Identifying Structural Flow Defects in Disordered Solids Using Machine-Learning Methods

E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras,
and A. J. Liu

Identifying structural flow defects in disordered solids using machine learning methods

E. D. Cubuk,[1, *] S. S. Schoenholz,[2, †] J. M. Rieser,[2] B. D. Malone,[1] J. Rottler,[3] D. J. Durian,[2] E. Kaxiras,[1] and A. J. Liu[2]

[1]*Department of Physics and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[3]*Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T1Z4, Canada*
(Dated: February 2, 2015)

We use machine learning methods on local structure to identify flow defects – or particles susceptible to rearrangement – in jammed and glassy systems. We apply this method successfully to two very different systems: a two dimensional experimental realization of a granular pillar under compression, and a Lennard-Jones glass in both two and three dimensions above and below its glass transition temperature. We also identify characteristics of flow defects that differentiate them from the rest of the sample. Our results show it is possible to discern subtle structural features responsible for heterogeneous dynamics observed across a broad range of disordered materials.

All solids flow at high enough applied stress and melt at high enough temperature. Crystalline solids flow [1] and premelt [2] via localized particle rearrangements that occur preferentially at structural defects known as dislocations. The population of dislocations therefore controls both how crystalline solids flow and how they melt. In disordered solids, it has long been hypothesized that localized particle rearrangements [3] induced by stress or temperature also occur at localized flow defects [4–6]. Like dislocations in crystals [7], flow defects in disordered solids are particularly effective in scattering sound waves, so analyses of the low-frequency vibrational modes [8] have been used successfully to demonstrate the existence of localized flow defects [7, 9–17]. However, all attempts to identify flow defects [18, 19] directly from the structure, without using knowledge of the inter-particle interactions, have failed [18, 19]. Likewise, in supercooled liquids, purely structural measures correlate only weakly with kinetic heterogeneities [17], although correlations between structure and dynamics have been established indirectly [20–24].

Here we introduce a novel application of machine learning (ML) methods to identify "soft" particles that are susceptible to rearrangement or, equivalently, that belong to flow defects, from the local structural geometry alone. Note that our goal is not to identify particles that undergo rearrangements–this requires studying their dynamics. Rather, our goal is to identify the structural characteristics that distinguish particles that are susceptible to rearrangement from those that are not. We apply the method to two very different systems–an experimental frictional granular packing under uniaxial compression and a model thermal Lennard-Jones glass in both two and three dimensions. The analysis of granular packing shows that our method succeeds even when previous methods based on vibrational modes [12] are inapplicable. The results for Lennard-Jones systems show that the correlation between structure and irreversible rearrangements does not degrade with increasing temperature, even above the dynamical glass transition, and is equally strong in two and three dimensions. Finally, we exploit the method to discover which structural properties distinguish soft particles from the rest of the system, and to understand why previous attempts to identify them by structural analysis have failed.

Physically-motivated quantities such as free volume or bond orientational order correlate with flow defects [12], but are insufficient to identify them *a priori*. We introduce a large set of quantities that are each less descriptive, but when used as a group provide a more complete and unbiased description of local structure. These quantities have been used to represent the potential energy landscape of complex materials from quantum mechanical calculations [25]. For a system composed of multiple species of particles, we define two families of structure functions for each particle i,

$$G_Y^X(i; \mu) = \sum_j e^{-(R_{ij}-\mu)^2/L^2} \tag{1}$$

$$\Psi_{YZ}^X(i; \xi, \lambda, \zeta) = \sum_j \sum_k e^{-(R_{ij}^2+R_{ik}^2+R_{jk}^2)/\xi^2}(1 + \lambda \cos\theta_{ijk})^\zeta \tag{2}$$

where R_{ij} is the distance between particles i and j; θ_{ijk} is the angle between particles i, j and k; L, μ, ξ, λ, and ζ are constants; X,Y,Z are labels that identify the different species of particles in the system, with the correspondence $i \leftrightarrow X$, $j \leftrightarrow Y$, $k \leftrightarrow Z$. By using many different values of the constants μ, ξ, λ, and ζ we generate many structure functions in each family that characterize different aspects of a particle's local configuration; for a list, see Ref. [26]. The first family of structure functions G characterizes radial density properties of the neighborhood, while the second family, Ψ, characterizes bond orientation properties. The sums are taken over particle

pairs whose distance is within a large cutoff R_c^S. Our results are qualitatively insensitive to the choice of R_c^S as long as it includes several neighbor shells.

