POLITECNICO DI TORINO
Repository ISTITUZIONALE

Classification and spatiotemporal correlation of dominant fluctuations in complex dynamical systems

Original
Classification and spatiotemporal correlation of dominant fluctuations in complex dynamical systems / Caruso, C., Crippa, M., Cardellini, A., Cioni, M., Perrone, M., Delle Piane, M., Pavan, G.M.. - In: PNAS NEXUS. - ISSN 2752-6542. - 4:2(2025). [10.1093/pnasnexus/pgaf038]

Availability:
This version is available at: 11583/2998443 since: 2025-03-20T13:54:40Z

Publisher:
National Academy of Science

Published
DOI:10.1093/pnasnexus/pgaf038

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)

26 November 2026

PNAS NEXUS

PNAS Nexus, 2025, 4, pgaf038
https://doi.org/10.1093/pnasnexus/pgaf038
Advance access publication 7 February 2025
Research Report

Classification and spatiotemporal correlation of dominant fluctuations in complex dynamical systems

Cristina Caruso [a], Martina Crippa [a], Annalisa Cardellini [b], Matteo Cioni [a], Mattia Perrone [a], Massimo Delle Piane [a] and Giovanni M. Pavan [a,b,*]

[a]Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy
[b]Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Polo Universitario Lugano, Campus Est, Via la Santa 1, Lugano-Viganello 6962, Switzerland
*To whom correspondence should be addressed: Email: giovanni.pavan@polito.it
Edited By Guruswami Ravichandran

Abstract

The behaviors of many complex systems, from nanostructured materials to animal colonies, are governed by local events/rearrangements that, while involving a restricted number of interacting units, may generate collective cascade phenomena. Tracking such local events and understanding their emergence and propagation in the system is often challenging. Common strategies consist, for example, in monitoring over time parameters (descriptors) that are designed ad hoc to analyze certain systems. However, such approaches typically require prior knowledge of the system's physics and are poorly transferable. Here, we present a general, transferable, and agnostic analysis approach that can reveal precious information on the physics of a variety of complex dynamical systems starting solely from the trajectories of their constitutive units. Built on a bivariate combination of two abstract descriptors, Local Environments and Neighbors Shuffling and *Time*Smooth Overlap of Atomic Position, such approach allows to (i) detect the emergence of local fluctuations in simulation or experimentally acquired trajectories of multibody dynamical systems, (ii) classify fluctuations into categories, and (iii) correlate them in space and time. We demonstrate how this method, based on the abstract concepts of local fluctuations and their spatiotemporal correlations, may reveal precious insights on the emergence and propagation of local and collective phenomena in a variety of complex systems ranging from the atomic- to the macroscopic-scale. This provides a general data-driven approach that we expect will be particularly helpful to study and understand the behavior of systems whose physics is unknown a priori, as well as to revisit a variety of physical phenomena under a new perspective.

Keywords: fluctuations, descriptors, spatiotemporal correlations, complex dynamical systems, local and collective dynamical events

Significance Statement

Many phenomena in Nature originate from local fluctuations or rare dynamical transitions that amplify in cascade events with drastic large-scale effects. From the brittle fracture of atomic materials to rapid turns in bird flocks and animal colonies, local events are dominant in complex systems but are also difficult to track and predict. Here, we describe a general and abstract data-driven analysis approach that, starting solely from experimental or simulation trajectories of the constitutive units, provides fundamental insights on the emergence and amplification of dominant fluctuations in complex multibody systems from the atomic- to the macroscale. Building on the general concepts of local fluctuations and spatiotemporal fluctuations' correlations, this method can help shedding light onto a variety of complex physical phenomena.

Introduction

Complex systems, as an ensemble of interacting units, are characterized by nontrivial internal dynamics that are often challenging to unveil. From atomic to macroscopic sizes, global trends often hide local events, fluctuations, or instantaneous perturbations that, although rare, can trigger interesting behaviors (1, 2). A beautiful example is the boson peak in amorphous solids (3), that is, an excess in the heat capacity whose microscopic origin is attributed to the emergence of dynamical defects or locally ordered structures in uniformly disordered systems (4, 5). Local perturbations of the atomic/molecular structure around a critical point are also at the origin of phase transition or nucleation phenomena (6, 7). Material properties, e.g. the brittle or ductile macroscopic deformation, are frequently controlled by the emergence and amplification of atomic-level defects (8). Even on larger scales, the collective behavior observed, for example, in active colloids converting selective energies into motion (9, 10), in bacterial colonies, fish banks, or bird flocks can be controlled by local perturbations or events that involve a restricted number of individuals (2, 11, 12). Understanding how local fluctuations may emerge, their correlation and amplification in space and time, and how they may determine emergent collective behaviors has, therefore,

Competing Interest: The authors declare no competing interests.
Received: January 20, 2025. **Accepted:** January 27, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

important implications in many fields and in the comprehension of complex systems in general.

Although crucial, localizing rare events and understanding the mechanisms at the root of emergent properties still present some unsolved challenges. Indeed, this requires to (i) unambiguously identify and (ii) classify such fluctuations according to their nature. In recent years, the increased volume of data generated by molecular simulations has led to a growing interest in the development of machine learning (ML)-based methods that are capable of identifying, within high-dimensional datasets, those local events revealing a disruption of structural and dynamical environments in both metals and soft matter (13–15). Numerous studies and consistent efforts have been focused on tracking local events/fluctuations rising from the time-space evolution of observed variables (2, 16, 17). Several methods, such as the change point detection (18), eigenvalue spectrum (19), critical slowing-down (20, 21), dynamic network biomarkers/markers (DNB) (22, 23), have been adopted to distinguish critical transition points between two steady states. Within this framework, the selection of an adequate *descriptor* emerges as a primary task for effectively detecting fluctuations. Successful results have been obtained from system-based descriptors, i.e. tailor-made on specific system properties (24, 25), thus dependent on prior knowledge about system's features. However, the goal is often to retrieve the dynamics of systems whose features are not a priori known, whereby general descriptors, e.g. based only on the mutual arrangements or movements of neighbor units, may show wider applicability (26–31). Among them, an advanced representation of atomic environments is provided by the Smooth Overlap of Atomic Position (SOAP) (32). Coupled with ML approaches, SOAP has enabled the characterization of diverse systems at equilibrium (33–35), including soft disordered assemblies (13, 30, 36–38) and metals (39–42). In the philosophy of tracking the time-space evolution of local events/fluctuations, two general descriptors have been recently developed: Local Environments and Neighbors Shuffling (LENS) (43) and *Time*SOAP (τSOAP) (44). By keeping track of units along molecular dynamics (MD) trajectories and any changes within their local environments, i.e. neighborhoods, LENS and τSOAP have been demonstrated to carefully characterize the dynamics of a wide range of systems albeit describing different local features (40, 43–45). LENS has been conceived to capture local reshuffling and permutation events that cannot be easily captured with other descriptors (e.g. SOAP) (43, 46). At the same time, LENS cannot capture local structural reconfigurations in the neighborhood of each unit in the system, being a *permutationally variant* and *structurally invariant* descriptor (43). On the other hand, τSOAP detects local variations in the order/disorder of the neighbors of every unit in the system. However, while keeping track of local structural reconfigurations, τSOAP cannot capture, e.g. local permutations and reshuffling: Vice versa, it is a *permutationally invariant* and *structurally variant* descriptor (44). The complementarity of LENS and τSOAP opens the opportunity to capture different types of local events and study their correlations, thus unveiling precious information on the physics of a variety of complex systems by an abstract and purely data-driven approach.

Here, we present an abstract and general analysis based on a bivariate time-series obtained via the combination of the **LE**NS and τSO**AP** descriptors, that we call **LEAP**. Using different case studies of systems on different scales and with different internal complexities, we demonstrate how the **LEAP** analysis allows to (i) detect dominant local fluctuations occurring in complex dynamical systems of any types, (ii) classify such fluctuations based on their physical nature, and (iii) correlate them in time and space. Simply starting from experimental or simulation trajectories of the constitutive units, such analysis can provide unique information on the physics of a variety of complex dynamical systems in an exquisitely agnostic and data-driven manner. We tested it to study various types of complex dynamical systems ranging from the atomic to the microscopic scale. The obtained results demonstrate how the presented bivariate approach can efficiently provide crucial insights, which can be useful to understand the mechanisms underpinning a variety of physical phenomena and, also, to trace them back to the spatiotemporal correlations between the local fluctuations that animate the systems. This provides us with a precious tool and a robust approach to explore and improve our understanding of complex systems whose physics is not a priori known, as well as to revisit known physical phenomena under a new light.

Results and discussion

The bivariate LEAP analysis

In this work, we demonstrate the potential of the bivariate **LE**NS-τSO**AP** (**LEAP**, for ease) time-series to unravel complex physical phenomena occurring in a range of systems with different inherent dynamical behaviors. As test cases, we study several types of prototypical systems in and out of equilibrium, dominated by rare-local, rather than collective nonlocal, events. We prove the generality of such an approach, analyzing both simulations and experimentally collected trajectories ranging from the atomic to the microscopic scales.

The philosophy of the analysis relies on considering systems as composed of N interacting units (IDs), and on monitoring each of them along the trajectory. We show how such "microscopic"—rather than "macroscopic/average"—approach allows to identify different types of local and collective events and, consequently, to correlate them in space and time thereby providing a thorough description of the inherent physics underlying disparate complex systems. Figure 1A shows a representation of two different types of local events that can be observed in the trajectories of complex systems. Considering ID 22 as an example, we report, on the left, an event where the unit is experiencing a "rigid" sliding on the other units, namely a change in its neighbors' identities (highlighted in magenta) without modifying its surrounding structural order: A permutationally variant, though structurally invariant, event. For the sake of simplicity, we refer to this type of event as a "local *diffusive* fluctuation." Differently, on the right, the same unit 22 undergoes a structural rearrangement of its closest neighbors, which yet preserve the same identities: Vice versa, a structurally variant but permutationally invariant event. This is an example of what we mean, for ease, with the term "local *structural* fluctuation." Noteworthy, in realistic systems, these two different types of events are not easy to capture, nor it is easy to distinguish between them. Such fundamental distinctions are relevant in the perspective of describing higher-scale dynamical events occurring in the systems.

Recently, LENS (43) and τSOAP (44) molecular descriptors have demonstrated to accurately detect these local dynamical events described above (see Materials and methods for details). As illustrated in Fig. 1B, LENS and τSOAP are perfectly tailored to capture purely *diffusive* (top) and purely *structural* (bottom) fluctuations, respectively. Although conceived to be different, these two novel descriptors are, at the same time, perfectly complementary to each other. Motivated by these observations, here we demonstrate how combining LENS and τSOAP into a bivariate time-series allows to get a unique characterization of distinct systems.

