Doctoral Dissertation

Elucidating Material Dynamics with Deep Learning-based Inference

DAO Duc Anh

Supervisor DAM Hieu Chi

Division of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Knowledge Science)

March 2026

Abstract

Scientific inquiry seeks to understand how objects behave under varying conditions, yet objects are never accessed directly. They are encountered only through observations obtained under specific experimental settings, each capturing a partial and condition-dependent manifestation. As a result, object behavior cannot be identified from individual observations alone. Instead, behaviors and behavior patterns are articulated as characteristic responses of an object, inferred from structured variability across collections of observations as observing conditions change.

This dissertation formulates object inquiry as a process of organizing and interpreting observational variability. Observations, representations, relations, and behaviors are assigned distinct roles. Observations constitute empirical records and exhibit substantial variability due to experimental conditions. Learned representations are introduced as organizational structures that arrange observations so that specific kinds of relations become examinable, including similarity, continuity, progression, and contextual dependence. These relations organize how observations vary, but do not themselves define behavior. Behaviors are characterized at the observation level as coherent and condition-dependent patterns of variability revealed through how such relations evolve under applied conditions. In this sense, inference refers to the construction of behavior-level understanding from structured observational variation, rather than to prediction or parameter estimation from isolated data instances.

Material dynamics are examined as a primary instantiation of this inquiry formulation. In material systems, behaviors such as diffusion, deformation, and transformation are distributed across time, scale, and measurement modality, and are manifested through complex and coupled sources of variability. Direct comparison of observations is therefore unreliable, and inquiry requires organizing large collections of heterogeneous observations to expose condition-dependent patterns of change. This requirement motivates the integration of deep-learning models as inferential instruments, selected according to the representational properties they provide.

Two complementary deep-learning–integrated inquiry approaches are developed. A generative-inference approach employs deep generative models to organize admissible variability and continuity among observations, enabling systematic exploration of plausible transformation pathways and statistical characterization of behavior patterns beyond direct observation. An

attentive-inference approach employs attention-based transformer models to organize contextual relations within observational data, emphasizing how localized features contribute to global responses across space, time, and modality. These approaches address the same notion of object behavior, while exposing complementary facets of how behaviors are manifested and constrained.

Across multiple case studies in material dynamics, the dissertation demonstrates how deep learning can support object inquiry not as a predictive endpoint, but as a means of organizing variability and relations in ways that make behaviors and behavior patterns interpretable. The central contribution is a clarified account of deep learning-based inference in object inquiry, specifying how representations support the organization of relations from which object behavior can be systematically articulated.

Keywords:: Scientific inquiry, Material dynamics, Time-resolved microscopy, Deep generative models, Attention mechanisms

Acknowledgment

First and foremost, I would like to express my sincere appreciation to my research supervisor, Professor Dam Hieu Chi, for granting me the opportunity to pursue this work and for providing steady guidance throughout the research. It has been a privilege to study under his supervision, and I am deeply grateful for his support.

I also wish to extend my gratitude to my collaborators, including members of Takahashi Laboratory at Tohoku University, the Tada Group at Nagoya University, and Professor Oshima Yoshifumi from the School of Materials Science at JAIST. Their helpful discussions and cooperation greatly contributed to the progress of this research. I am further thankful to Professor Huynh Van Nam for his instruction on the foundations of Knowledge Science.

In addition, I would like to thank my friends and research colleagues—Dr. Nguyen D. N., Dr. Vu T. S., Dr. Ha M. Q.—for their continual encouragement and support during this work.

Finally, my sincere thanks go to all those who have supported this research, whether directly or indirectly.

Contents

List of Figures

List of Tables

Chapter 1

Conceptual Object Inquiry into Material Dynamics

This chapter introduces the conceptual foundations of the dissertation by clarifying how dynamic behavior is approached in scientific settings where objects are accessed indirectly through observations. In many domains, and particularly in the study of evolving material systems, individual measurements capture only partial and condition-dependent manifestations of an object. Dynamic behavior cannot be identified from any single observation, but must instead be articulated through relationships that connect observations across time, scale, and experimental context. Establishing how such articulation is possible, and how it can be carried out coherently, is the central aim of this chapter.

The section *Research Background* situates this aim within a broader view of scientific inquiry. It develops a perspective in which observations serve as the primary empirical interface to objects, while behaviors are understood as relational descriptions that emerge through systematic examination of multiple observations. Within this context, the subsection *Observations and Behaviors in Scientific Inquiry* emphasizes why variability across observations is an intrinsic feature of empirical data rather than a secondary complication, and why understanding behavior requires attention to how observations relate rather than to how any single observation appears in isolation.

Building on this foundation, the subsection *Object-Inquiry Framework as a Conceptual Process* introduces a structured perspective for reasoning about objects studied through variable observations. This framework clarifies how observations can be organized and how relational structure can be interpreted so that object behavior is coherently articulated. Rather than advancing specific analytical or computational techniques, the framework establishes a conceptual vocabulary and process that guide how observations, representations, relationships, and behaviors are distinguished and connected throughout the dissertation.

These ideas are then grounded in a concrete scientific context through the

subsection *Material Dynamics as an Object of Inquiry*. Material dynamics provide a representative domain in which the challenges motivating object-centered inquiry are particularly evident, as material systems are accessed through time-resolved observations that are inherently variable, multiscale, and often multimodal. This subsection explains why such systems require careful organization of observations and systematic examination of relationships, and it introduces learning-based representations as practical means of instantiating the object-inquiry perspective when observational complexity exceeds what can be handled through direct inspection.

The chapter concludes by articulating the *Research Purposes* of the dissertation and outlining the *Organization of the Dissertation*. These final sections connect the conceptual perspective developed in the research background to the representational and methodological developments pursued in subsequent chapters, positioning Chapter 1 as the conceptual entry point for the work as a whole.

