015).
21. Santillán, D., Juanes, R. & Cueto-Felgueroso, L. Phase field model of fluid-driven fracture in elastic media: Immersed-fracture formulation and validation with analytical solutions. *Journal of Geophysical Research: Solid Earth* **122**, 2565–2589, <https://doi.org/10.1002/2016JB013572> (2017).
22. Natarajan, S., Annabattula, R. K. & Martínez-Pañeda, E. Phase field modelling of crack propagation in functionally graded materials. *Composites Part B: Engineering* **169**, 239–248 (2019).
23. Zhou, S. & Zhuang, X. Phase field modeling of hydraulic fracture propagation in transversely isotropic poroelastic media. *Acta Geotechnica* **15**, 2599–2618, <https://doi.org/10.1007/s11440-020-00913-z> (2020).
24. Mises, R. The mechanics of solids in the plastically-deformable state. Tech. Rep. (1986).
25. Knight, E. E. *et al.* Hoss: an implementation of the combined finite-discrete element method. *Computational Particle Mechanics* **7**, 765–787, <https://doi.org/10.1007/s40571-020-00349-y> (2020).
26. Munjiza, A., Knight, E. E. & Rougier, E. *The combined finite-discrete element method* (Wiley, Hoboken, NJ, 2011).
27. Lei, Z., Rougier, E., Knight, E. E., Zang, M. & Munjiza, A. Impact fracture and fragmentation of glass via the 3d combined finite-discrete element method. *Applied Sciences* **11**, 2484 (2021).
28. Lei, Z., Rougier, E., Knight, E. E. & Munjiza, A. A framework for grand scale parallelization of the combined finite discrete element method in 2d. *Computational Particle Mechanics* **1**, 307–319 (2014).
29. Lei, Z., Rougier, E., Knight, E. E., Munjiza, A. & Viswanathan, H. A generalized anisotropic deformation formulation for geomaterials. *Computational Particle Mechanics* **3**, 215–228 (2016).
30. *3D Numerical Modeling of Hydraulic Fracture Branching*, Vol. 55th U.S. Rock Mechanics/Geomechanics Symposium of U.S. Rock Mechanics/Geomechanics Symposium.
31. Munjiza, A., Rougier, E., Lei, Z. & Knight, E. E. FSIS: a novel fluid-solid interaction solver for fracturing and fragmenting solids. *Computational Particle Mechanics* **7**, 789–805, <https://doi.org/10.1007/s40571-020-00314-9> (2020).
32. Henderson, S. F. *et al.* Modeling the formation of sedan crater using the flag and hoss codes. *Scientific Reports* **15**, 21466 (2025).
33. Lei, Z. & Zang, M. An approach to combining 3d discrete and finite element methods based on penalty function method. *Computational Mechanics* **46**, 609–619 (2010).
34. Lei, Z., Knight, E. E., Munjiza, A. & Rougier, E. Unified cohesive zone model (uczm) for fracturing and fragmenting solids. *Engineering Fracture Mechanics* **312**, 110598 (2024).
35. Munjiza, A., Rougier, E. & Knight, E. E. *Large strain finite element method: a practical course* (Wiley, Chichester, West Sussex, UK, 2015).
36. Munjiza, A. *Computational Mechanics of Discontinua* (Wiley, Chichester, West Sussex, PO19 8SQ, United Kingdom, 2004).
37. Lei, Z., Rougier, E., Euser, B. & Munjiza, A. A smooth contact algorithm for the combined finite discrete element method. *Computational Particle Mechanics* **7**, 807–821 (2020).
38. Rougier, E., Munjiza, A. & John, N. Numerical comparison of some explicit time integration schemes used in dem, fem/dem and molecular dynamics. *International journal for numerical methods in engineering* **61**, 856–879 (2004).
39. Rougier, E. & Munjiza, A. *Mrck\_3d contact detection algorithm* (2010).
40. Munjiza, A. & Andrews, K. Nbs contact detection algorithm for bodies of similar size. *International Journal for Numerical Methods in Engineering* **43**, 131–149 (1998).
41. Scotto, M. R. *et al.* Pyvista: 3d visualization and analysis of scientific data <https://pyvista.org>. Accessed: 2025-07-20 (2024).
42. Komatitsch, D. & Tromp, J. Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International* **139**, 806–822, <https://doi.org/10.1046/j.1365-246x.1999.00967.x> (1999).
43. Hill, R. G. *et al.* smartFRACS/material\_fracturing (Revision adflfd2). Hugging Face. <https://doi.org/10.57967/hf/6069> (2025).
44. Ren, H., Zhuang, X., Anitescu, C. & Rabczuk, T. An explicit phase field method for brittle dynamic fracture. *Computers & Structures* **217**, 45–56, <https://doi.org/10.1016/j.compstruc.2019.03.005> (2019).
45. Abanto-Bueno, J. & Lambros, J. An experimental study of mixed mode crack initiation and growth in functionally graded materials. *Experimental Mechanics* **46**, 179–196, <https://doi.org/10.1007/s11340-006-6416-6> (2006).
46. Lin, Q. *et al.* Crack coalescence in rock-like specimens with two dissimilar layers and pre-existing double parallel joints under uniaxial compression. *International Journal of Rock Mechanics and Mining Sciences* **139**, 104621, <https://doi.org/10.1016/j.ijrmm.2021.104621> (2021).
47. Tatone, B. & Grasselli, G. A calibration procedure for two-dimensional laboratory-scale hybrid finite-discrete element simulations. *International Journal of Rock Mechanics and Mining Sciences* **75**, 56–72 (2015).
48. Euser, B. *et al.* Simulation of fracture coalescence in granite via the combined finite–discrete element method. *Rock Mechanics and Rock Engineering* **52**, 3213–3227 (2019).

## Acknowledgements

This research work was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20250637DI. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program. We gratefully acknowledge partial support from the Department of Energy, Office of Science, Office of Basic Energy Sciences, Geoscience Research program under Award Number LANLECA1. Approved for unlimited release LA-UR-25-22599.

## Author contributions

R.G.H.: Formal analysis, Investigation, Data Acquisition, Data curation, Writing - original draft, Writing - review & editing, Visualization; K.G.: Formal analysis, Software, Investigation, Data Acquisition, Data curation, Writing - original draft, Visualization; A.P.: Validation, Data curation, Writing - original draft, Writing - review & editing, Visualization; A.M.: Validation, Data curation, Writing - review & editing; X.W.: Investigation, Data Acquisition, Data curation, Visualization; P.G.: Data curation, Visualization; E.R.: HOSS technical advice and guidance, Software, Writing - review & editing; Z.L.: HOSS technical advice and guidance, Software; J.E.S.: Conceptualization; V.A.: Methodology, Writing - review & editing; Q.K.: Writing - review & editing; J.H.: Writing - review & editing; A.H.: Writing - review & editing; C.M.S.: Data Curation; Writing - review & editing; N.D.: Writing - review & editing; E.L.: Resources, Funding acquisition, Conceptualization; H.V.: Conceptualization, Writing - review & editing, Funding acquisition; D.O.: Conceptualization, Methodology, Writing - review & editing, Funding acquisition.

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to R.G.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025