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

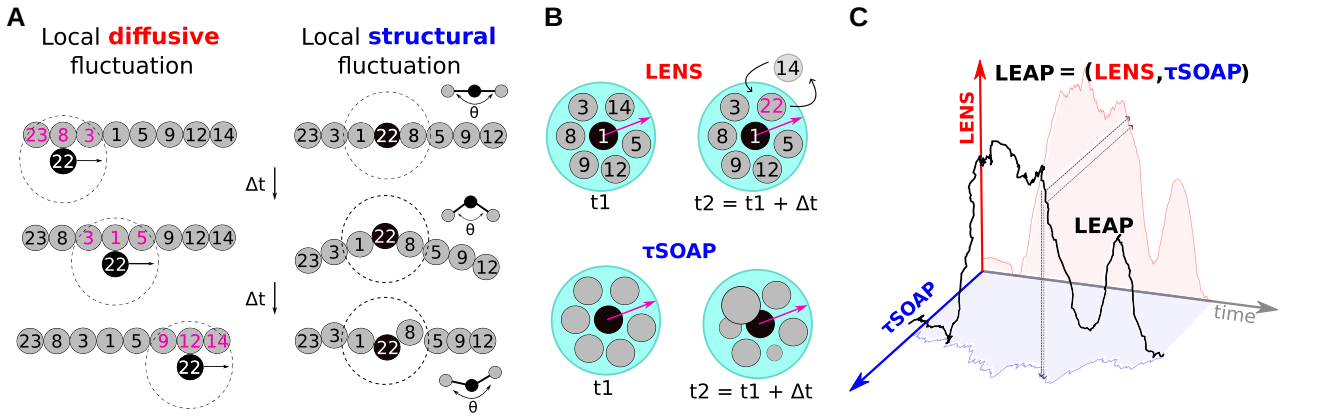


Fig. 1. The bivariate LENS-τSOAP (**LEAP**) analysis scheme. A) Schematic illustration of two types of local events, i.e. occurring among units (IDs) within a neighborhood (outlined with dot circles), that can be observed in the trajectories of complex systems. Local *diffusive* fluctuation (left): ID 22 changes its neighbor's identities yet preserving the structural order in its surrounding. Local *structural* fluctuation (right): ID 22 undergoes a structural rearrangement without changing the identities of its closest neighbors. B) Intrinsically distinct insights at the root of LENS and τSOAP molecular descriptors. Over time, and within a certain cutoff neighborhood, LENS (top) keeps track of changes in each ID's neighbor list, while τSOAP (bottom) monitors variations in ID's local order. C) **LEAP** as a bivariate time-series incorporating LENS and τSOAP components.

An illustration of the bivariate time-series (black) defined as **LEAP** = (*LENS*, τ*SOAP*), namely composed of LENS (red) and τSOAP (blue) components, is reported in Fig. 1C. Noteworthy, both LENS and τSOAP have been normalized from 0 to 1 in the **LEAP** definition (additional details are provided in the Materials and methods section). In the following, we prove the broad applicability of our analysis not only in classifying local events occurring in the system into different types of local fluctuations but also in correlating such different fluctuations in space and time. This allows unveiling the overarching behavior in several prototypical systems dominated by diverse inherent physics with a growing level of internal complexity.

LENS-τSOAP fluctuations at the ice/liquid water equilibrium interface

As a first case study, we show the results of the proposed analysis on the trajectories of water molecules in a system where ice and liquid water coexist in a dynamic equilibrium (Fig. 2).

Figure 2A (top) shows an MD periodic simulation box containing 2,048 **TIP4P/Ice** water molecules, of which 50% are arranged in the crystalline hexagonal ice configuration and the other 50% are in the liquid phase. This system is simulated for 100 ns, using a sampling time interval of $\Delta t = 0.001$ ns, at the melting temperature for the employed **TIP4P/Ice** water model (47, 48). Herein, we focus on 3 ns, extracted from the last part of the 100 ns-long MD simulation, and we compute the bivariate **LEAP** time-series for each water molecule (see Materials and methods for computational details). The bivariate time-series dataset, related to all the water molecules in the system, clearly shows a *linear* trend (Fig. 2A, Bottom). This means that, in a system characterized by two coexisting phases in an unstable equilibrium, the metastable interface fluctuations of atoms undergoing phase transitions can be described by concurrently occurring *diffusive* -detected by LENS —and *structural*—captured by τSOAP- fluctuations. The description of such metastable interface fluctuations from the bivariate **LEAP** time-series perspective is detailed hereafter.

Calculating the LENS and τSOAP time-series for each ID in the system, similar Kernel Density Estimation (KDE) distributions can be observed (Fig. 2B): Two peaks, at ~0.2 and at ~0.8 intensity, emerge in both signals. Therefore, two main domains presenting clear dynamic fingerprints can be identified in the system. As a result, an univariate Onion clustering analysis (45) carried out separately on LENS and τSOAP KDEs (see Fig. S1) detects, in both cases, ice (lowest descriptors' intensities), liquid water (highest descriptors' intensity), and solid–liquid interface (intermediate values). A detail of the LENS and τSOAP signals related to an example water molecule, ID 595, is highlighted on each plot in red and blue, respectively. The marked signals display a clear correspondence between the two time-series: At $t \sim 0.5$ ns, the ID 595 simultaneously undergoes a LENS and τSOAP transition. Confirming our previous findings (43, 44, 46), the correspondence of the identified domains is a first evidence that both LENS and τSOAP descriptors come up to be equivalent in identifying water states in phase coexistence.

The projection of the bivariate time-series onto its 2D LENS-τSOAP phase space is shown in Fig. 2C for each molecule in the system. In such 2D plot, each single point displayed in gray is the **LEAP** value related to a specific water molecule in a precise MD time step (3,000 time steps for 2,048 water molecules, for a total of $\sim 6 \times 10^6$ data points). Projected on this plot, the LENS-τSOAP path of the ID 595, chosen as a representative example, is colored from blue to yellow as time increases. The same water molecule (shown with the analogous color code, together with its neighborhood in light cyan) is depicted in the three MD snapshots in Fig. 2C while diffusing from the ice (1) to the liquid phase (3). During the transition, the water molecule moves on the 2D phase space along the diagonal meaning that, in this case, to every local structural rearrangement (τSOAP) corresponds an equivalent change in the local reshuffling dynamics (LENS). Similar results have been found by analyzing a longer (50 ns) MD trajectory, corresponding to the whole second half of such 100 ns-long MD simulation (see Fig. S2).

The pathway that molecules' trajectories follow onto such LENS-τSOAP phase space clearly determines the type of correlation between *diffusive* and *structural* fluctuations. To quantitatively describe this correlation, we define the following parameter:

$$\chi = \frac{LENS - \tau SOAP}{LENS + \tau SOAP}. \quad (1)$$

Over time, χ allows to measure to what extent the system's dynamics is driven by a specific type of fluctuation. When $\chi \sim 0$,

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

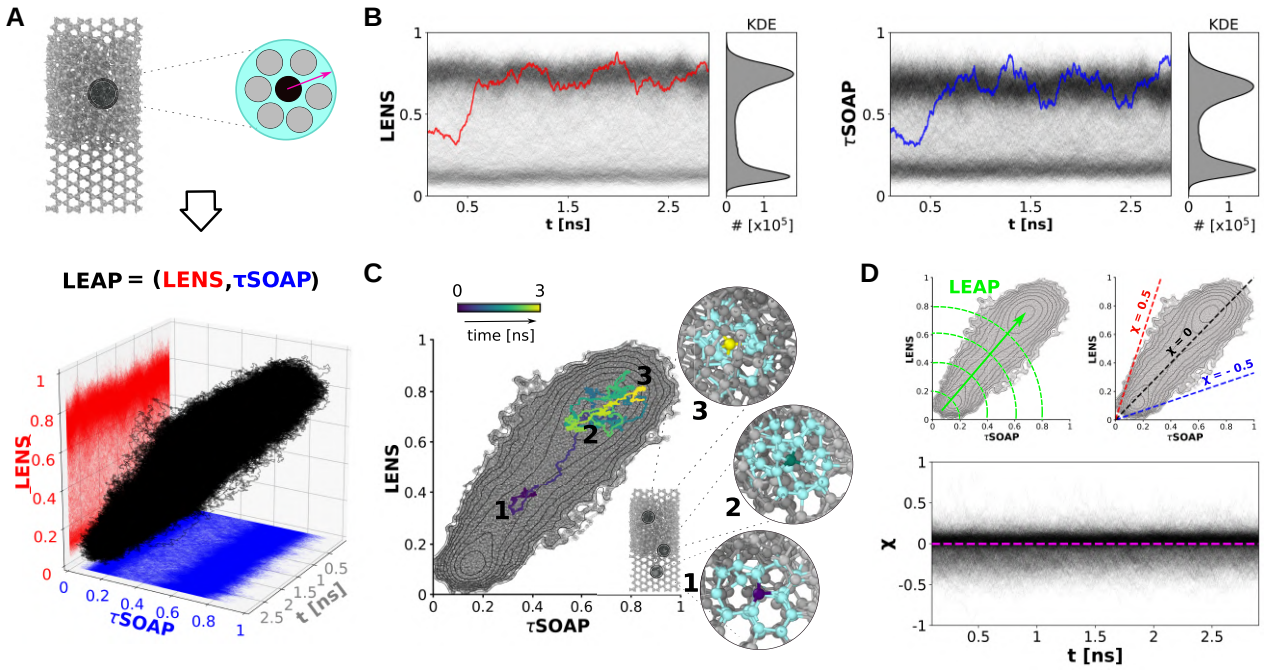


Fig. 2. LEAP analysis in Ice/liquid water phase coexistence. A) **LEAP** time-series dataset related to 3 ns extracted from a 100 ns-long MD trajectory (from 95 ns to 98 ns, sampling time Δt= 0.001 ns) composed of 2,048 water molecules (**TIP4P/Ice** water model) (47), whose 50% arranged in the crystalline hexagonal ice configuration and the remaining 50% in the liquid phase, coexisting in a dynamic equilibrium. Oxygen atoms (OW) are considered as centers to compute both LENS and τSOAP. B) LENS and τSOAP time-series, with the related Kernel Density Estimation (KDE) distributions, for all the water molecules in the system. Signals related to an example water molecule (ID 595) are highlighted on both LENS and τSOAP components. C) Projection of the whole **LEAP** dataset on the 2D LENS-τSOAP phase space (3,000 time steps for 2,048 water molecules, for a total of ~6×10^6 data points). The **LEAP** path related to the ID 595 is colored from blue to yellow as time increases. An analogous color code is used for the example water molecule (ID 595) in the representative MD snapshots taken at $t = 0$ ns (1), $t \sim 0.5$ ns (2) and $t \sim 2.8$ ns (3), while its neighborhood is colored in cyan. (D) χ parameter. Top: **LEAP** magnitude (left) and χ (right), represented on the 2D **LEAP** phase space. Bottom: χ parameter over time is reported for all the water molecules in the system, with the mean value ($\chi \sim 0$) highlighted in magenta.