Having characterized local structure through $G_Y^X(i; \mu)$ and $\Psi_{YZ}^X(i; \xi, \lambda, \zeta)$, we introduce a method to infer from this information the location of flow defects in disordered solids. Generically, we begin with a set of N particles to be classified as "soft" or "hard." Each particle is described by a set of M variables derived from the two families of functions G_Y^X and Ψ_{YZ}^X by varying the constants μ, ξ, λ, and ζ (here $M = 160$); this is represented by the set of vectors $\{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_N\}$, where $\boldsymbol{F}_i$ constitutes an embedding of the local environment of a particle i, constructed at a time t_i, in $\mathbb{R}^M$. We select at random a subset of n of these particles (the "training set") and label the particles with an index $r = 0, 1$. A particle i is labelled as $r_i = 1$ if it rearranges (the details of which will be discussed below) between a time t_i, when the structure is characterized, and a time $t_i + \Delta t$, and with $r_i = 0$ if it does not rearrange in this time interval.

The next step is to use the particles in the training set, along with their labels r_i, to classify a particle as soft or hard based on the structure of its neighborhood. We use the support vector machine (SVM) method [27], which constructs a hyperplane in $\mathbb{R}^M$ that best separates particles with $r_i = 1$ from those with $r_i = 0$ in the high-dimensional space characterizing local structure. Once this hyperplane has been established for the training set, we classify particles not in the training set (either from the same configuration or from other configurations at the same conditions) as soft if they fall on the $r = 1$ side of the hyperplane, and as hard if they fall on the $r = 0$ side. Generically, no exact separation is possible so the SVM method is adapted to penalize particles whose classification is incorrect; the degree of penalty is controlled by a parameter C where larger values of C allow for fewer incorrect classifications. We find that the quality of our classifications is insensitive to C for $C > 0.1$ [26]. The SVM algorithm was implemented using the LIBSVM package [28].

Note that although we are using the dynamics of particles in the training set to construct the hyperplane, the hyperplane is not a dynamical quantity but a structural one. It identifies which aspects of the structural environment are different for particles that rearrange and for those that do not. Once the hyperplane is constructed, it can therefore be used to classify particles by their local structure as soft (susceptible to rearrangement) or hard.

To identify rearrangements we calculate, for each particle i, the widely-used quantity [5]

$$D_{\min}^2(i) \equiv \min_\Lambda \left\{ \frac{1}{z} \sum_j (R_{ij}(t + \Delta t) - \Lambda R_{ij}(t))^2 \right\}, \quad (3)$$

which characterizes the magnitude of non-affine displacement during a time interval Δt. Here the sum runs over

neighbors j within a distance of R_c^D of particle i, R_{ij} is the center-center displacement between particles i and j, z is the number of neighbors within R_c^D. The quantity is minimized over choices of the local strain tensor Λ. We find $D_{\min}^2$ is insensitive to the choice of R_c^D and Δt so long as R_c^D is large enough to capture the particles local neighborhood and Δt is longer than the ballistic timescale. A particle is said to have undergone a rearrangement if $D_{\min}^2 \geq D_{\min,0}^2$. We choose $D_{\min,0}^2$ such that approximately 0.15% of the particles from each species in each system has gone through a rearrangement although the results depend little on the specific choice of $D_{\min,0}^2$ [26].
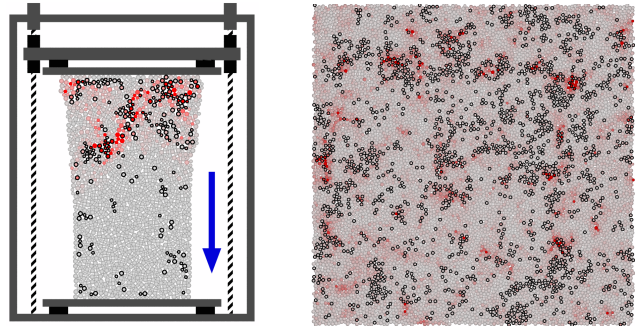