1.1 Research Background

This section provides the conceptual background motivating the object-inquiry framework. It examines how scientific inquiry proceeds when objects cannot be accessed directly but are studied through collections of observations obtained under varying conditions. The section focuses on the relationship between observations and behaviors, emphasizing that behavior is not a directly observable entity but a relational construct that emerges through systematic examination of multiple observations. The discussion clarifies why variability among observations plays a central role in shaping scientific understanding. Differences across observations—arising from changes in time, experimental conditions, spatial location, or measurement modality—complicate direct comparison and obscure behavioral structure. Rather than treating variability as noise to be eliminated, the section frames it as an intrinsic feature that must be addressed explicitly within inquiry. By articulating how behaviors are understood through relationships among observations, this section establishes the conceptual motivation for a structured framework that can organize observations and support coherent articulation of behavior. The considerations developed here prepare the ground for introducing the object-inquiry framework in the following section.

1.1.1 Observations and Behaviors in Scientific Inquiry

This subsection examines the foundational roles of observations and behaviors in scientific inquiry. Its purpose is to clarify how objects are accessed empirically through observations and how behaviors are articulated

through relationships among those observations. Rather than treating observations as direct representations of objects or behaviors as intrinsic properties of single states, the subsection emphasizes the relational nature of scientific understanding. The discussion highlights why individual observations are insufficient for characterizing object behavior and why collections of observations must be examined collectively. By focusing on how similarities, differences, and progressions among observations give rise to behavioral descriptions, the subsection establishes a relational perspective that underlies object-centered inquiry. This treatment also foregrounds the role of variability in shaping how behaviors are identified and described. As observations become more diverse and distributed across conditions, behaviors increasingly emerge through structured relational examination rather than local inspection. These considerations motivate the need for explicit organizational and interpretive processes, which are developed further in the subsequent subsection.

1.1.1.1 Scientific Inquiry through Observations and Behaviors

Scientific inquiry is motivated by our interest in objects and by the practical need to use, manipulate, or rely on them in specific situations. When working with an object, we often seek to know in advance how it will perform, whether it will remain stable, or how it will respond to external influences. Such questions arise naturally from anticipated use, design, or intervention, rather than from a desire to describe the object exhaustively in all its possible aspects. Scientific inquiry provides a systematic way to address these concerns by examining how objects behave under conditions that are relevant to their intended application.

Empirical observations play an essential role in motivating and guiding this inquiry. Initial observations—whether encountered in everyday practice, exploratory experiments, or prior studies—may reveal repetitions, sensitivities, or contrasts in how an object responds. For example, repeated deformation of a metal specimen may suggest a tendency toward work hardening, or changes in optical response under illumination may hint at underlying electronic structure. Such observations do not by themselves establish general conclusions, but they prompt further investigation by indicating that an object exhibits responses that depend on external conditions. In this way, observations motivate scientific inquiry by revealing aspects of object behavior that warrant systematic examination.

To support this aim, scientific inquiry considers situations in which an object may be placed or operated. These situations may involve specific mechanical, thermal, chemical, or electromagnetic conditions, and they are not always encountered directly or repeatedly in practice. Instead, scientific

inquiry recreates or stages such situations in a controlled manner so that object responses can be examined deliberately. The outcomes of these staged situations constitute what are referred to as observations within scientific inquiry.

Within this framework, observations take the form of measurements, images, or signals obtained under specified experimental conditions [1]. Each observation records how the object behaves when subjected to a particular configuration of conditions, such as a given temperature, stress state, chemical environment, or illumination. In this sense, an observation represents empirical evidence of the object in a specific situation constructed to reflect prospective use. It does not aim to describe the object in general, but to document how it responds under that particular set of conditions [2].

Because each observation corresponds to a specific situation, a single observation provides information only about the object under that condition. It does not, by itself, indicate how the object would respond if conditions were changed, nor does it establish whether the observed response is typical across situations. For instance, a single diffraction pattern records the structure of a crystal at a particular orientation and wavelength, but does not by itself determine crystallographic symmetry [3]. Similarly, an individual photoemission spectrum reflects electronic states under specific measurement settings, but does not alone reveal the full band structure [4]. Relying on isolated observations therefore limits what can be concluded about object performance.

Scientific inquiry organizes observations with a broader objective in mind. Rather than treating observations as independent records, it considers observations obtained under controlled changes in experimental conditions as complementary evidence of object response. By examining how observations differ as conditions are varied, inquiry moves beyond the particulars of any single situation. This organization allows observations to inform expectations about how the object is likely to behave in situations beyond those directly examined, which is essential for reliable use and manipulation.

Within this perspective, the consistent ways in which an object responds across different conditions are referred to as its behaviors [5, 6]. Behaviors are not identified with any single observation, but with what observations collectively indicate about object response. They summarize how observable manifestations change as parameters such as temperature, stress, or environment are modified. In this sense, behaviors capture the aspects of the object that scientific inquiry seeks to establish in order to guide future use, design, and intervention.

Scientific inquiry can therefore be understood as a process that proceeds from observations toward knowledge of object behavior. Objects are charac-

terized not solely by individual observed instances, but by the responses they exhibit across a range of conditions relevant to anticipated situations. By organizing observations around this response-oriented perspective, scientific inquiry provides a practical and systematic basis for reasoning about object performance. This perspective sets the stage for examining how information about behavior is derived from observations, which is addressed in the following subsection.

1.1.1.2 Deriving Behaviors from Relations of Observations

Observations obtained under different experimental conditions provide multiple perspectives on how an object responds in specific situations. Each observation reflects the object under a particular configuration of parameters and records only how the object behaves in that situation. To understand how the object behaves beyond any single instance, scientific inquiry must therefore attend to how observations change as conditions are varied. What matters is not the absolute content of an individual observation, but how observations differ relative to one another under controlled variation.

When observations produced from the same object are examined together, their variation is not arbitrary. As experimental conditions such as temperature, stress, or chemical environment are modified, the object responds in ways that reflect its material constitution and functional characteristics. These responses give rise to structured patterns among observations, including systematic dependence on control parameters, recurring correspondences across experimental settings, and consistent organization across measurement modalities. For example, diffraction patterns collected under different loading conditions may show coordinated shifts in peak positions, while spectroscopic measurements may exhibit systematic changes in intensity or energy that track environmental variation [7, 8].

