Cite this: DOI: 00.0000/xxxxxxxxxx

Electronic Supplementary Information

A structural approach to vibrational properties ranging from crystals to disordered systems

Xin Tan,[a] Ying Guo,[b] Duan Huang,[a] and Ling Zhang*[b]

1 Simulation details

In our simulation, the aspect ratio of the box $L_x : L_y = 2 : \sqrt{3}$ to ensure perfect crystalline order, with periodic boundary conditions in both directions to minimize the boundary effect of the system[1,2]. The interaction between particle i and its nearest neighbor j is via a finite-range repulsive harmonic potential[3]:

$$V_{ij}(r) = \varepsilon(1-\frac{r_{ij}}{R_{ij}})^2\Theta(1-\frac{r_{ij}}{R_{ij}}), \tag{1}$$

where r_{ij} is the distance between the center of disk i and disk j, and R_{ij} is the sum of their radius. The length, energy and mass are measured in units of the small particle diameter σ,characteristic energy scale ε and mass m. Time and temperature have units $\tau = \sqrt{\varepsilon/m\sigma^2}$ and ε/k_b, where k_b is the Boltzmann constant.

All of our calculations are based on molecular dynamics simulator LAMMPS[4]. First, the system is relaxed to its minimum energy by using the FIRE algorithm[5] with a timestep dt=0.005τ (A long enough relaxation time is assured to reach the steady state.). Simulations proceed at a fixed temperature $T = 0.005$ with a Nose-Hoover thermostat and a fixed pressure $P = 100$ with a Berendsen barostat coupling within an NPT ensemble, while the NVT ensemble simulated under a fixed volume fraction. In all cases, data are collected for analysis after the system reaches the target temperature and pressure, and a timescale $\Delta t = 2\tau$ units of time to allow the system to complete a rearrangement, which means we dump a configuration every 400 timeteps.

2 ML details

2.1 Calculation of the structure functions

To construct an M-dimensional space $\mathbb{R}^{\mathbb{M}}$, two families of functions which describe the local structural environments around each particle are introduced to generate the array of M structure functions. As in literatures[6,7], we describe a structural function of M =160, including radial density and bond angle information of the whole system. The first family of structure functions describes the radial density properties:

$$G_X(i;\mu) = \sum_{j\in X} e^{-(R_{ij}-\mu)^2/2\sigma^2} \tag{2}$$

where X denotes a species whose density we wish to probe, μ is varied over the range from $0.3d_{AA}$ to $5.0d_{AA}$ in increments of $0.1d_{AA}$, where d_{AA} refers to the large grain diameter, thus a total of 94 radial density features are obtained for each particle. The second class of functions includes the bond angle information between particles within a distance ξ of one another. It is defined as:

$$\Psi_{XY}(i;\xi;\lambda;\zeta) = \sum_{j\in X}\sum_{k\in Y} e^{-(R_{ij}^2+R_{jk}^2+R_{ik}^2)/\xi^2}(1+\lambda\cos\theta_{ijk})^\zeta \tag{3}$$

where θ_{ijk} is the angle measured between R_{ij} and R_{ik}. By varying the parameters ξ, λ and ζ, the angular structure functions carry different information of a particle's angular neighborhood, without loss of generality, here we take the values shown by Cubuk et al.[6] to yield 66 additional angular structure features for each particle,where the different parameter combinations are shown in STable 1 1 . Taken together, we then construct an M-dimensional space $\mathbb{R}^{\mathbb{M}}$ with M=160 structure functions, and the local structure for a center particle i is captured as a point in the space $\mathbb{R}^{\mathbb{M}}$.

2.2 Details of the SVM method

The training set can be written as $\{(\mathbf{F}_1,y_1),...,(\mathbf{F}_n,y_n)\}$, where $\mathbf{F}_i$=$\{F_i^1,...,F_i^M\}$ are the values of the M structure functions that describe the local neighborhood of particle i. Correspondingly, a particle i is labeled as y_i=1 when it rearrange within the next time step and conversely y_i=-1. We then use the LIBSVM package to find a hyperplane $\omega\cdot F - b = 0$ separating rearranging and non-rearranging particles[8,9]. Since the data set is not linearly separable, it is impossible to find a hyperplane that exactly separates the two classes. Therefore, We use penalty parameter C and find the optimal hyperplane equation by minimizing