FIG. 1. (color online) Snapshot configurations of the two systems studied. Particles are colored gray to red according to their $D_{\min}^2$ value. Particles identified as soft by the SVM are outlined in black. (a) A snapshot of the pillar system. Compression occurs in the direction indicated. (b) A snapshot of the $d = 2$ sheared, thermal Lennard-Jones system.

We first test our approach on an experimental system of two-dimensional ($d = 2$) "pillars" of particles. A bidisperse pillar made up of grains (plastic cylinders resting upright on a horizontal substrate) is situated between two plates in a custom-built apparatus. The bottom plate is fixed and the top plate is driven into the pillar at a constant speed of $v_0 = 0.085$ mm/s. The pillars are composed of a bidisperse mixture of approximately 1500 rigid grains with size ratio 3:4 and the large particles having a radius of $d_{AA} = 0.3175$cm. These particles have elastic and frictional interactions with each other, as well as frictional interactions with the substrate, making the identification of flow defects using vibrational modes impossible. A camera is mounted above and captures images at 7 Hz throughout the compression.

We construct our training set from compression experiments performed on ten different pillars. We select 6,000 particles at random from the entire duration of the experiments that undergo a rearrangement in the next 1.43 seconds and an equal number of particles that do not. To identify rearrangements we calculate $D_{\min}^2$ with $R_c^D = 1.5 \cdot d_{AA}$ and $D_{\min,0}^2 = 0.25 d_{AA}^2$. Compression of the mechanical pillar from the top only affects particles above a certain "front" that starts at the top and advances towards the bottom with time. Our training set

contains only particles within this activated front. Particles in a horizontal slice between y and $y + \delta y$ are said to be within the activated front if the average speed of particles in the slice exceeds $v_{\text{thresh}} \sim 0.04$ mm/s.

As a second test, we apply our approach to a model glass in both $d = 2$ and $d = 3$. We study a 65:35 binary Lennard-Jones (LJ) mixture with $\sigma_{AA} = 1.0$, $\sigma_{AB} = 0.88$, $\sigma_{BB} = 0.8$, $\epsilon_{AA} = 1.0$, $\epsilon_{AB} = 1.5$, and $\epsilon_{BB} = 0.5$ [29]. The LJ potential is cut off at $2.5\sigma_{AA}$ and smoothed so that both first and second derivatives go continuously to zero at the cutoff. The natural units for the simulation are σ_{AA} for distances, ϵ_{AA} for energies, and $\tau = \sqrt{m\sigma_{AA}^2/\epsilon_{AA}}$ for times. We perform molecular dynamics simulations using LAMMPS [30] with a time step of $5 \times 10^{-3}\tau$ at density $\rho = 1.2$, using a Nosé-Hoover thermostat with a time constant of τ. Here, temperature is in units of ϵ_{AA}/k_B, where k_B is the Boltzmann constant. In $d = 2$, we consider a system of 10,000 particles at temperatures $T = 0.1, 0.2, 0.3$, and 0.4 at a single strain rate, $\dot{\gamma} = 10^{-4}/\tau$. In all cases data was collected after allowing the system to reach steady state by shearing up to 20% strain. In $d = 3$, we use a collection of 30,000 particles at temperatures $T = 0.4, 0.5, 0.6$ with no strain. The quiescent system has a glass transition at $T_G = 0.33$ in $d = 2$ and $T_G = 0.58$ in $d = 3$ [29]. Therefore, in both dimensions we study the system both above and below its glass transition temperature.