LENS and τSOAP changes simultaneously occur. On the other hand, $\chi < 0$ and $\chi > 0$ indicate that the unit ID is interested by τSOAP-dominated and LENS-dominated fluctuations, respectively.

In such a specific case (the simplest we investigate, where LENS and τSOAP provide equivalent insights), the average χ is roughly 0, as displayed in magenta in Fig. 2D. This points out the correspondence between fluctuations in the structural order and in the neighbor identities while molecules move across phase transitions. However, there are systems exhibiting nontrivial behaviors which convey in nonsimultaneous *diffusive* and *structural* fluctuations and, hence, in a peculiar LENS-τSOAP mismatch, where further investigations are needed.

Nontrivial dynamical fluctuations on an atomic metal surface

As a second case study, we show the results of a **LEAP** analysis related to a metal copper (Cu) surface which exhibits a peculiar dynamics even well below the melting temperature (49, 50). The system consists of a portion of **Cu(211)** face-centered cubic (FCC) surface, containing 2,400 atoms, simulated for 150 ns using a DeepMD-based potential (51) which allows to reach length and time scales typically not affordable to DFT calculations (41). In particular, this **Cu(211)** FCC slab has been simulated at $T = 600$ K ($T \sim 1/3$ of the Cu melting temperature, see Materials and methods for simulation details), thus near the Hüttig temperature where metal surface dynamics can be observed (Fig. 3A). Recently, LENS-based analyses have highlighted, for the first time, the presence of sparse atoms sliding in a "rigid" manner on this surface (43, 46). Interestingly, such fascinating events had not been captured before by either looking at the average system's properties or using traditional pattern recognition analyses based on structural descriptors, e.g. SOAP (41). This is thus an ideal, prototypical example system where local (purely *diffusive*) fluctuations emerge only (predominantly) in the LENS dimension, while no evident fluctuations can be observed in the τSOAP one.

Herein, we consider the three top-most layers (995 atoms) of the **Cu(211)** FCC slab described above (see the snapshot in Fig. 3B). For each Cu atom considered, we analyze 24 ns extracted from the last part of the 150 ns-long MD simulation, sampled every $\Delta t = 12$ ps. In Fig. 3B (right), we report the whole time-series dataset including the 995 LENS, τSOAP, and **LEAP** signals plotted in red, blue, and black, respectively. Contrary to the water case of Fig. 2, here the **LEAP** dataset is far from being linear in the LENS-τSOAP space. Such a peculiar nonlinearity is investigated by first decoupling the **LEAP** time-series in its LENS and τSOAP components (see Fig. 3C, left). Clearly, the two descriptors display diverse trends, as also emphasized by the distinct fingerprints of LENS (in red, top-left) and τSOAP (in blue, bottom-left) corresponding to the same representative atom (ID 460). Indeed, while LENS deviates from the average behavior after ~15 ns, τSOAP slightly enhances. This represents an evident example where the two types of local events (diffusive and structural) do not simultaneously occur.

To deepen this concept, in Fig. 3C (right), we plot the whole **LEAP** data set projected onto LENS-τSOAP phase space (995 Cu atoms for 2,000 time steps, for a total of ~2×10^6 data points

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

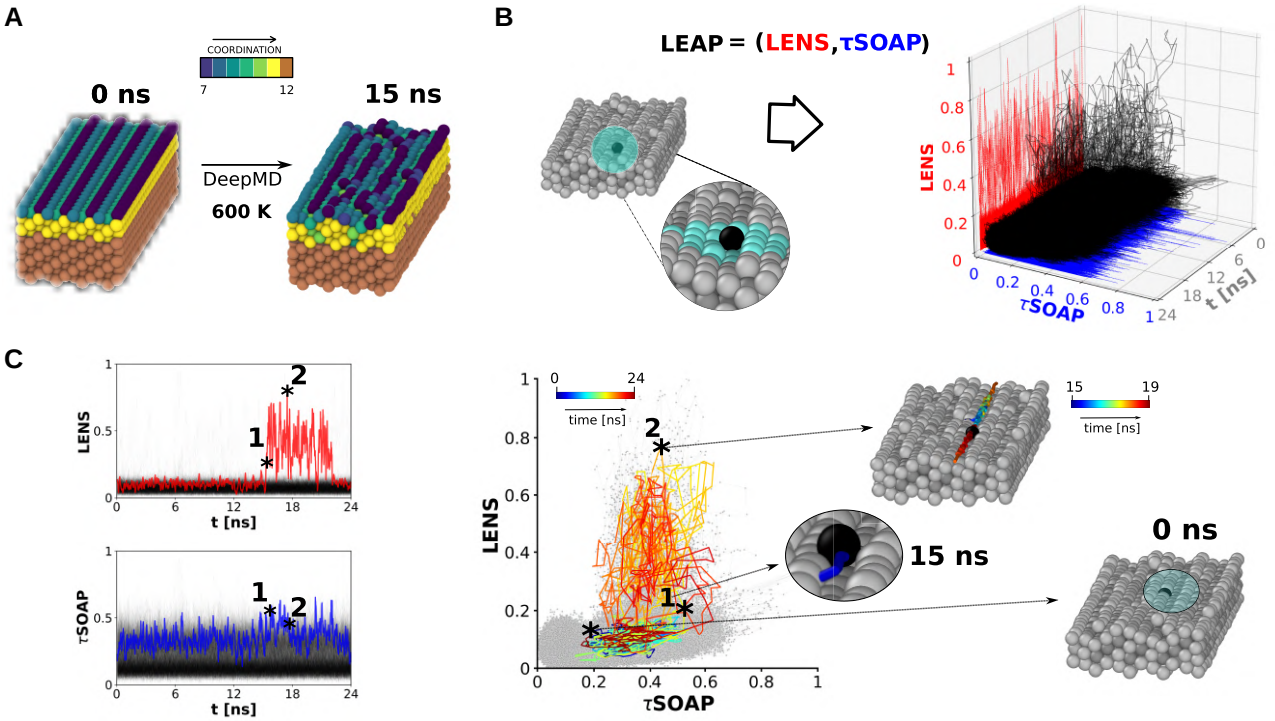


Fig. 3. LEAP fluctuations in a metal surface dominated by rare events. A) Portion of **Cu(211)** face-centered cubic (FCC) surface, composed of 2,400 atoms, simulated by DeepMD-based potential at $T = 600$ K. Two snapshots, taken at $t = 0$ ns and $t = 15$ ns, are colored according to their coordination number. B) **LEAP** time-series dataset related to the three top-most layers (995 atoms) of the **Cu(211)** FCC surface. Left: Focus on the local neighborhood representation of a central Cu atom and its closest neighbors. Right: LENS, τSOAP, and **LEAP** time-series plotted along the considered 24 ns of the MD trajectory. C) Left: LENS and τSOAP time-series, corresponding to all the 995 Cu atoms, are reported. The LENS and τSOAP signals of an example Cu atom (ID 460) are highlighted on each plot. Right: 2D **LEAP** dataset related to the Cu atoms on the three top-most layers (995 atoms for 2,000 time steps, for a total of $\sim 2{\times}10^6$ data points). Peculiar trajectory path, related to the example Cu atom (ID 460), colored from blue to red as time increases. Three MD snapshots taken at $t = 0$ ns (atom well incorporated in the surface), $t = 15$ ns (point (1), atom detaching from the surface) and 15 ns $< t <$ 19 ns (point (2), sliding after the detachment), are reported.

displayed in gray), and we select the peculiar path corresponding to the Cu atom ID 460. Coloring the selected trajectory from blue to red as a function of simulation time, we observe that during the early MD steps the atom path lies on a domain characterized by LENS <0.2 and τSOAP ~0.2, which describes the atom incorporated (and vibrating) in the top-most surface layer (see the MD snapshot at $t = 0$ ns). At roughly 15 ns, the **LEAP** values increase dramatically: Denoted by (1) in Fig. 3C, the τSOAP signal increases up to ~0.5, while LENS remains ~0.2. This indicates a first structural transition, namely the detachment of the atom from the surface (MD snapshot at 15 ns in Fig. 3C). From 15 to 19 ns, data show a sharp enhancement of the LENS component, while the τSOAP value remains substantially constant. Denoted by (2), we point out the atom sliding along the surface, after the detachment occurring at $t = 15$ ns. In such a "rigid" event, the sliding atom continuously changes the IDs of its neighbors (the LENS signal reaches ~0.8) while the structural order in the surrounding atoms exhibits only slight variations (sliding trajectory line displayed in the snapshot (2), colored according to the time evolution from 15 to 19 ns). After ~20 ns of MD, the atom comes back to the starting phase space domain (dark red line, overlapped to the dark blue one), meaning that it is stably reincorporated in the surface. This example reveals rare atomic events occurring on a metal surface at $T = 600$ K, which have a characteristic physics. Such an atomic surface diffusion is, in fact, characterized by sharp LENS signals, not accompanied by appreciable τSOAP fluctuations. At the same time, the trajectory analysis reveals that a sliding motion is conditioned by a preliminary structural transition, namely a sharp τSOAP variation with negligible LENS change (1). Interestingly, a sharp τSOAP variation occurs both when the atom jumps out and when it is reincorporated in the surface. The sliding mechanism follows a τSOAP⟶LENS⟶τSOAP sequence. As reported in Fig. S3, predominantly LENS or predominantly τSOAP fluctuations may emerge also for other atom ID trajectories in this system.

In cases like this, where the internal dynamical events are characterized by nonsimultaneous LENS and τSOAP fluctuations, the sequence and localization of events might be crucial to investigate the physics of the system. This opens, indeed, the possibility to understand (i) if there is a correlation between the number of structural fluctuations needed to generate diffusive LENS events (or vice versa) and (ii) how spatially correlated they have to be in such a way to generate a dynamical transition in the material.