From this perspective, behaviors are expressed through the regular ways in which observations relate to one another across conditions. An object's behavior determines which kinds of variation are observed and how changes under one condition are connected to changes under another. If observations varied independently with no discernible structure, it would not be possible to form a coherent account of how the object responds. The emergence of reproducible patterns across observations therefore indicates that the object exhibits characteristic modes of response as conditions are varied.

The role of any individual observation is thus defined by how it participates in this broader pattern of variation. An observation that appears atypical when considered in isolation may be fully consistent with object behavior once it is situated among other observations obtained under related conditions. Conversely, observations that appear similar at first glance may

correspond to different behaviors if they arise from distinct response patterns across the explored range of conditions [7, 8]. Understanding behavior therefore requires situating observations within the larger context of how responses are organized across variation, rather than relying on surface-level similarity.

As observations are collected over broader ranges of experimental parameters, these response patterns become more clearly articulated. Relations that appear incomplete or ambiguous when only a narrow set of conditions is examined may resolve into well-defined trends or distinct response regimes when additional observations are incorporated [9,10]. In such cases, behavior is not associated with any particular observation, but with the structured manner in which observations change together as conditions are varied.

In many experimental settings, information about behavior is therefore distributed across collections of observations rather than localized within individual measurements. For example, responses inferred from combined imaging, diffraction, and spectroscopic data often become apparent only when observations from different modalities are considered together [11, 12]. Each observation contributes partial evidence, and behavior is derived through the organization of these contributions into coherent patterns of response.

Deriving behavior from observations can thus be understood as recognizing and characterizing the non-arbitrary patterns that persist as conditions are varied. Scientific inquiry does not impose these patterns, but seeks to make them explicit by organizing observations so that consistent response tendencies can be identified. This need to systematically organize observations in order to reveal how behaviors are expressed across variation motivates the development of formal approaches for structuring observational data, which is taken up in the following section.

1.1.2 Object-Inquiry Framework as a Conceptual Process

This subsection introduces the object-inquiry framework as a conceptual process for reasoning about objects studied through variable observations. Building on the preceding discussion of observations and behaviors, it articulates how inquiry proceeds when behavior must be constructed from relations among observations rather than directly observed. The framework is presented as a structured way of organizing observations and interpreting relational structure so that behavior can be articulated in a coherent and empirically grounded manner. The subsection clarifies the complementary roles of organizational constructs, such as representations, and interpretive

steps that connect abstract relational structure back to observable manifestations. By framing object inquiry as a process that integrates organization and interpretation, this subsection establishes the conceptual architecture that guides the remainder of the dissertation.

1.1.2.1 Conceptual Basis for Object Inquiry

Scientific inquiry frequently investigates objects through collections of observations obtained under varying experimental conditions. Each observation records how an object manifests in a specific situation, while information about object behavior is carried by how such observations are related across conditions. Meaningful descriptions of behavior therefore depend not on individual observations, but on identifying and interpreting relations among them. In practice, however, establishing and understanding these relations is nontrivial, giving rise to fundamental challenges in object-centered inquiry.

The first challenge concerns how observations are organized so that relations among them can be identified. Observations associated with the same object may differ substantially due to changes in experimental conditions, measurement settings, or modes of observation [7, 13]. As a result, direct comparison between observations is often unreliable. Apparent differences between observations do not necessarily correspond to different object behaviors, while apparent similarities may obscure meaningful distinctions. Without systematic organization, observations remain a heterogeneous collection of instances, and relations that reflect how the object responds across conditions remain difficult to discern. This difficulty is especially pronounced when observations are high-dimensional, multimodal, or capture complex manifestations of the object.

This challenge becomes evident when considering repeated examinations of the same object under different conditions. Each observation captures a particular manifestation and may emphasize different attributes or aspects of the object. When examined individually, such observations can appear inconsistent or difficult to reconcile. Even pairwise comparison may be dominated by superficial variation, preventing relations that span across multiple conditions from becoming apparent. Without an organizational structure that situates observations within a broader context, relations that carry information about object behavior are easily obscured.

The second challenge concerns how identified relations are interpreted as meaningful descriptions of object behavior. Relations among observations are often abstract in form, reflecting similarity, correspondence, or structured arrangement rather than directly observable characteristics. While such relations may be formally well defined, they do not automatically indicate what they imply about how the object behaves [14, 15]. There is therefore

a gap between recognizing relational structure and understanding how that structure reflects consistent ways in which the object responds across conditions.

This interpretative challenge is closely tied to the manner in which relations are obtained. Relations are typically identified within some organized view of the observations rather than directly from raw observational instances. However, relations expressed in such organized forms may lack immediate physical or intuitive meaning [16, 17]. Interpreting them requires reconnecting relational structure to observable manifestations of the object, ensuring that descriptions of behavior remain grounded in empirical evidence. Without this connection, relations risk remaining formal constructs rather than informative accounts of object behavior.

Together, these considerations define the foundational requirements for object-centered inquiry. Scientific inquiry must both organize observations so that relations among them can be made explicit and interpret those relations in terms of how the object responds across conditions. Addressing only one of these requirements is insufficient: organization without interpretation yields abstract structure without behavioral meaning, while interpretation without systematic organization lacks a reliable basis for identifying relations. These requirements motivate the need for a structured approach that treats organization and interpretation as complementary aspects of inquiry, which is developed in the following subsection.

1.1.2.2 Construction of Object-Inquiry Framework

The object-inquiry framework provides a structured way to organize observations and examine the relations among them so that object behavior can be articulated in a coherent and empirically grounded manner. Rather than treating observations as isolated records, the framework treats them as interconnected manifestations of the same object obtained under varying conditions. Its construction is guided by the need to make relations among observations explicit and to ensure that these relations can be interpreted as meaningful descriptions of how the object responds across conditions.

Access to an object within this framework is mediated through a collection of observations obtained under different experimental conditions. Let $\mathcal{X}$ denote the space of observations, and let $\{x_i\}_{i=1}^{N} \subset \mathcal{X}$ represent the set of observations associated with a given object. Each observation captures a particular manifestation of the object, reflecting how it appears under a specific configuration of conditions. Object inquiry therefore proceeds by examining how observations relate to one another across the collection, rather than by assigning interpretive significance to individual observations in isolation.