At each temperature we construct training sets of 6,000 and 10,000 particles, in $d = 2$ and 3, respectively, selected at random from the entire run that undergo a rearrangement in the following $\Delta t = 2\tau$ units of time, and an equal number of particles that do not undergo a rearrangement. To identify rearrangements, we calculate $D_{\min}^2$ with $R_c^D = 2.5\sigma_{AA}$ to be the same as the range of the truncated LJ potential and $\Delta t = 2\tau$. In $d = 3$ the $D_{\min}^2$ distributions of the species A and species B particles differ significantly and so are treated separately throughout the analysis.

We now test our classifying hyperplane on the three systems outlined above. For each system we construct a test set of particles that were not used in training the SVM; for the pillar this test set consists of 100,000 particle environments from many snapshots of ten additional compression experiments. For the $d = 2$ and $d = 3$ LJ glasses we use 2×10^7 and 75×10^7 unseen particle environments (composed of each particle's environment aggregated over every snapshot of the simulation) respectively. Fig. 2 (a)-(d) shows the probability, $P(D_{\min}^2)$, that a particle with an observed value of $D_{\min}^2$ was identified as soft by our classification *a priori* for each system. Fig 2 (c)-(d) treat the particles of species A and species B separately for the $d = 3$ LJ glass. In all cases we see that $P(D_{\min}^2)$ rises with increasing plastic activity, $D_{\min}^2$. This implies that the particles identified as soft by the SVM are more likely to be involved in plastic flow. For the granular pillar, 21% of the particles are classified as soft,
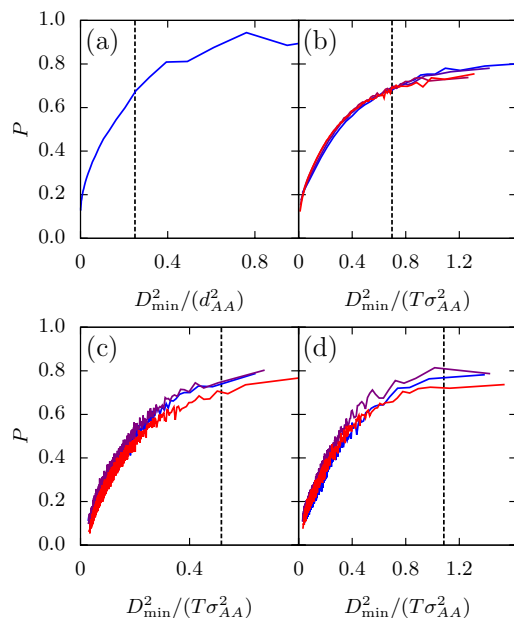

FIG. 2. (color online) Probability that a particle of a given $D_{\min}^2$ value is soft. The vertical dashed lines are corresponding $D_{\min,0}^2$ values. (a) The result for the pillar system, where d_{AA} refers to the large grain diameter (since this is a granular system with macroscopic grains, thermal fluctuations are negligible). (b) The result of using an SVM trained at a temperature T ($T = 0.1, 0.2, 0.3$ and 0.4 shown in different colors) to classify data at the same temperature for the $d = 2$ LJ glass. (c,d) Results for species A and B, respectively, for the $d = 3$ system at $T = 0.4, 0.5$ and 0.6.

and these particles capture 80% of the rearrangements. For the $d = 2$ and $d = 3$ LJ systems, these numbers are 26% and 73%, and 24% and 72%, respectively, at the highest temperatures studied. Thus, we consistently find rearrangements to occur at soft particles about 3-4 times more frequently than if the soft particles were randomly chosen. Finally, Fig. 2(b,c,d) show that $P(D_{\min}^2)$ collapses for different T both for 2D and 3D systems, when $D_{\min}^2$ is scaled by $T\sigma_{AA}^2$; this scaling arises since for particles not undergoing rearrangement, $D_{\min}^2 \sim \langle v^2 \rangle \sim T$ by the equipartition theorem [16].