Space and time correlations between different fluctuation types

A more detailed analysis of the dynamics of the **Cu(211)** FCC surface can be obtained by (i) classifying the local fluctuations emerging in the system into structural (τSOAP) and diffusive (LENS) fluctuations and (ii) analyzing their correlation in space and time. First, in Fig. 4A, we report the LENS and τSOAP time-series with the corresponding KDE distributions. An univariate Onion clustering analysis (45) of the individual LENS and τSOAP KDEs identifies *different* clusters in the two components, displayed in white, light red/blue, and dark red/blue colors (Fig. 4A). Except

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

for the crystalline bulk atoms characterized by the lowest descriptors' values (first KDE peaks and white atoms in both MD snapshots), the other domains do not match each other. For example, the two clusters corresponding to the highest LENS and τSOAP values (dark red and blue, respectively) do not include the same atoms (compare the red and blue atoms in the MD snapshots in Fig. 4A, right). While the τSOAP intermediate cluster ($0.25 \leqslant \tau$SOAP < 0.5) well identifies the top-most surface layer, this is not the same in the LENS component. This indicates that the LENS and τSOAP descriptors detect different types of dynamical events on this metal surface. Figure 4B, indeed, shows that dominant LENS and τSOAP fluctuations are well distinguishable and detectable as two distinct and separated domains in the **LEAP** 2D space. Projected on such phase space, the data points that are classified with a CI >95% as LENS outliers are colored in red (top-most red cluster in Fig. 4A), while the data points identified as τSOAP outliers with a CI >95% are shown in blue (top-most blue cluster in Fig. 4A).

It is interesting to look into the temporal correlations between LENS (red) and τSOAP (blue) fluctuations. Going back to the **LEAP** time-series, this classification (τSOAP vs. LENS fluctuations) allows us to color atoms based on the dynamical event they are experiencing over time (predominantly τSOAP, predominantly LENS or simultaneously occurring). On the right side of Fig. 4B, two example **LEAP** time-series (magnitude) for two representative atoms are reported in black: LENS and τSOAP (outliers) fluctuations are identified in red and blue bands, respectively. As clearly shown in both plots (see also Fig. S4), the selected IDs experience several structural fluctuations (sequence of blue bands) before sliding on the metal surface (red bands). Extending this analysis to all the atoms visiting the outlier (relevant fluctuation) regions, we can obtain a quantitative characteristic time estimation for the LENS and τSOAP fluctuations occurring in the system. For each atom crossing the τSOAP fluctuation region (blue domain in Fig. 4B), we compute the mean time interval (t) between two successive τSOAP events, thereby obtaining the cumulative distribution functions (CDFs) shown in Fig. 4C (see Materials and methods for technical details). The characteristic transition times ($\tau*$, Fig. 4C, left) of the structural reordering are reported by distinguishing atom trajectories visiting the τSOAP outlier region (dashed blue line) from their subset also experiencing the sliding event (solid blue line). In addition, the characteristic curve related to the sliding events is displayed (in red). It is worth noting that the probability to observe τSOAP (blue) or LENS (red) fluctuations in the system follows the typical Poisson distribution, which is expected for rare transition events. In detail, Fig. 4C (left) confirms the rare nature of sliding events which, indeed, happen less frequently than the structural reordering: 19 ns and 0.5 ns are the estimated characteristic times of sliding and rearrangement before sliding, respectively. Furthermore, the blue CDFs reveal that the characteristic transition time scale of structural fluctuations is slightly shorter when atom's rigid surface motion (LENS fluctuation) follows after ($\tau* = 0.3$ ns for the solid blue, while $\tau* = 0.5$ ns for the dashed blue line). This shows how, in this system, a local diffusive LENS event is always anticipated by a set of blue fluctuations occurring with a frequency that is ~65–70% higher than the normal one (namely, compared to the average frequency observed for blue fluctuations in general in the system). Figure 4C (right) also shows the number (~49) of τSOAP structural reordering transitions preceding the observation of every atomic sliding (LENS event) on the surface. This indicates that in a surface where structural fluctuations are ubiquitous, observing one rigid atomic sliding on the surface is a rare event (see also Materials and methods) that, statistically, is observed when one atom undergoes a set of τSOAP that are considerably more packed in time (more frequent) than what happens on average.

These data provide a mechanistic picture in terms of temporal correlation between the different types of fluctuations characterizing such a system. As a next step, we also investigate the spatial correlation between them. Figure 4D reports the probability density to observe structural events within a certain distance from diffusive events. Data show the presence of clear peaks, with the first one in correspondence of the interatomic distance (a ~3.6 Å in FCC Cu). This means that atoms, during the sliding motion, are close to IDs which undergo a reconstruction transition. A closer look at this phenomenon is provided in the Movie S1. For each ID in each MD time step, the plot of Fig. 4E shows (on the left) the relationship between the ID LENS value and the τSOAP mean value of its neighbors, and vice versa (on the right), between the ID τSOAP value and the LENS mean value of its neighbors (additional correlations are reported in Fig. S5). The two distributions show how, while LENS fluctuations require, in general, neighbors' τSOAP values higher than the average, τSOAP fluctuations do not necessarily need high neighbors' LENS values. Altogether, this data demonstrate how the sliding events identified as LENS fluctuations are (i) anticipated in time by structural (τSOAP) reconfigurations being, in intensity and frequency, higher than the average and (ii) characterized in space by local rearrangements of close neighborhoods.

The data reported herein demonstrate how classifying fluctuations into different categories and correlating them in space and time hold a great potential in revealing the physics underpinning complex system. The potential of the proposed approach can be emphasized by the comparison with other approaches often used to study the dynamics, e.g. the time-lagged Independent component analysis (tICA) (52, 53). By projecting the high-dimensional input space—typically coming from MD trajectories—to a low-dimensional linear subspace which maximizes the autocorrelation of the input coordinates, tICA is an outstanding method to shed light on the dynamics of many complex systems. Since conceived to best differentiate between slowly equilibrating populations, however, tICA might lead to discard the fastest degrees of freedom, such that the slowest timescales can be better estimated (see Figs. S6 and S7). As a result, the fastest transitions/fluctuations might not be correctly represented. This is exactly where the potential of the analysis herein presented comes in. LENS and τSOAP descriptors are, per se, already capable of preserving any changes of local environments, from the slowest to the fastest ones. Incorporating them into a bivariate time-series allows to take a further step, that is, the classification and correlation of dynamic domains of different nature (*diffusive* and *structural*). This is particularly relevant for characterizing the behavior of systems whose underlying physics is unknown, but also to revisit well-established physical phenomena, as shown in the following.

Plastic deformation of metals seen under the LEAP analysis

To prove the generality of the **LEAP** analysis, we tested it to revisit a well-known phenomenon controlled by the formation and amplification of local defects/fluctuations, namely the deformation and fracture of metals under tensile stress.

As a reference case, we consider a bulk of copper (Cu) FCC crystal containing 2,744 atoms and subjected to a constant strain rate at $T = 300$ K. Figure 5A shows the stress–strain curve alongside the potential energy profile obtained from the MD simulation of a periodic box using an Embedded atom method (EAM) potential

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

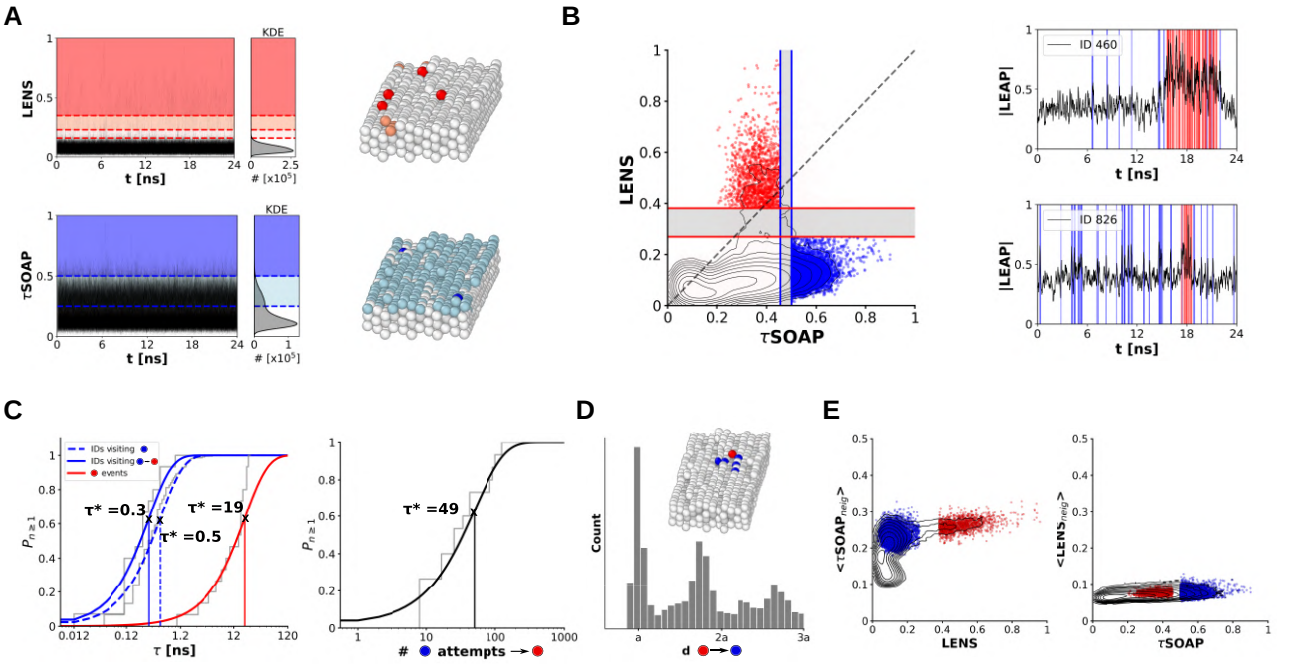


Fig. 4. Time-space correlation between diffusive and structural fluctuations using **LEAP**. A) LENS and τSOAP time-series, computed for the three top-most layers (995 Cu atoms) of the **Cu(211)** FCC surface, with the corresponding KDE distributions. Univariate Onion clustering (45) is applied. The identified clusters are reported using different colors in both LENS (top) and τSOAP (bottom). The MD snapshots reported on the right, taken at the same time step ($t \sim 11$ ns), are colored according to the identified clusters, respectively. B) Left: Projection of the **LEAP** dataset on the 2D LENS-τSOAP phase space. LENS and τSOAP outlier domains are colored in red and blue, respectively. Right: Time-series of the **LEAP** magnitude (black) for two example IDs. The transit through LENS and τSOAP outlier domains (red and blue regions shown on the left) are identified with red and blue bands, respectively. C) Characteristic time scale ($\tau*$) estimation for LENS and τSOAP fluctuations. Left: Cumulative distribution functions (CDFs) for the mean time interval (τ) between successive τSOAP fluctuations related to (i) trajectories visiting the blue region in (B) (dashed blue line), (ii) the subset of (i) crossing the red region in (B) after the blue one (solid blue line); (iii) CDF related to the time interval between LENS fluctuations occurring in the system (red line). Right: CDF for the number of τSOAP fluctuations preceding a LENS fluctuation. D) Space correlation between LENS and τSOAP fluctuations. x-axis: Distance between atoms lying in the LENS outlier region and atoms in the τSOAP outlier region, where (a) is the interatomic distance (a ~3.6 Å in FCC Cu). E) Left: ID LENS value vs. τSOAP mean value of its neighbors. Right: ID τSOAP value vs. LENS mean value of its neighbors. Blue and red points correspond, respectively, to the blue and red domains in (B).