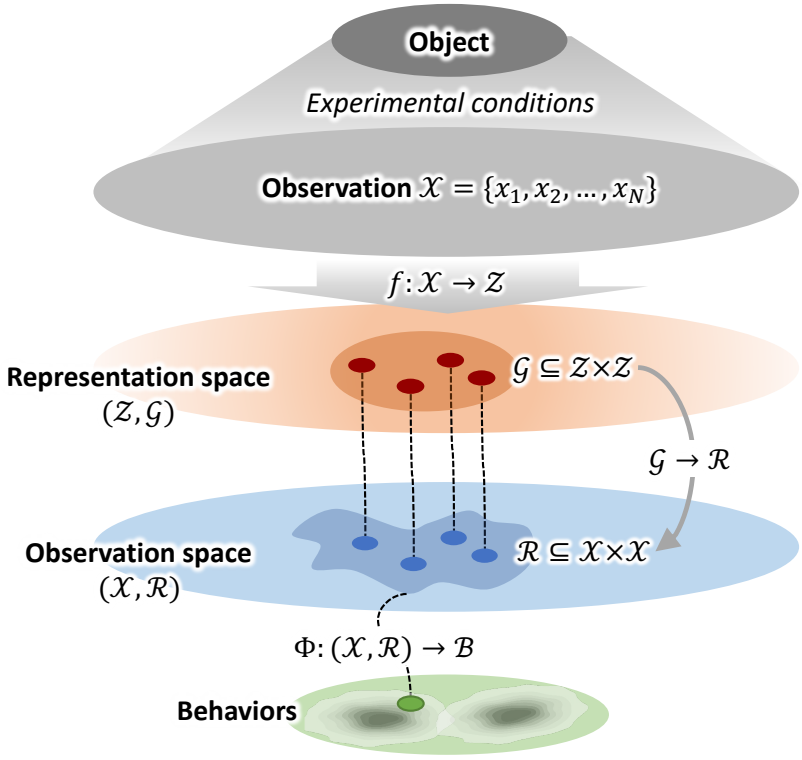

Figure 1.1: Schematic diagram of the conceptual object-inquiry framework from observations of an object in various experimental conditions.

Because observations may differ substantially in form, scale, or appearance, identifying meaningful relations directly in the observation space $\mathcal{X}$ is often unreliable. Superficial variation can dominate comparison, obscuring patterns that reflect how the object responds across conditions. To address this, the object-inquiry framework introduces an intermediate organization of observations that supports systematic relational examination. Let $\mathcal{Z}$ denote a representation space, and let $f : \mathcal{X} \to \mathcal{Z}$ map each observation to a representation. This organization places observations into a structured form in which relations among them can be more consistently examined across the collection.

Within this organized representation, relations among observations become more clearly expressed. Observations corresponding to similar manifestations of the object may be brought into closer correspondence within $\mathcal{Z}$, while observations associated with distinct manifestations may be separated. Relations expressed at this level can be written as $\mathcal{G} \subset \mathcal{Z} \times \mathcal{Z}$, capturing how representations are connected across the collection. At this stage, these relations describe how observations are arranged relative to one another,

providing a stable basis for examining relational structure without yet assigning behavioral interpretation.

Relational structure identified in representation space, however, does not by itself constitute an account of object behavior. Relations expressed in $\mathcal{Z}$ are abstract and must be interpreted in relation to observable manifestations of the object. The object-inquiry framework therefore emphasizes an interpretive step that reconnects relational structure to the observation space. Conceptually, this involves examining how relations among representations correspond to relations among observations, that is, how elements of $\mathcal{G}$ reflect relations $\mathcal{R} \subset \mathcal{X} \times \mathcal{X}$ grounded in observable characteristics and experimental conditions.

Object behavior is articulated through this interpretive connection. Rather than equating relations themselves with behavior, the framework treats behavior as a description of how the object manifests across the collection of observations. Consistent relational organization is interpreted in terms of observable variation, recurrence, or systematic change across conditions. This may involve identifying which aspects of observation change together, which remain stable, and how such patterns are distributed across the range of conditions examined. In this way, behavior is understood as a collective property revealed through relations among observations, while remaining anchored in empirical evidence.

By maintaining a balance between organization and interpretation, the object-inquiry framework avoids two complementary pitfalls. Organization without interpretation risks producing abstract relational structure with little connection to observable object behavior, while interpretation without systematic organization lacks a reliable basis for identifying relations across variable observations. The framework instead treats organization and interpretation as mutually dependent components of inquiry, jointly enabling the transition from collections of observations to interpretable descriptions of object behavior. With this structure in place, the framework can be adapted to specific domains by specifying the nature of observations, the form of their organization, and the manner in which relations are interpreted in domain-relevant terms.

1.1.3 Material Dynamics as an Object of Inquiry

Material dynamics concern how material systems evolve through changes in structure, composition, or configuration over time and across varying conditions [13,18]. Unlike static material characterization, which emphasizes equilibrium states or fixed descriptors, the study of material dynamics treats evolution itself as the defining feature of the object under investigation. Behaviors of interest are expressed through transformation pathways, in-

termediate configurations, and transient states that cannot be fully captured by isolated measurements.

This perspective has become increasingly important in contemporary materials science. Many properties relevant to technological performance, including mechanical response, stability, and functionality [19, 20], depend not only on material structure but on how that structure develops under external stimuli or operational environments [21, 22]. At microscopic and nanoscopic scales, transformations such as phase transitions, structural deformation, and chemical reaction pathways play a central role in determining material behavior [23–28]. Subtle differences in transformation pathways can lead to markedly different outcomes, even when composition is identical, as illustrated in systems exhibiting stereochemical sensitivity such as molecular chirality [29].

Advances in experimental instrumentation have reinforced this shift in emphasis. Modern *in situ* and time-resolved techniques enable materials to be observed repeatedly as they evolve, providing empirical access to dynamic processes that were previously inferred only indirectly [7, 30]. Rather than relying on before-and-after comparisons, investigators can now examine how materials transform through successive stages, motivating an object-centered view in which dynamic behavior is treated as a primary subject of inquiry [1, 30–33].