Remarkably, our ability to identify soft particles does not decrease with increasing temperature or dimension. For the $d = 2$ LJ system over the same temperature range, the accuracy of the vibrational mode method decreases by over 50% [16]. The key difference between the two methods is that we construct local environments from the actual particle positions in snapshots of the thermal system, while soft particles from vibrational modes are extracted from particle positions in the inherent structures obtained by quenching to $T = 0$ [26].

The hyperplane that divides soft from hard particles tells us which features of the local structural environment are important in distinguishing between the hard and soft particles. To illustrate this, we focus on the
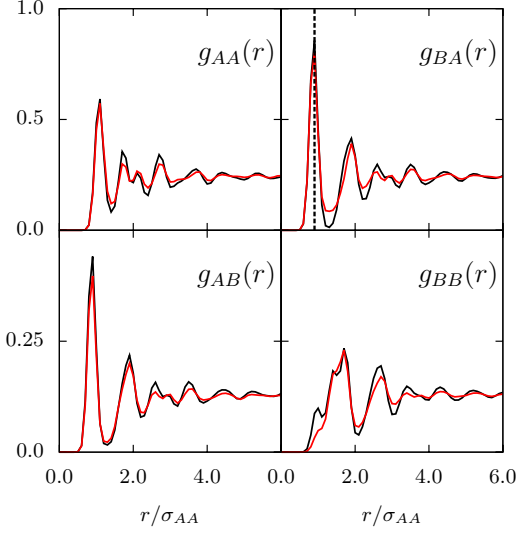
FIG. 3. (color online) Radial distribution functions averaged over hard (black lines) or soft (red lines) particles. g_{AB} and g_{BA} of soft particles are not equal to each other since they refer to different kinds of regions: neighbors of soft particles from species A and neighbors of soft particles from species B, respectively.
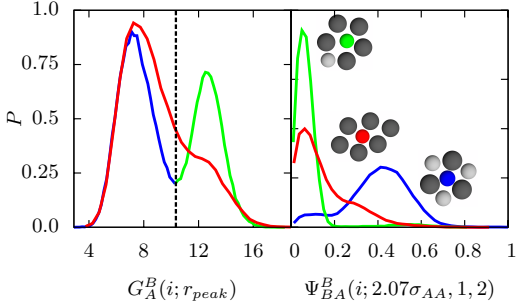

FIG. 4. (color online) (a) Distribution of $G_B^A(i; r_{\text{peak}}^{AB})$, proportional to the gaussian weighted density at r_{peak}^{AB}, for soft (red) and hard (blue/green) particles. r_{peak}^{AB} corresponds to the first peak of g_{AB} or g_{BA}. (b) Distribution of $\Psi_{AB}^B(i; 2.07\sigma_{AA}, 1, 2)$, proportional to the density of neighbors with small bond angles near a particle i, for soft (red) and hard (blue/green) particles. The inset shows examples of configurations with corresponding radial and bond orientation properties, where dark (light) gray neighbors are of species A (B).

sheared $d = 2$ LJ glass at $T = 0.1$. The first set of structure functions, $G_Y^X(i; \mu)$ defined in Eq. (1), has a familiar physical interpretation in terms of the radial distribution function: $g_{XY}(r) = \lim_{L \to 0} \langle G_Y^X(i; r) \rangle / 2\pi r$. This family of structure functions is essentially a parameterization of a particle's local contribution to $g(r)$. In Fig. 3 we show the approximations to various $g(r)$ constructed in this way, using $L = 0.1\sigma_{AA}$, for particles identified as hard (black lines) and soft (red lines).