(54)(see Materials and methods section for details). The plot clearly shows the transition from the elastic to the plastic deformation phase (after ~110 ns of MD and $\epsilon > 0.11$). We use the **LEAP** analysis to gain insights into one reference plastic event (occurring at ~130–145 ns, as highlighted in Fig. 5A). Figure 5B shows a zoom in the temporal evolution of the **LEAP** (magnitude) time-series, revealing a pronounced peak at ~138.88 ns. Interestingly, the decomposition of such **LEAP** signal in its LENS (Fig. 5C) and τSOAP (Fig. S8) components demonstrates that such **LEAP** fluctuation is governed by the LENS component, that is, by diffusive events (neighbors reshuffling) rather than by structural reconfigurations. This is consistent with the traditional view of plastic deformations in metals that proceed predominantly through dislocation planes, as also supported by the findings in Fig. 5D, where temporal and spatial correlation analyses are reported. The data demonstrate the collective and concerted nature of the fluctuations that control the deformation of the material under tensile stress after entering the plastic region. In particular, the plastic deformation, controlled by the motion of the dislocation planes respect to each other, is reflected by sharp LENS fluctuations that are substantially simultaneous in time and correlated in space: They occur within ~5 ps (see Fig. 5D) and with a distance corresponding to the interatomic spacing (a), as shown in the inset of Fig. 5D. Figure 5E shows that the dislocation first nucleates in the system and, when the nuclei grow, collective motions of planes emerge (identified by the red domains in the central green rectangle); Then, after the sliding of the dislocation planes, the system comes back and rests. Lastly, Fig. 5F provides a 3D representation of the sliding regions, identifying rows (in red) of high-LENS atoms corresponding to the dislocation vectors.

The system herein discussed, used as a proof of concept case, proves the reliability of the **LEAP** analysis scheme. This example demonstrates how, in a purely agnostic manner and simply relying on the concept of local fluctuations and their spatial and temporal correlations, it is possible to revisit and reconstruct nontrivial complex and collective phenomena such as those underpinning the well-known dislocation motion that control the plastic deformation in metals.

Into the physics of complex active matter systems with LEAP

Lastly, we show how a **LEAP** analysis can provide insights into the physics of complex active matter systems whose trajectory may be obtained experimentally rather than by MD simulations.

As a case study, we use Quincke rollers: Dielectric colloidal (polystyrene) particles immersed in a conducting fluid subjected to an external vertical DC electric field (**E**), as illustrated in Fig. 6A (17). In this system, an increase of **E** enhances the fluctuations of particle polarization vectors **P**, promoting particles rolling on the *xy* plane. Recent studies have demonstrated how interesting collective phenomena, such as activity waves, vortices, etc., may emerge and propagate in a quiescent population of such colloidal particles in conditions where **E** is weaker than $\mathbf{E}_c$, the

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

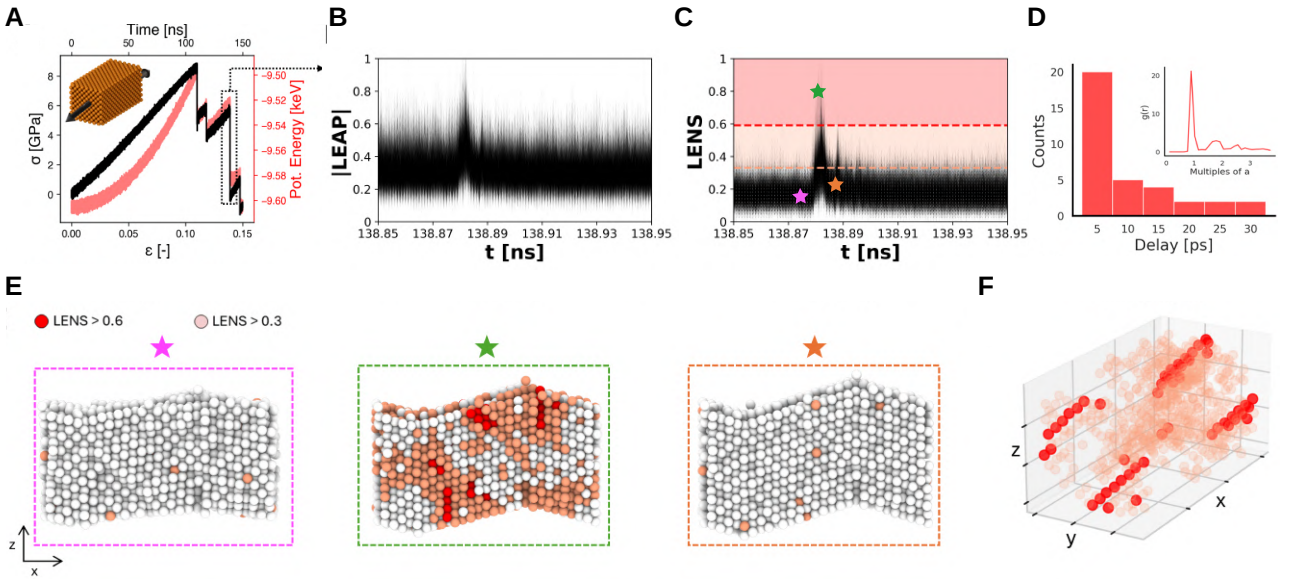


Fig. 5. LEAP analysis of bulk metal (Cu) during constant strain rate at $T = 300$ K. A) Stress–strain curve (black) over time with corresponding potential energy profile (red), indicating the regions of elastic and plastic deformation. Inset showing the simulated bulk copper structure, with arrows indicating the direction of strain. B) Time-series of the LEAP signal identifying one detail of the plastic phase highlighted in (A). C) Time-series of the LENS signal within the same timeframe, color-coded by intensity thresholds (LENS >0.3 in pink, LENS >0.6 in red); stars mark snapshots detailed in (E). This panel reveals how changes in LENS values dominate the LEAP signal over time. D) Temporal and spatial correlation analysis of high-intensity LENS signals, showing that most LENS spikes occur within 5 ps of each other and tend to cluster spatially (the inset shows the radial distribution function, $g(r)$, of atoms with high LENS signals, illustrating their spatial proximity). E) Snapshots corresponding to the marked events in (C), color-coded to show the intensity of their LENS signals, highlighting both spatial and temporal correlations. F) 3D representation of atomic positions with high LENS signals, emphasizing regions of coherent plastic deformation, identified as dislocation lines.

threshold value promoting the motion of an isolated particle (17). Here, we consider the optical microscope movie reported by Liu et al. (17), where an activity wave emerges and flows directionally, through a field of view of 700 × 700 μm^2 containing n= 6,921 particles over 0.25 s of image acquisition. From this movie, we obtain a trajectory of $T = 312$ frames using the Python package Trackpy (see snapshot in Fig. 6A) (55). Of these, we retain only the first 200 frames in the analysis (the portion of the movie in which the wave passage is observed). From the IDs' positions along the trajectory, we compute the LENS and τSOAP time-series for each particle, as plotted in Fig. 6B (complete technical details are provided in Materials and methods).

In both time-series, the high-density peaks in the KDEs can be isolated from the low-density high-intensity domains depicted by the red and the blue regions of Fig. 6B (see Fig. S9 for technical details). The high-density peaks, characterized by low-intensity signals, are related to the characteristic diffusive (top) and structural (bottom) vibrations of the particles in the system, which represent the "background" LENS and τSOAP noise, respectively. Conversely, the low-density domains of the time-series identify fewer particles undergoing intense LENS or τSOAP fluctuations in the system (high-intensity signals). Coloring these LENS/τSOAP-fluctuating particles in each frame (in red and blue, respectively) allows visualizing the collective wave flowing left-to-right throughout the system (Fig. 6B: right). Both LENS and τSOAP components capture the wave propagation but not exactly in an identical manner.

The projection of the **LEAP** dataset on the 2D LENS-τSOAP phase space (6,921 particles for 200 frames, for a total of ∼ 1.3×10^6 data points, Fig. 6C) allows to distinguish four different domains based on the thresholds discriminating the relevant fluctuations from noise in the LENS and τSOAP dimensions (clustering details in Fig. S9). Coloring the trajectory frames of Fig. 6D according to such a classification shows the quiescent particles in the system as depicted in white (corresponding to the bottom-left region of Fig. 6C: LENS ⩽0.056 and τSOAP ⩽ 0.169). These are particles that just vibrate, whose motion constitutes the intrinsic noise in the system. The lilac domain, which includes all the particles in the core of the wave, corresponds to the region near the diagonal in the LENS-τSOAP phase space of Fig. 6C (LENS >0.056 and τSOAP >0.169). In this region, every particle undergoing a LENS fluctuation is also undergoing a τSOAP fluctuation (structural+diffusive fluctuations). The two most interesting domains are those colored in red (LENS >0.056 and τSOAP ⩽0.169) and blue (LENS ⩽0.056 and τSOAP >0.169), which identify particles undergoing fluctuations that are LENS- and τSOAP-dominated, respectively. In particular, the red cluster (LENS fluctuations) corresponds to the wavefront (the wave moves left-to-right, see also Movie S2), a region where LENS-dominated fluctuations appear as simultaneous in time and correlated in space. This indicates that the first event that anticipates the passage of the wave in this system is a LENS event (namely, every particle that becomes part of the wave, first undergoes a LENS fluctuation).