Within the object-inquiry framework, material dynamics are naturally treated as object-centered phenomena. The material system constitutes the object, while its behavior is revealed only through collections of observations acquired across time or conditions. No single observation suffices to characterize such behavior; instead, behavior emerges from how observations relate across sequences. This makes material dynamics a particularly demanding application domain, as both the organization of observations and the articulation of behavior are central concerns.

1.1.3.1 Time-Resolved Observations of Material Dynamics

Access to material dynamics is mediated through time-resolved and condition-dependent observations. Advances in experimental techniques have enabled materials to be observed at spatial and temporal resolutions that approach the intrinsic scales of structural transformation [34,35]. As a result, dynamic processes can now be recorded with unprecedented detail, producing dense sequences of observations that capture intermediate states and transient configurations as materials evolve. Rather than limiting investigation, this abundance of observation shifts the central challenge toward making sense of increasingly complex observational records.

Time-resolved observations emphasize progression rather than static

state [30, 32]. Each observation represents a partial manifestation of an ongoing process, capturing only a snapshot of material evolution [13]. Dynamic behavior is therefore distributed across sequences of observations rather than contained within any individual measurement. Understanding material dynamics requires examining how observations relate across time or conditions, rather than interpreting them independently. The object of inquiry is accessed through collections of temporally ordered observations, whose relations encode aspects of dynamic behavior.

Despite increased resolution and sampling frequency, time-resolved observations remain inherently discrete [7, 33]. Continuous material evolution is sampled at finite intervals determined by experimental design and instrumental limits. As observations become denser, successive measurements may appear more closely related, yet they still represent distinct instances separated by unobserved intervals. Apparent continuity must therefore be understood through relations among observations rather than through direct observation of continuous change.

In addition to temporal sequencing, time-resolved material data often span multiple spatial and temporal scales [1, 26]. Observations may capture global material configurations at coarse resolution while simultaneously resolving localized features at finer scales. Temporal sampling may likewise vary, with fast processes embedded within slower evolutionary trends. As a result, relations of interest are not limited to those between whole observations, but also arise across spatial and temporal decompositions of the data [10, 15]. Understanding material dynamics therefore requires relating observations not only across time, but also across scales at which different aspects of material behavior become visible.

Time-resolved studies increasingly incorporate multiple synchronized measurement modalities. Imaging, diffraction, and spectroscopic signals may be acquired concurrently, each providing a distinct perspective on the evolving material system [20, 36, 37]. These modalities capture complementary aspects of material behavior, such as structural configuration, chemical composition, or electronic state. Consequently, relations of interest extend beyond comparisons between observations of the same type, encompassing correspondences across measurement channels and their respective representations.

The richness of such data introduces substantial variability and ambiguity across observations. Material systems often exhibit heterogeneous or localized changes, with different regions or features evolving in distinct ways. Observables such as image contrast, diffraction phase, or spectral intensity may vary across observations due to multiple interacting factors, including changes in geometry, strain, or composition [20, 36, 37]. As

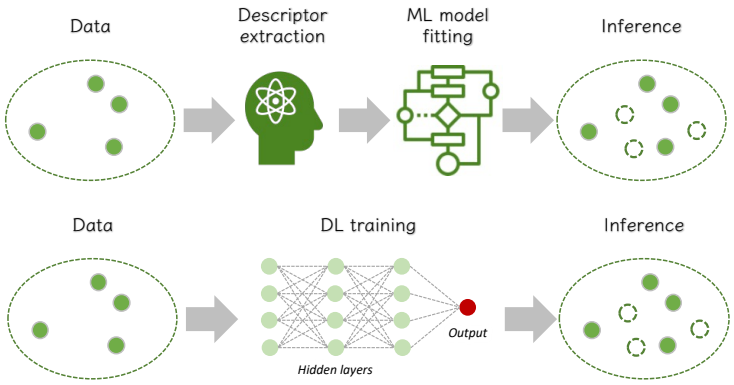

Figure 1.2: Comparison between conventional machine-learning and deep-learning frameworks in deriving inference from data. Conventional ML relies on handcrafted descriptors while the DL internally learn representations of data to perform inference.

a result, observations originating from the same underlying process may appear markedly different, while similar appearances may arise from distinct material configurations.

As datasets become richer, spanning multiple scales and modalities, the space of possible relations expands rapidly. Increased observational detail enhances descriptive fidelity, but it also amplifies the diversity of ways in which observations, decompositions, and measurement channels may be related across time and conditions [34, 35]. Making sense of time-resolved material data therefore requires approaches that can accommodate high variability while supporting coherent examination of relations across collections of observations, scales, and modalities.

These characteristics make time-resolved observations of material dynamics a paradigmatic challenge for object-centered inquiry. Observations must be examined collectively, relations must be established across sequences, scales, and measurement channels, and behavior must be articulated from how manifestations evolve over time. The following subsection discusses representational approaches that are suited to organizing such complex observational data and supporting structured examination of relations in the study of material dynamics.

1.1.3.2 Learning Representations for Material Dynamics

The high variability and complexity of time-resolved material observations motivate the use of learning-based representations as tools for organizing data and examining relations. As discussed in the previous subsection,

material dynamics are accessed through collections of observations that span multiple spatial and temporal scales and may involve multiple synchronized measurement modalities. Direct comparison of such observations is often unreliable due to nonlinear coupling among structural, chemical, and geometrical factors [7, 26]. Learning-based representations provide a means of organizing these observations into structured forms that support systematic examination of relations across time, scale, and modality [15, 38].

Conventional machine learning approaches have long been employed to construct representations of material data through feature extraction and statistical modeling [39–41]. In these approaches, observations are mapped into feature spaces designed to capture aspects of material structure or evolution deemed relevant by prior knowledge. Relations among observations are then examined through similarities, distances, or correlations defined within these spaces. Such representations can be effective when the dominant sources of variation are well understood and can be expressed through handcrafted descriptors. However, their applicability becomes limited as material behavior arises from interacting mechanisms that are nonlinear, multiscale, and difficult to parameterize explicitly.