In all cases, we see that soft particles feature slightly

lower peaks and higher troughs in $g(r)$. To see whether this difference is sufficient to identify soft particles, we expand our analysis beyond average values of $G_Y^X(i; \mu)$, to the *distributions* of the different structure functions. In Fig. 4(a) we show the distribution of values of $G_A^B(i; r_{\text{peak}}^{AB})$ where r_{peak}^{AB} is the location of the first peak of $g_{BA}(r)$, for soft particles (red) and hard particles (blue/green). While the distribution for soft particles features a single peak, that for hard particles is bimodal. This indicates the existence of (at least) two distinct populations of hard particles which we divide into two groups: one with $G_A^B(i; r_{\text{peak}}^{AB})/r < 1/2$ (blue) that we will call H_0-type, and one with $G_A^B(i; r_{\text{peak}}^{AB})/r > 1/2$ (green) that we will call H_1-type. Radial information therefore distinguishes between soft particles and H_1 hard particles but not between soft and H_0 hard particles.

We now consider the distribution of $\Psi_{AB}^B(i; 2.07\sigma_{AA}, 1, 2)$ for soft particles (red), H_0 hard particles (blue), and H_1 hard particles (green), shown in Fig. 3(c). Physically, $\Psi_{AB}^B(i; 2.07r_{\text{peak}}^{BA}, 1, 2)$ is large when there are many pairs of neighbors of the central particle that lie within a distance ξ with small angles between them, such that one is of species A and one is of species B. The soft particles fall into a single category (one peak, red) while H_0-type and H_1-type hard particles–defined from radial information above–have very different distributions (blue and green peaks). Unlike before, here the soft particles and the H_0 hard particles have very different distributions while soft particles and H_1 hard particles have similar distributions. Bond-angle information therefore distinguishes between soft particles and H_0-type hard particles but not between soft particles and H_1-type hard particles. To fully distinguish between soft and hard particles, both radial and bond-angle information is needed. Soft particles have environments that–at a minimum–have fewer particles in their nearest neighbor shell *and* larger angles between adjacent particles.

In summary, we have presented a novel ML method for identifying flow defects in disordered solids. We note that we have focused on the short-time correlation of structure with particle rearrangements. However, our method should shed light on the connection between local structural evolution and the correlation of rearrangements in time and space [31] over longer time scales in glassforming liquids. We also note that we cannot predict the specific particles that will participate in rearrangements at a later time; rather, we identify a *population* of particles that are likely to rearrange. The latter quantity is more useful in thermal and/or sheared systems, since fluctuations lead to stochasticity in rearrangements.

Our method relies on local structure alone, and can be applied directly to snapshots of experimental systems, in contrast to previous methods [12]. Our approach also scales linearly with the number of particles, N, while vibrational mode approaches scale as N^3. The

efficient identification of flow defects is critical to testing phenomenological approaches to plasticity based on flow defects [6, 32–34]. Previous applications of machine learning methods in physics have focused on approximation [25, 35, 36] or on optimization and design tools [37–40]. Our approach shows that such methods–designed for detecting subtle correlations–can also be used directly to gain conceptual understanding not achieved with conventional approaches.

---

[*] cubuk@fas.harvard.edu
[†] schsam@sas.upenn.edu

[1] G. I. Taylor, Proc. R. Soc. Lond. A Math. Phys. Sci. **145**, 362 (1934).
[2] A. M. Alsayed, M. F. Islam, J. Zhang, P. J. Collings, and A. G. Yodh, Science **309**, 1207 (2005).
[3] A. S. Argon and H. Y. Kuo, Materials science and Engineering **39**, 101 (1979).
[4] A. S. Argon, Acta metallurgica **27**.
[5] M. L. Falk and J. S. Langer, Phys. Rev. E **57**, 7192 (1998).
[6] L. Bocquet, A. Colin, and A. Ajdari, Phys. Rev. Lett. **103**, 036001 (2009).
[7] J. Rottler, S. S. Schoenholz, and A. J. Liu, Phys. Rev. E **89**, 042304 (2014).
[8] C. Brito and M. Wyart, J. Stat. Mech. , L08003 (2007).
[9] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman, Nat. Phys. **4**, 711 (2008).
[10] A. Widmer-Cooper, H. Perry, P. Harrowell, D. R. Reichman, *et al.*, J. Chem. Phys. **131**, 194508 (2009).
[11] R. Candelier, A. Widmer-Cooper, J. K. Kummerfeld, O. Dauchot, G. Biroli, P. Harrowell, and D. R. Reichman, Phys. Rev. Lett. **105**, 135702 (2010).
[12] M. L. Manning and A. J. Liu, Phys. Rev. Lett. **107**, 108302 (2011).
[13] K. Chen, M. L. Manning, P. J. Yunker, W. G. Ellenbroek, Z. Zhang, A. J. Liu, and A. G. Yodh, Phys. Rev. Lett. **107**, 108301 (2011).