$$\frac{1}{2}\omega^{\mathbf{T}}\cdot\omega + C\sum_{i=1}^{N}\xi_i, \tag{4}$$

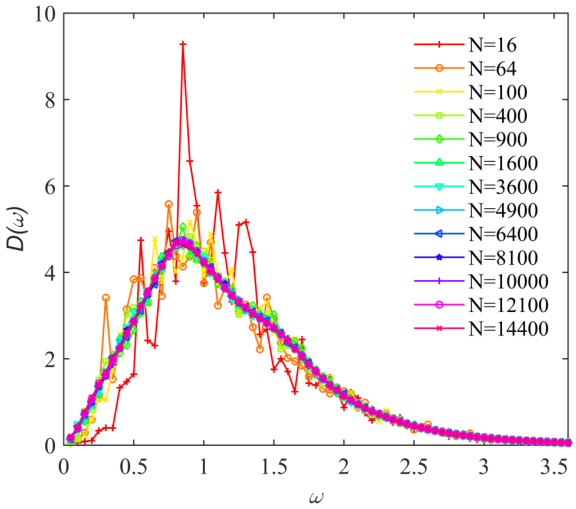
[a] School of Computer Science and Engineering, Central South University, Changsha 410083, China.
[b] School of Automation, Central South University, Changsha 410083, China. E-mail: lingzhang2019@csu.edu.cn

STable 1 Parameter combinations for angular structure function[6].

	ξ	ζ	λ
1	14.633	1	-1
2	14.633	1	1
3	14.638	2	-1
4	14.638	2	1
5	2.554	1	-1
6	2.554	1	1
7	2.554	2	-1
8	2.554	2	1
9	1.648	1	1
10	1.648	2	1
11	1.204	1	1
12	1.204	2	1
13	1.204	4	1
14	1.204	16	1
15	0.933	1	1
16	0.933	2	1
17	0.933	4	1
18	0.933	16	1
19	0.695	1	1
20	0.695	2	1
21	0.695	4	1
22	0.695	16	1

which is subject to the constraint that $y_i \cdot (\omega^{\mathbf{T}} \cdot \mathbf{F_i} + b) \geq 1 - \chi_i$ and $\chi_i \geq 0$, where χ_i are the slack variables. The parameters are optimized by cross-validation to obtain the best generalization ability. Hence we need to find the hyperplane that makes the margin between the training points for class y_i=1 and y_i=-1 maximized, and also the possibility of the points on the wrong side of their margin minimized. Once we have the hyperplane, we calculate the local structural functions of each particle, which is encoded as a point in the space $\mathbb{R}^{\mathbb{M}}$ to be classified as soft or hard[8,9].

3 Selection of the system size



SFig 1 The density of states $D(\omega)$ for different system size N. Note that here each curve is emsemble averaged for 5 realizations at $\eta = 1$ and $\phi = 0.943$.

To avoid the possibility of low-frequency artifacts produced by the boundary conditions, we analyzed the DOS for different system size N shown in SFig. 1. Evidently, there is a relatively strong fluctuation in the DOS curve when the system size is smaller than 1600, then with N increasing, the curve remains nearly unchanged, indicating that our result about the vibrational spectrum would not be affected by the finite size effect if $N > 1600$. Meanwhile, to satisfy the identification accuracy of the soft particles by SVM, we thus consider the system size as 10000 particles which is in agreement with literatures[10–13].

Notes and references

1 J. L. Beeby, *Physics Bulletin*, 1977, **28**, 325–326.
2 Allen and M. P., *Liquid Crystals*, 1990, **8**, 499–511.
3 C. S. O'Hern, S. A. Langer, A. J. Liu and S. R. Nagel, *Phys. Rev. Lett.*, 2002, **88**, 075507.
4 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.
5 E. Bitzek, P. Koskinen, F. Gähler, M. Moseler and P. Gumbsch, *Phys. Rev. Lett.*, 2006, **97**, 170201.
6 E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras and A. J. Liu, *Phys. Rev. Lett.*, 2015, **114**, 108001.
7 S. S. Schoenholz, *J. Phys.: Conf. Ser.*, 2018, **1036**, 012021.
8 C.-C. Chang and C.-J. Lin, *ACM Trans. Intell. Syst. Technol.*, 2011, **2**, 27:1–27:27.
9 T. Hastie, R. Tibshirani and J. H. Friedman, *The Mathematical Intelligencer*, 2005, **27**, 83–85.
10 S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras and A. J. Liu, *Nat. Phys.*, 2016, **12**, 469.
11 E. D. Cubuk, S. S. Schoenholz, E. Kaxiras and A. J. Liu, *J. Phys. Chem. B*, 2016, **120**, 6139–6146.
12 J. Rottler, S. S. Schoenholz and A. J. Liu, *Phys. Rev. E*, 2014, **89**, 042304.
13 R. J. S. Ivancic and R. A. Riggleman, *Soft Matter*, 2019, **15**, 4548–4561.