We can follow the trajectory of every particle in the system over time on the LENS-τSOAP phase space. Figure 6E shows one prototypical trajectory of one example particle (ID 106): Very similar paths are observed for the other particles in the system. In general, the trajectories follow a characteristic path on the 2D **LEAP** phase space, where the particles move sequentially red ⟶ blue ⟶ lilac (where they spend ∼20–40 ms) ⟶ blue ⟶ white (where they return to a quiescent configuration). In this system, where the raw trajectory obtained from the movie can be sampled at most every 1 ms (17), the first LENS events (i.e. the passage through the red region) are very short-lived and last at most ∼1–2 ms. For every particle in the system, we calculated the χ parameter (as defined in Eq. 1) over time. Figure 6F reports the χ parameter over time for all the

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

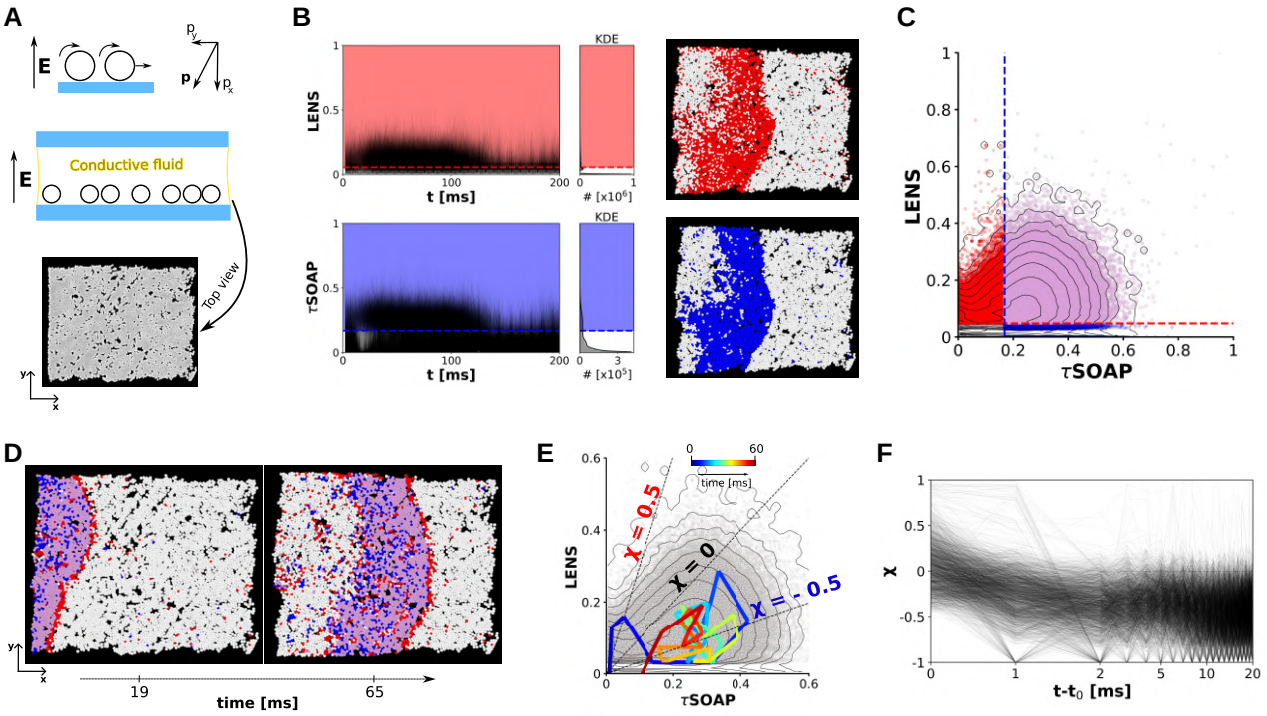


Fig. 6. LEAP analysis of a microscopic experimental system: Quincke rollers. A) Top: Schematic illustration of Quincke rollers, i.e. dielectric colloidal particles in a weakly conductive fluid and exposed to a vertical DC electric field, **E**. Bottom: Top view of the experimental system, extracted from an optical microscope movie (17). n= 6,921 particles are tracked for 312 ms, with a sampling time of 1 ms (45). B) LENS and τSOAP time-series related to all the tracked particles, with the corresponding KDEs. The first 200 ms are considered. The tails of both the KDEs are isolated: LENS threshold = 0.056 (dashed line), τSOAP threshold = 0.169 (dashed line). They allow visualizing the collective wave flowing left-to-right throughout the system, shown in red for LENS (top-right) and in blue for τSOAP (bottom-right) in the two example snapshots (t = 50 ms). C) Projection of the **LEAP** dataset on the 2D LENS-τSOAP phase space (6,921 particles for 200 time steps, for a total of $\sim 1.3\times10^6$ data points). Based on the thresholds reported in (B), four different domains can be distinguished. D) Two example snapshots are colored according to the identified domains shown in (C). E) Representative example trajectory (particle ID 106) following a characteristic path on the **LEAP** 2D phase space. Sixty milliseconds of the trajectory are shown, colored as time increases. F) χ parameter related to all the particles in the system remaining in the core of the wave (lilac domain in (C)) for at least 20 ms (2,680 particles in total). t_0 corresponds to the first instant in which every particle enters the red region in (C), that is, the front of the wave.

particles in the system that remain in the core of the wave (lilac region) for at least 20 ms. All trajectories are phased in such a way that the point in $t - t_0 = 0$ (in the x-axis) corresponds to the instant in which every particle undergoes the first LENS event, entering the front of the wave. The data show how for all particles entering the wave, the first LENS-dominated fluctuations annihilate within ~3 ms (the majority of them, within 1 ms). Only sparse new LENS fluctuations are seen to emerge after 5 ms of observation. These identify the sparse, local (nonspatially correlated) LENS fluctuations that may emerge in the body of the wave (red particles within the lilac domain in the snapshots of Fig. 6D). Interestingly, when the LENS fluctuations annihilate, every particle undergoes a sharp τSOAP fluctuation ($\chi \sim -1$). This demonstrates how all the particles in the system follow a red ⟶ blue ⟶ lilac pathway (like the trajectory of ID: 106 shown in Fig. 6E).

These results provide a demonstrative example of how such abstract detection–classification–correlation of local fluctuations **LEAP** approach can provide precious insights into the behavior and the internal phenomena occurring in complex dynamical systems whose physics is unknown a priori.

Conclusions

In this article, we demonstrate how the combination of complementary general descriptors allows elucidating the physical behavior of diverse complex systems by leveraging the concept of local fluctuations. In the philosophy of monitoring units along the trajectory, the bivariate **LEAP** time-series incorporates two complementary fingerprints, LENS and τSOAP, which detect, respectively, permutationally variant/structurally invariant (*diffusive*) and permutationally invariant/structurally variant (*structural*) local events (Fig. 1). Essentially, they are fluctuations that typically characterize the local neighborhoods in the trajectory of complex systems.

Trajectories linearly moving on the 2D **LEAP** phase space identify molecules (or, in general, units) whose neighbors undergo local structural rearrangements while changing, at the same time, their local dynamics (in terms of dynamical reshuffling). This is the case, for instance, of phase transitions in the MD simulation of ice/liquid water at phase coexistence (Fig. 2), where structural and diffusive local events simultaneously occur. However, some physical phenomena may be governed by predominantly diffusive (LENS-dominated) or by predominantly structural (τSOAP-dominated) fluctuations, namely by local dynamical events that can be captured exclusively by the LENS or τSOAP component of our **LEAP** bivariate time-series. Interestingly, such phenomena can be described by strongly nonlinear trajectory paths on the 2D **LEAP** phase space, thus identifying nontrivial dynamical events (from Figs. 3 to 6). In particular, LENS-dominated fluctuations identify "rigid" diffusive dynamical events, where units slide while changing, at every time step, the identity of their closest neighbors which, nonetheless, preserve their structural order.

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

On the other hand, τSOAP-dominated fluctuations identify events where units do not dynamically reshuffle within close neighborhoods, but structurally rearrange (local rattling events). Leveraging such a classification in different types of fluctuations allows correlating such different events in space and time, obtaining precious information for understanding a variety of phenomena. For example, in the case of the atoms sliding on the **Cu(211)** FCC surface seen in Figs. 3 and 4, our data demonstrate how every Cu atom undergoing a surface diffusive LENS event (Fig. 4B: in red) sees a drastic increase in the frequency of the structural τSOAP fluctuations (blue) preceding such event (Fig. 4C: frequency augmented by ∼65% compared to the average τSOAP fluctuations in the system).

At the same time, the spatial correlation between the local events is also fundamental. In several cases discussed herein, it is interesting to observe where local fluctuations may occur. Besides the temporal one, a spatial relationship may emerge between diffusive and structural fluctuations, e.g. in the case of **Cu(211)** FCC surface (Fig. 4D, E). Or even, local fluctuations may occur simultaneously in time and well localized in space. This identifies collective events, as demonstrated by the LENS-dominated fluctuations which describe the sliding of the dislocation planes in metals entering the plastic region (Fig. 5). Lastly, the active matter system case study proves how such a data-driven approach, based on the very simple and general concept of local events and their correlations, may help in elucidating the behavior of complex systems whose trajectories are experimentally obtained (Fig. 6). Our **LEAP** analysis on Quincke rollers, indeed, unveils a well-defined sequential mechanism followed by all the particles involved in the wave passage: First, colloids undergo a local LENS-dominated fluctuation; then, the τSOAP-dominated fluctuations grow in the system; finally, particles return to a quiescent state *via* τSOAP fluctuations (Fig. 6E, F). This provides relevant insights into the origin and mechanisms underpinning the evolution and annihilation of such a phenomenon.

With a similar spirit as some causality detection methods (56–61), these results demonstrate how such abstract **LEAP** analyses can provide crucial insights useful to understand the mechanisms underlying a variety of physical phenomena and to predict their emergence in the system in space and time. In general, we envisage that this **LEAP** analysis approach will constitute a precious tool to explore and understand complex systems whose physics is not a priori known, as well as to revisit known physical phenomena under a new light.

Materials and methods

Trajectories and preprocessing

Complete data and details concerning all molecular models and simulation parameters used to get the MD trajectories, as well as the complete **LEAP** analysis code, are available at https://github.com/GMPavanLab/LEAP and at https://doi.org/10.5281/zenodo.14851069.

All the setup details described in the following have been summarized in the Table S1.