[14] G. M. Hocky and D. R. Reichman, J. Chem. Phys. **138**, 12A537 (2013).
[15] K. Chen, T. Still, S. Schoenholz, K. B. Aptowicz, M. Schindler, A. C. Maggs, A. J. Liu, and A. G. Yodh, Phys. Rev. E **88**, 022315 (2013).
[16] S. S. Schoenholz, A. J. Liu, R. A. Riggleman, and J. Rottler, Phys. Rev. X **4**, 031014 (2014).
[17] R. L. Jack, A. J. Dunleavy, and C. P. Royall, Phys. Rev. Lett. **113**, 095703 (2014).
[18] J. Gilman, Phys. Today **28**, 46 (1975).
[19] P. Chaudhari, A. Levi, and P. Steinhardt, Phys. Rev. Lett. **43**, 1517 (1979).
[20] A. Widmer-Cooper, P. Harrowell, and H. Fynewever, Phys. Rev. Lett. **93**, 135701 (2004).
[21] A. Widmer-Cooper and P. Harrowell, J. Chem. Phys. **126**, 154503 (2007).
[22] D. Coslovich and G. Pastore, J. Chem. Phys. **75**, 784 (2006).
[23] D. Coslovich and G. Pastore, J. Chem. Phys. **127**, 124504 (2007).
[24] A. Malins, J. Eggers, H. Tanaka, and C. P. Royall, Faraday Discuss. **167**, 405 (2013).
[25] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[26] See Supplemental Material at [URL will be inserted by publisher] for details, which includes Refs [7, 12, 13, 15, 16, 28].
[27] C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995).
[28] C.-C. Chang and C.-J. Lin, ACM Transactions on Intelligent Systems and Technology **2**, 27:1 (2011).
[29] R. Brüning, D. A. St-Onge, S. Patterson, and W. Kob, J. Phys.: Condens. Matter **21**, 035117 (2009).
[30] S. Plimpton, J. Comp. Phys. **117**, 1 (1995).
[31] A. S. Keys, L. O. Hedges, J. P. Garrahan, S. C. Glotzer, and D. Chandler, Phys. Rev. X. **1**, 021013 (2011).
[32] M. L. Falk and J. S. Langer, Ann. Rev. of Cond. Matt. Phys. **2**, 353 (2011).
[33] V. Mansard, A. Colin, P. Chauduri, and L. Bocquet, Soft Matter **7** (2011).
[34] J. Lin, A. Saade, E. Lerner, A. Rosso, and M. Wyart, Europhys. Lett. **105**, 26003 (2014).
[35] G. L. Hart, V. Blum, M. J. Walorski, and A. Zunger, Nat. Mater. **4**, 391 (2005).
[36] R. Z. Khaliullin, H. Eshet, T. D. Kühne, J. Behler, and M. Parrinello, Nat. Mater. **10**, 693 (2011).
[37] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, Nat. Mater. **5**, 641 (2006).
[38] C. J. Pickard and R. Needs, Nat. Mater. **9**, 624 (2010).
[39] B. Huskinson, M. P. Marshak, C. Suh, S. Er, M. R. Gerhardt, C. J. Galvin, X. Chen, A. Aspuru-Guzik, R. G. Gordon, and M. J. Aziz, Nature (London) **505**, 195 (2014).
[40] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, Nat. Mater. **12**, 191 (2013).