Time-resolved material data frequently exhibit such complexity. Observed changes may reflect the coupled influence of diffusion, deformation, phase transformation, and surface processes occurring concurrently and across different scales [36, 37, 42]. In these settings, constructing representations that meaningfully organize observational variability becomes particularly challenging. Descriptors designed without sufficient understanding of the underlying mechanisms may fail to capture important aspects of material evolution, while descriptors derived from incomplete observational perspectives may obscure relations that are crucial for articulating behavior.

This difficulty reflects a broader methodological tension in scientific inquiry. Meaningful representations are required in order to organize observations and reveal relations that characterize object behavior. Yet constructing such representations typically presupposes some prior understanding of the phenomena under investigation. In the study of complex material dynamics, this creates a practical paradox: understanding is needed to design representations that reveal behavior, while those representations are themselves required to obtain that understanding. Scientific inquiry therefore advances iteratively, through successive refinement of representations and interpretations as observational evidence accumulates. Provisional representations organize observations sufficiently to expose partial relational structures, which in turn inform improved understanding of the system and guide subsequent refinement of the representations used to examine it.

Learning-based representations provide practical mechanisms for support-

ing this iterative process in data-rich scientific settings. Deep learning, in particular, offers a flexible approach to constructing representations suited to complex observational data [43, 44]. Rather than relying on predefined features, deep models learn hierarchical representations directly from data through compositions of nonlinear transformations. These representations can capture intricate dependencies among observables and integrate information across spatial and temporal scales. For time-resolved material observations, such flexibility enables representations to adapt to heterogeneous patterns of evolution and to organize observations whose relations are not well described by simple similarity measures.

A key characteristic of deep learning representations is their ability to reorganize observational data into latent spaces in which complex relational structure becomes more coherent [45]. Observations that appear dissimilar in raw form may be organized closely in representation space if they correspond to related material states, while observations reflecting distinct modes of behavior may be separated. This reorganization supports examination of relations that reflect underlying material evolution, even when such relations are obscured by variability in the original observation space.

Different classes of deep learning models emphasize complementary aspects of representational organization, reflecting distinct requirements on how relations among observations are to be expressed and examined. Importantly, these representations are not fixed descriptors but are *learned*, adapting to prior understanding and to the specific relational criteria imposed by the inquiry. At a basic level, encoder-based representations provide a common space in which observations can be organized and compared. Contrastive and metric-learning approaches further refine this organization by stabilizing similarity relations, encouraging observations deemed related under given criteria to be placed in close proximity while separating others [38, 46].

Beyond similarity-based organization, other model classes encode richer forms of relational structure. Generative representations focus on capturing admissible variation within observational data, learning distributions that organize observations according to shared patterns of transformation and variability [37, 43, 47, 48]. Such representations are well suited to structuring large collections of time-resolved observations, as they support coherent comparison across sequences and conditions. Attention-based representations, in contrast, emphasize contextual interaction by selectively weighting and coupling features across space, time, or modality [16, 49]. By modulating which aspects of observations are related, attention mechanisms enable focused examination of relational dependencies within complex observational settings. Temporal models, including recurrent and sequence-based architec-

tures, further extend this organization by explicitly encoding progression and correlation across ordered observations [50, 51].

These representational strategies are particularly relevant for material dynamics, where meaningful behavior often emerges only through the joint consideration of multiple scales, temporal progressions, and measurement channels [20,37,52,53]. Within the object-inquiry framework, generative and attention-based representations serve as prominent examples of how learned representations can be tailored to specific relational goals, while remaining part of a broader hierarchy of models that collectively support the systematic organization and examination of relations among observations.

Within this framework, learning-based representations are not treated as explanatory models of material behavior but as organizational constructs that enable relational examination. Their role is to structure observations so that relations across time, scale, and modality can be examined coherently, allowing representations and interpretations to evolve together as inquiry progresses. The subsequent chapters build on this perspective by examining generative and attentive representational strategies as concrete instantiations of object-centered inquiry in the study of material dynamics.

1.2 Research Purposes

The preceding sections established a conceptual object-inquiry framework in which objects are accessed through observations, and behavior is articulated through relations among observations once they are appropriately organized and interpreted. This framework provides a structured perspective for reasoning about object behavior in settings where observations are heterogeneous, condition-dependent, and distributed across time, scale, or measurement modality. The present dissertation builds on this conceptual foundation by examining how object inquiry can be carried out in practical scientific settings where such observational conditions are prevalent.

Contemporary materials research provides a particularly relevant context for this examination. Advances in experimental techniques have enabled material systems to be observed through large collections of time-resolved and multimodal observations that capture evolving structures and heterogeneous processes. In such settings, material behavior is not evident within individual observations but must be articulated through relations across collections of data. Material dynamics is therefore adopted in this dissertation as a representative and demanding domain in which the instantiation of object inquiry can be systematically examined.

The general purpose of this dissertation is to examine how an object-inquiry framework can be instantiated in practice to articulate material be-

havior from complex observational data. This purpose emphasizes practical inquiry: how observations are organized, how relations among them are exposed, and how these relations are interpreted as descriptions of object behavior. Rather than seeking to explain underlying mechanisms or to make predictive claims, the dissertation focuses on clarifying how behavior can be coherently articulated from observational evidence within the object-inquiry perspective.

To achieve this general purpose, the dissertation pursues the following specific objectives:

- **Instantiation of object inquiry through representation:**

  The first objective is to investigate how learning-based representations can be used to instantiate the object-inquiry framework by organizing material observations in ways that make behavior-informing relations explicit. This objective examines representations as organizational structures that shape how observations are related across varying conditions, rather than as explanatory models or predictive tools. Attention is given to how different representational choices influence the stability and accessibility of relations among observations in high-dimensional and heterogeneous datasets.

- **Articulation of material behavior through relational interpretation:**

  The second objective is to articulate dynamic material behaviors by interpreting relations among observations as structured within representation spaces. This objective focuses on how behaviors can be described as consistent patterns of response across varying experimental conditions, as revealed through relational organization rather than isolated observations. By examining how relational structure supports different descriptions of behavior, the dissertation clarifies how object inquiry operates in practice within the context of material dynamics.