Ice/liquid water phase coexistence

The atomistic ice/liquid water phase coexistence system is simulated employing the direct coexistence approach at the solid/liquid transition temperature, as recently used (43, 44, 46). The **TIP4P/ICE** water model (47) is chosen to model both the ice I_h and the liquid water phase. In the coexistence model, the two phases are put in contact in the same simulation box. In the initial configuration, half of the water molecules ($n = 1{,}024$) are in the solid ice phase, the other half ($n = 1{,}024$) in the liquid phase. The MD trajectory analyzed herein is obtained by simulating the system at constant pressure (1 atm) at the melting temperature for the employed water model ($T = 267.5$ K). The v-rescale thermostat with a time constant of $t = 20$ ps is used. For the pressure coupling, the c-rescale barostat is employed with a time constant of $t = 20$ ps (48). The GROMACS software is used (62). A 100 ns-long MD production run is performed using the same setup of Ref. (44), with a sampling time interval of $\Delta t = 0.001$ ns and an integration time step of 2 fs. The simulation is carried out in semiisotropic conditions, applying the pressure only in the direction perpendicular to the ice/water interface, thus reproducing the strictly correct ensemble for the liquid–solid equilibrium simulation according to the direct coexistence technique (63). The results reported in Fig. 2 are obtained by extracting, from the 100 ns-long simulation, 3 ns (from 95 ns to 98 ns, for a total of 3,000 frames). Firstly, the MD trajectory has been preprocessed by considering only the oxygen atoms (OW) of the water molecules in the system as representative centers in the **LEAP** analysis, i.e. for the calculation of LENS (43) and τSOAP (44) time-series. For each of the 2,048 oxygen atoms in the system, LENS and τSOAP are computed on the sampled MD configurations using a cutoff of 6 Å (second minimum of the OW-OW radial distribution function, thus enclosing the first two solvation shells) and smoothed using a moving average with width = 0.2 ns (200 frames). This analysis has been repeated also analyzing the whole second half of the same 100 ns-long MD trajectory, i.e. from 50 to 100 ns, using a sampling time interval of $\Delta t = 0.1$ ns (for a total of 500 frames) and a smoothing width = 8 ns (80 frames), as reported in Fig. S2.

FCC Cu(211) surface

The atomistic model of copper FCC surface **Cu(211)** studied herein is composed of $n = 2{,}400$ atoms arranged in eight layers that model a portion of an infinite surface through periodic boundary conditions. A 150 ns-long MD trajectory is obtained using a deep neural network potential of the **Cu(211)** surface built training a Neural Network with the DeepMD platform (51) on DFT data (for details, see Ref. (41)). The MD simulation is carried out at $T = 600$ K and with a sampling time interval of $\Delta t = 6$ ps using the LAMMPS software (64). The MD simulation setup is described in detail in Ref. (41). For the analysis reported in Figs. 3 and 4, 24 ns are considered (96–120 ns, for a total of 2,000 frames extracted every $\Delta t = 12$ ps along the MD trajectory). The analysis is conducted on the three top-most layers (995 atoms), since we are interested in the dynamics of the surface and most of the bulk remains substantially immobile during the MD simulation (41). For each of the 995 surface Cu atoms, a cutoff of 6 Å is used to compute the LENS and τSOAP descriptors using a Δt time lapse of 12 ps. Both signals are smoothed by a moving average, using a width = 120 ps (10 frames).

Plastic deformation of bulk Cu

The third system studied is a bulk of copper (Cu) FCC crystal composed of 2,744 atoms (see Fig. 5) and simulated at $T = 300$ K. The MD simulations are conducted with the LAMMPS software package (64) using an EAM potential specifically designed for copper (54). The MD trajectories are saved using a sampling time interval of $\Delta t = 5$ ps. This potential has been proved to be reliable in mimicking the phenomena occurring within these materials during a

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

tensile fracture test (e.g. it reproduces a Young modulus for copper consistent with the experiments, etc.) (65). The system undergoes minimization and then equilibration at 300 K for 2 ns, employing a Nose–Hoover thermostat and barostat to control the temperature and pressure (66). A time step of 1 fs is used in the simulation. Following equilibration, a uniaxial tensile deformation is applied along the x-direction with a strain rate of 0.1% strain per ns. The pressure in the y- and z-directions is maintained at zero during this deformation. The deformation simulation has been carried out for a total of 150 ns. The **LEAP** analysis is conducted for the entire simulation duration. However, due to the nature of the specific events being investigated, a brief time interval of 0.1 ns (from 138.85 ns to 138.95 ns) is reported in Fig. 4, showing details of the signals corresponding to the raw (**LEAP** and LENS) data. A cutoff of 8 Å is used in the **LEAP** analysis, which has been found as the best compromise between computational efficiency and the amount of information retained.

Experimental Quincke rollers trajectory

The Quincke rollers trajectory analyzed (in Fig. 6) is obtained from an optical microscope movie published by Liu et al. (17). Employing image recognition and the Trackpy tracking code (55), the *x* and *y* coordinates related to 6,921 colloidal particles for 312 consecutive frames have been obtained, as described in detail in Ref. (45). Since the collective propagation occurs in the first part of the trajectory, in the analysis we considered the first 200 consecutive frames (0–200 ms). For each particle in the system, LENS and τSOAP are computed using a cutoff of 40 pixels ($\sim 56\mu$, that takes into account the neighborhood at least up to the third minimum of the radial distribution function). A moving average is used to smooth the signal with width = 2 ms (two frames).

LENS and τSOAP data analysis

For each individual representative center *i* in each system, the LENS (43) and τSOAP (44) values are computed over time along the trajectory.

As detailed in Ref. (43), being C_i^t an array containing all the individual identities (IDs) of the particles/individuals surrounding the center *i* within a sphere of radius r_{cut} at the time step t, the LENS value, indicated by δ_i, is defined as

$$\delta_i^{t+\Delta t} = \frac{\#(C_i^t \bigcup C_i^{t+\Delta t} - C_i^t \bigcap C_i^{t+\Delta t})}{\#(C_i^t + C_i^{t+\Delta t})}, \quad (2)$$

where Δt is the time interval between two consecutive sampled time steps. The first and the second terms of the numerator are the mathematical union and intersection, respectively, of the neighbor IDs within r_{cut} from the center *i* at the time *t* and $t + \Delta t$. Thus, for each individual center *i*, $\delta_i(t)$ monitors the *i*th local environment changes in terms of neighbor identities/individuals along the trajectory, ranging from 0 to 1 for persistent to highly dynamic neighborhoods, respectively.

Indicated by λ_i, the instantaneous τSOAP value is defined as

$$\lambda_i^{t+\Delta t} \propto \sqrt{2 - 2\mathbf{p}_i^t \mathbf{p}_i^{t+\Delta t}}, \quad (3)$$

where $\mathbf{p}_i^t$ is the full SOAP feature vector associated to the *i*th individual center within r_{cut} at the time step t, as described in detail in Ref. (44). In a nutshell, $\lambda_i(t)$ tracks the variations of the *i*th SOAP vector over time, that is, to what extent the atomic environment related to each center in the system changes at every consecutive time interval Δt in terms of SOAP power spectrum.

For each system's individual center, therefore, two time-series were obtained, LENS ($\delta_i(t)$) and τSOAP ($\lambda_i(t)$), tracking over time the neighbor list and the structural variations, respectively. In order to reduce the noise, both $\delta_i(t)$ and $\lambda_i(t)$ time-series are smoothed by using a moving average with different time widths depending on the analyzed system (apart where explicitly stated otherwise). A time width of 200 frames is used for the ice/liquid water phase coexistence. To better detect the emergence of rare and often short-time fluctuations, smaller widths are chosen for the other systems. A time width of 10 and 2 frames have been chosen for the **Cu(211)** surface and the experimental Quincke rollers system, respectively. Thus, smoothed $\langle\delta_i(t)\rangle$ and $\langle\lambda_i(t)\rangle$ time-series were obtained. No smoothing is applied in the case of bulk Cu during constant strain rate. For the sake of simplicity, we refer to $\langle\delta_i(t)\rangle$ as $\delta_i(t)$ and to $\langle\lambda_i(t)\rangle$ as $\lambda_i(t)$.

For each trajectory, both $\delta_i(t)$ and $\lambda_i(t)$ are first normalized from 0 to 1 and, then, combined in the **LEAP** bivariate time-series defined as

$$\mathbf{LEAP}_i(t) = (\delta_i(t), \lambda_i(t)). \quad (4)$$

$\mathbf{LEAP}_i(t)$, thus, is a bicomponent array related to the individual *i*, which keeps track of neighbor's identity changes (first component) and of structural rearrangement (second component) over time.

Characteristic time estimation for LENS and τSOAP fluctuations

A quantitative characteristic time estimation of the diffusive and structural events occurring in the FCC **Cu(211)** surface is provided in Fig. 4C. In detail, for each Cu atom experiencing a τSOAP fluctuation, i.e. crossing the τSOAP outlier domain depicted in blue in Fig. 4B, we compute the mean time interval (τ) between two successive structural events. Thus, we obtain (i) the probability distribution function of the structural fluctuation frequency. For those atoms visiting the LENS outlier domain (Fig. 4B, red), the τSOAP fluctuations are only considered before the diffusive (LENS) event. Furthermore, (ii) a subset consisting of Cu atoms that visit the LENS outlier domain after undergoing τSOAP fluctuations is also considered. This allows to distinguish τ in the atoms undergoing LENS fluctuations after several structural rearrangements. The mean time intervals (τ) between consecutive τSOAP events are used to build the cumulative distribution functions (CDFs) $P_{n\geqslant 1}$:

$$P_{n\geqslant 1} = 1 - e^{-\tau/\tau*}, \quad (5)$$

where $\tau*$ is the characteristic time scale of structural (τSOAP) fluctuations. The obtained CDFs are thus reported Fig. 4C, on the left (dashed blue line for (i), solid blue line for the subset (ii)). In addition, (iii) the probability distribution of LENS fluctuations occurring in the system is obtained and the related CDF is reported with a red line (Fig. 4C, left): In this case, $\tau*$ represents the characteristic time scale of diffusive (LENS) fluctuations. The same approach is used to compute the CDF related to the number of structural fluctuations needed before observing a diffusive event (see Fig. 4C, right). All these $P_{n\geqslant 1}$ distributions turned out to be well fitted by the typical Poisson distribution expected for rare events.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

Funding

G.M.P. acknowledges the support received by the European Research Council (ERC) under the Horizon 2020 research and innovation program (grant agreement no. 818776 - DYNAPOL).

Author Contributions

G.M.P. designed research; C.C., M.C., A.C., M.C., and M.P. performed research; C.C., M.C., A.C., M.C., M.P., M.D.P., and G.M.P. analyzed data; C.C, A.C., and G.M.P. wrote the article.

Data Availability

Details on the molecular models and on the MD simulations, and additional MD data are provided in the Supplementary material. Complete details of all molecular models used for the simulations, and of the simulation parameters (input files, etc.), as well as the complete LEAP analysis code, are available at: https://github.com/GMPavanLab/LEAP and at https://doi.org/10.5281/zenodo.14851069.