Together, these objectives define the scope of the dissertation. The work concentrates on examining how a conceptual object-inquiry framework can be practically realized through learning-based representations and relational interpretation, using material dynamics as a representative domain of application. Through this focus, the dissertation aims to provide methodological clarity for conducting object-centered inquiry in scientific settings characterized by rich, variable, and distributed observational data.

1.3 Organization of the Dissertation

This dissertation is organized to develop and demonstrate an object-inquiry approach for studying material systems through complex, time-resolved observations. The overall structure reflects a progression from conceptual framing to representational realization, moving from clarifying how dynamic behavior can be articulated from observations to examining how learning-based representations support this articulation in practice. Rather than advancing a single algorithmic pipeline, the dissertation builds a coherent perspective on how objects, observations, representations, relations, and behaviors are connected in the study of evolving material systems.

Chapter 1: Conceptual Object Inquiry of Material Systems. Chapter 1 establishes the conceptual foundation of the dissertation. It introduces object inquiry as a general framework for reasoning about dynamic systems accessed through variable and indirect observations. The chapter discusses the challenges posed by observational variability, multiscale structure, and temporal evolution, and motivates material systems as a representative scientific domain in which these challenges are especially prominent. Time-resolved observations and learning-based representations are positioned as essential components of object inquiry, providing the conceptual basis for the methodological developments that follow.

Chapter 2: Representational Foundations for Object Inquiry. Chapter 2 develops the representational perspective underlying the dissertation. It examines how observations are organized into representations, how relations among observations are encoded, and how dynamic behavior can be articulated from these relations. The chapter reviews canonical scientific approaches alongside modern learning-based models, emphasizing how representational capacity and relational expressiveness have evolved to accommodate increasingly complex observations. This chapter serves as a conceptual and methodological bridge between the inquiry framework introduced in Chapter 1 and its concrete realizations in later chapters.

Chapter 3: Generative Inquiry Framework of Material Dynamics. Chapter 3 investigates deep generative modeling as a means of organizing time-resolved material observations into structured representational spaces. The focus is on how generative representations extend the scope of observation by populating spaces of plausible variation consistent with experimental data, thereby enabling examination of dynamic behavior distributed across sequences of configurations. Rather than treating generative models as predictive or explanatory devices, the chapter analyzes their role as representational mechanisms that organize variability and support relational examination of material dynamics.

Chapter 4: Attention-Based Inquiry Framework of Material Dynamics. Chapter 4 introduces attention-based modeling as a complementary realization of object inquiry for settings in which observations are dense but structurally complex. Attention mechanisms are examined as representational tools that expose relations between localized structural variation and measured material response. Through case studies spanning nanoscale deformation and macroscale interfacial contact, the chapter demonstrates how attention-based representations organize relevance across space and time and enable articulation of dynamic behavior from structured relations among observations.

Chapter 5: Discussion and Conclusions. Chapter 5 synthesizes the conceptual and methodological contributions of the dissertation. It reflects on object inquiry as a unifying perspective for studying dynamic material systems and evaluates generative and attention-based representations as complementary mechanisms for organizing observations and articulating behavior. The chapter discusses the scope and limitations of the proposed approaches, considers their generality beyond the specific material systems studied, and outlines directions for future work in data-driven investigation of complex dynamical phenomena.

Through this organization, the dissertation advances a unified narrative: from defining how dynamic behavior can be reasoned about under complex observation, to demonstrating how learning-based representations can be used to organize observations, expose relations, and articulate behavior in evolving material systems. The structure emphasizes conceptual clarity and representational coherence, providing a disciplined framework for engaging with complexity in contemporary materials research.

Chapter 2

Scientific Inquiry and the Role of Representation

This chapter establishes the technical and methodological grounding for the analyses developed in the remainder of this dissertation by examining how object inquiry has been realized in practice across physics and materials science. The focus is not on introducing a new mode of scientific reasoning, but on clarifying how enduring inquiry practices—centered on objects, observations, representations, relations, and the articulation of behavior—have been implemented through evolving experimental and computational tools. By situating contemporary learning-based methods within this broader lineage, the chapter emphasizes continuity in the structure of inquiry alongside evolution in its technical realization.

The chapter opens with the section *Evolution of Object Inquiry in Physics and Materials Science*, which reviews canonical implementations of object inquiry and traces how its constituent modules have developed historically. This section discusses how objects of interest are defined in physics and materials research, how observational access has expanded through advances in experimental instrumentation, and how representational constructions have evolved to organize increasingly rich observations. Particular attention is given to the study of material dynamics, where representations are used to relate observations across time, space, and experimental condition in order to articulate recurring patterns of behavior. Throughout this discussion, the underlying object-inquiry framework remains stable, while the technical means of realizing its components become progressively more expressive.

The chapter then turns to the section *Learning-Based Representations for Object Inquiry*, which examines how contemporary machine-learning and deep-learning methods extend canonical representational practices. This section introduces deep learning as a statistically optimized approach to constructing representations that can accommodate high-dimensional, heterogeneous observations without altering the fundamental logic of inquiry. Within this framework, different representational families are reviewed according to the kinds of relations they encode among observations. Deep generative

models are discussed as similarity-based representations that organize observational variability into coherent latent spaces, while attention-based transformer architectures are examined as token-level relational representations that explicitly model associations among observational components such as spatial regions, temporal segments, or measurement channels.

The distinctions developed across these sections clarify how different representational strategies organize observations and express relations in materially distinct ways. This framing provides the technical context for the methodological chapters that follow, where specific modeling choices are introduced and examined in detail. In particular, the contrast between similarity-based organization and explicit relational modulation serves as a guiding theme for understanding how learning-based representations can be employed to study material dynamics while remaining aligned with long-standing scientific practices.

2.1 Evolution of Object Inquiry in Physics and Materials Science

Scientific inquiry in the physical sciences can be formulated as an object-inquiry process composed of several interrelated modules: the definition of an object of interest, the acquisition of observations, the construction of representations to organize those observations, the examination of relations among them, and the articulation of object behavior through recurring patterns. This modular structure has remained largely stable across the history of physics and materials research. What has evolved over time are the technical realizations of these modules, driven primarily by advances in experimental capability and analytical methodology, rather than by changes in the fundamental manner of inquiry.