References

1 Cavagna A, *et al.* 2023. Natural swarms in 3.99 dimensions. *Nat Phys*. 19(7):1043–1049.
2 Ballerini M, *et al.* 2008. Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proc Natl Acad Sci U S A*. 105(4):1232–1237.
3 Nakayama T. 2002. Boson peak and terahertz frequency dynamics of vitreous silica. *Rep Prog Phys*. 65(8):1195.
4 Hu Y-C, Tanaka H. 2022. Origin of the boson peak in amorphous solids. *Nat Phys*. 18(6):669–677.
5 González-Jiménez M, *et al.* 2023. Understanding the emergence of the boson peak in molecular glasses. *Nat Commun*. 14(1):215.
6 Sosso GC, *et al.* 2022. The role of structural order in heterogeneous ice nucleation. *Chem Sci*. 13(17):5014–5026.
7 Tao P, Du C, Xiao Y, Zeng C. 2023. Data-driven detection of critical points of phase transitions in complex systems. *Commun Phys*. 6(1):311.
8 O'Ghaffari H, *et al.* 2023. Microscopic defect dynamics during a brittle-to-ductile transition. *Proc Natl Acad Sci U S A*. 120(42):e2305667120.
9 Reyes Garza R, Kyriakopoulos N, Cenev ZM, Rigoni C, Timonen JV. 2023. Magnetic quincke rollers with tunable single-particle dynamics and collective states. *Sci Adv*. 9(26):eadh2522.
10 Fasano M, *et al.* 2019. Thermally triggered nanorocket from double-walled carbon nanotube in water. *Mol Simul*. 45(4–5):417–424.
11 Cavagna A, Giardina I. 2010. Large-scale behaviour in animal groups. *Behav Processes*. 84(3):653–656.
12 Bialek W, *et al.* 2014. Social interactions dominate speed control in poising natural flocks near criticality. *Proc Natl Acad Sci U S A*. 111(20):7212–7217.
13 Cardellini A, *et al.* 2023. Unsupervised data-driven reconstruction of molecular motifs in simple to complex dynamic micelles. *J Phys Chem B*. 127(11):2595–2608.
14 Kashiwa K, Kikuchi Y, Tomiya A. 2019. Phase transition encoded in neural network. *Prog Theor Exp Phys*. 2019(8):083A04.
15 Arnold J, Schäfer F. 2022. Replacing neural networks by optimal analytical predictors for the detection of phase transitions. *Phys Rev X*. 12(3):031044.
16 Cavagna A, Giardina I. 2014. Bird flocks as condensed matter. *Annu Rev Condens Matter Phys*. 5(1):183–207.
17 Liu ZT, *et al.* 2021. Activity waves and freestanding vortices in populations of subcritical quincke rollers. *Proc Natl Acad Sci U S A*. 118(40):e2104724118.
18 Aminikhanghahi S, Cook DJ. 2017. A survey of methods for time series change point detection. *Knowl Inf Syst*. 51(2):339–367.
19 Trefethen LN, Trefethen AE, Reddy SC, Driscoll TA. 1993. Hydrodynamic stability without eigenvalues. *Science*. 261(5121):578–584.
20 Dakos V, *et al.* 2008. Slowing down as an early warning signal for abrupt climate change. *Proc Natl Acad Sci U S A*. 105(38):14308–14312.
21 Marconi M, *et al.* 2020. Testing critical slowing down as a bifurcation indicator in a low-dissipation dynamical system. *Phys Rev Lett*. 125(13):134102.
22 Liu R, Wang X, Aihara K, Chen L. 2014. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Med Res Rev*. 34(3):455–478.
23 Yang B, *et al.* 2018. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nat Commun*. 9(1):678.
24 Steinhardt PJ, Chaudhari P. 1981. Point and line defects in glasses. *Philosop Mag A*. 44(6):1375–1381.
25 Stukowski A. 2012. Structure identification methods for atomistic simulations of crystalline materials. *Model Simul Mat Sci Eng*. 20(4):045021.
26 Pietrucci F, Martoňák R. 2015. Systematic comparison of crystalline and amorphous phases: charting the landscape of water structures and transformations. *J Chem Phys*. 142(10):104704.
27 Behler J. 2011. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J Chem Phys*. 134(7):074106.
28 Drautz R. 2019. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys Rev B*. 99(1):014104.
29 Faber F, Lindmaa A, Von Lilienfeld OA, Armiento R. 2015. Crystal structure representations for machine learning models of formation energies. *Int J Quantum Chem*. 115(16):1094–1101.
30 Gasparotto P, Bochicchio D, Ceriotti M, Pavan GM. 2019. Identifying and tracking defects in dynamic supramolecular polymers. *J Phys Chem B*. 124(3):589–599.
31 Musil F, *et al.* 2021. Physics-inspired structural representations for molecules and materials. *Chem Rev*. 121(16):9759–9815.
32 Bartók AP, Kondor R, Csányi G. 2013. On representing chemical environments. *Phys Rev B*. 87(18):184115.
33 Monserrat B, Brandenburg JG, Engel EA, Cheng B. 2020. Liquid water contains the building blocks of diverse ice phases. *Nat Commun*. 11(1):5757.
34 Offei-Danso A, Hassanali A, Rodriguez A. 2022. High-dimensional fluctuations in liquid water: combining chemical intuition with unsupervised learning. *J Chem Theory Comput*. 18(5):3136–3150.
35 De S, Bartók AP, Csányi G, Ceriotti M. 2016. Comparing molecules and solids across structural and alchemical space. *Phys Chem Chem Phys*. 18(20):13754–13769.
36 Capelli R, Gardin A, Empereur-Mot C, Doni G, Pavan GM. 2021. A data-driven dimensionality reduction approach to compare and classify lipid force fields. *J Phys Chem B*. 125(28):7785–7796.
37 Lionello C, Perego C, Gardin A, Klajn R, Pavan GM. 2022. Supramolecular semiconductivity through emerging ionic gates in ion–nanoparticle superlattices. *ACS Nano*. 17(1):275–287.

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025

38 Gardin A, Perego C, Doni G, Pavan GM. 2022. Classifying soft self-assembled materials via unsupervised machine learning of defects. *Commun Chem*. 5(1):82.

39 Reinhardt A, Pickard CJ, Cheng B. 2020. Predicting the phase diagram of titanium dioxide with random search and pattern recognition. *Phys Chem Chem Phys*. 22(22):12697–12705.

40 Cioni M, *et al*. 2024. Sampling real-time atomic dynamics in metal nanoparticles by combining experiments, simulations, and machine learning. *Adv Sci*. 11(25):2307261.

41 Cioni M, *et al*. 2023. Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments. *J Chem Phys*. 158(12):124701.

42 Rapetti D, *et al*. 2023. Machine learning of atomic dynamics and statistical surface identities in gold nanoparticles. *Commun Chem*. 6(1):143.

43 Crippa M, Cardellini A, Caruso C, Pavan GM. 2023. Detecting dynamic domains and local fluctuations in complex molecular systems via timelapse neighbors shuffling. *Proc Natl Acad Sci U S A*. 120(30):e2300565120.

44 Caruso C, Cardellini A, Crippa M, Rapetti D, Pavan GM. 2023. Timesoap: tracking high-dimensional fluctuations in complex molecular systems via time variations of soap spectra. *J Chem Phys*. 158(21):214302.

45 Becchi M, Fantolino F, Pavan GM. 2024. Layer-by-layer unsupervised clustering of statistically relevant fluctuations in noisy time-series data of complex dynamical systems. *Proc Natl Acad Sci U S A*. 121(33):e2403771121.

46 Crippa M, Cardellini A, Cioni M, Csányi G, Pavan GM. 2023. Machine learning of microscopic structure-dynamics relationships in complex molecular systems. *Mach Learn Sci Technol*. 4(4):045044.

47 Abascal J, Sanz E, García Fernández R, Vega C. 2005. A potential model for the study of ices and amorphous water: Tip4p/ice. *J Chem Phys*. 122(23):234511.

48 García Fernández R, Abascal JL, Vega C. 2006. The melting point of ice Ih for common water models calculated from direct coexistence of the solid-liquid interface. *J Chem Phys*. 124(14):144506.

49 Spencer M. 1986. Stable and metastable metal surfaces in heterogeneous catalysis. *Nature*. 323(6090):685–687.

50 Jayanthi C, Tosatti E, Pietronero L. 1985. Surface melting of copper. *Phys Rev B*. 31(6):3456.

51 Wang H, Zhang L, Han J, Weinan E. 2018. Deepmd-kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Comput Phys Commun*. 228:178–184.

52 Molgedey L, Schuster HG. 1994. Separation of a mixture of independent signals using time delayed correlations. *Phys Rev Lett*. 72(23):3634.

53 Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. 2013. Identification of slow molecular order parameters for Markov model construction. *J Chem Phys*. 139(1):015102.

54 Mishin Y, Mehl M, Papaconstantopoulos D, Voter A, Kress J. 2001. Structural stability and lattice defects in copper: ab initio, tight-binding, and embedded-atom calculations. *Phys Rev B*. 63(22):224106.

55 Allan DB, Caswell T, Keim NC, van der Wel CM, Verweij RW. 2024. soft-matter/trackpy: v0.6.3. https://doi.org/10.5281/zenodo.11522100.

56 Granger CW. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 37(3):424–438.

57 Chen Y, Rangarajan G, Feng J, Ding M. 2004. Analyzing multiple nonlinear time series with extended granger causality. *Phys Lett A*. 324(1):26–35.

58 Schreiber T. 2000. Measuring information transfer. *Phys Rev Lett*. 85(2):461.

59 Paluš M, Komárek V, Hrnčíř Z, Štěrbová K. 2001. Synchronization as adjustment of information rates: detection from bivariate time series. *Phys Rev E*. 63(4):046211.

60 Glielmo A, Zeni C, Cheng B, Csányi G, Laio A. 2022. Ranking the information content of distance measures. *PNAS Nexus*. 1(2):pgac039.

61 Del Tatto V, Fortunato G, Bueti D, Laio A. 2024. Robust inference of causality in high-dimensional dynamical processes from the information imbalance of distance ranks. *Proc Natl Acad Sci U S A*. 121(19):e2317256121.

62 Hess B, Kutzner C, Van Der Spoel D, Lindahl E. 2008. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 4(3):435–447.

63 Frenkel D. 2013. Simulations: the dark side. *Eur Phys J Plus*. 128:1–21.

64 Thompson AP, *et al*. 2022. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput Phys Commun*. 271:108171.

65 Rassoulinejad-Mousavi SM, Mao Y, Zhang Y. 2016. Evaluation of copper, aluminum, and nickel interatomic potentials on predicting the elastic properties. *J Appl Phys*. 119(24):244304.

66 Evans DJ, Holian BL. 1985. The nose–hoover thermostat. *J Chem Phys*. 83(8):4069–4074.

Downloaded from https://academic.oup.com/pnasnexus/article/4/2/pgaf038/8003892 by Politecnico di Torino user on 20 March 2025