In early physical sciences, objects were often idealized entities such as particles, fields, or continua, and observations were limited in both resolution and scope. Measurements were mapped into analytically defined representational spaces—coordinate systems, state variables, and phase spaces—that allowed relationships among observables to be expressed explicitly. Within these representations, recurring patterns such as trajectories, equilibria, or conservation laws could be identified, enabling articulation of object behavior and extrapolation beyond the specific conditions directly observed. The predictive power of such inquiry relied on the coherence between observation, representation, and relational structure.

As materials science emerged as a distinct discipline, the objects of inquiry became structurally and chemically more complex, and observational tech-

niques diversified accordingly. Diffraction experiments organized scattering intensities into reciprocal-space representations to resolve crystallographic symmetry and atomic arrangement [3], while spectroscopic measurements introduced representations tied to electronic and chemical degrees of freedom [4]. These advances extended observational access to finer structural details and additional physical dimensions, necessitating richer representational constructs such as order parameters, structural descriptors, and reduced coordinates. Representations thus evolved alongside observational capability, providing structured domains in which material-specific patterns could be identified and related to macroscopic behavior.

The investigation of material dynamics further amplified this co-evolution. Dynamic phenomena are inherently associated with temporal progression, spatial heterogeneity, and sensitivity to external stimuli. Historically, limitations in temporal resolution constrained dynamic studies to comparisons between discrete initial and final states. Nevertheless, even under such constraints, representations enabled the identification of recurring modes of evolution—such as phase transformation pathways or defect migration patterns—through relational comparison across configurations [13,18]. Here, behavior was articulated not through direct observation of continuous change, but through structured interpretation of patterns spanning multiple observations.

Ongoing advances in experimental instrumentation have progressively reduced observational barriers, extending access to faster dynamics, smaller length scales, and increasingly localized phenomena. Ultrafast electron and X-ray techniques probe structural evolution on femtosecond timescales [7,30], while atomic-resolution microscopy captures defect motion and lattice distortions at the level of individual atomic columns [54]. At the same time, multiscale and multimodal approaches integrate information across spatial resolutions and measurement types, combining structural, chemical, electronic, and mechanical observations within synchronized experimental frameworks [1, 26, 32, 33, 55]. These developments expand not only the quantity of observations but also their structural diversity and relational richness.

As observational access advances toward ever finer temporal and spatial scales, representations must correspondingly evolve to organize increasingly complex observational ensembles. Representational constructions have been extended to incorporate temporal indexing, spatial decomposition, and alignment across modalities, enabling relations to be examined not only among observations themselves but also across scales and measurement channels. In this context, representations function as integrative structures that mediate between raw observations and the identification of coherent

patterns associated with dynamic behavior.

Importantly, these developments reflect a progressive enrichment of the technical realization of object inquiry rather than a transformation of its underlying logic. Objects continue to be investigated through organized collections of observations, and behavior continues to be articulated through recognition of relational patterns that generalize beyond specific measurements. What changes is the expressive capacity required of representational constructs. As observational complexity increases, representations increasingly absorb the burden of organizing variability, aligning heterogeneous data, and supporting systematic examination of relations.

This historical co-evolution of observation and representation naturally motivates approaches that construct representational spaces through systematic and adaptive means. Learning-based representations can thus be understood as a continuation of this trajectory, providing statistically optimized mechanisms for organizing complex observations while preserving the object-inquiry framework. At the same time, the shift toward highly expressive representations introduces new challenges in maintaining coherent and interpretable relations among observations. The following section examines pre–deep learning approaches that bridge analytically designed representations and fully learning-based systems, highlighting how relational structure has been modeled prior to contemporary deep learning methods.

2.2 Deep Learning Representations as an Extension of Object Inquiry

The progression from canonical and pre–deep learning representations toward deep learning reflects a continued evolution in the technical realization of object inquiry. Rather than introducing a new manner of scientific reasoning, deep learning provides a means of constructing highly expressive representation spaces through statistical optimization. In this setting, representations are no longer specified analytically or through handcrafted features alone, but are learned directly from data in a manner that adapts to the complexity, dimensionality, and heterogeneity of modern material observations [38, 44, 46].

Classical machine-learning approaches, including regression models, kernel methods, and support vector machines, can be understood as statistically optimized extensions of canonical inquiry [46, 56, 57]. These methods rely on predefined descriptors and feature spaces to organize observations and model relationships among them. While effective within constrained settings, their representational capacity is limited by the structure imposed a priori. As ex-

perimental datasets expanded to include high-dimensional image sequences, hyperspectral measurements, and synchronized multimodal observations, the rigidity of feature-based representations increasingly restricted their ability to capture complex and coupled variation present in material dynamics.

Deep learning addresses this limitation by learning representations and relational structure jointly. Neural networks construct nonlinear mappings from observations to latent representation spaces whose geometry reflects statistical regularities in the data [38, 44]. In materials research, such representations have been shown to organize time-resolved and spatially resolved observations into coherent embeddings that encode correlations across successive configurations, spatial neighborhoods, and measurement modalities [15]. Convolutional and spatiotemporal architectures, including two- and three-dimensional convolutional neural networks and recurrent variants, are particularly effective at capturing localized evolution patterns in dynamic material data [50–52, 58]. Through these constructions, deep learning representations absorb substantial observational variability while preserving continuity across space and time.

The expressive capacity of deep learning representations enables accurate prediction and reconstruction of dynamic material responses directly from observational data. Models trained on time-resolved microscopy sequences have successfully organized complex patterns associated with defect motion, domain switching, phase-boundary migration, and other dynamic phenomena [36, 59–65]. In such applications, learned representations implicitly encode high-dimensional relationships among observations, linking successive configurations into structured trajectories within representation space.

At the same time, this representational power introduces a fundamental challenge for object inquiry. While deep learning embeddings effectively organize observations and accommodate nonlinear variability, the relations encoded within these spaces are often difficult to interpret explicitly. Proximity, similarity, or transition structure in latent space does not necessarily correspond to clearly articulated physical relationships among observations. As a result, although deep learning representations support pattern recognition and generalization, the coherence of the modeled relations may be insufficient for directly articulating behavior patterns in a transparent and interpretable manner.

This tension between expressive capacity and relational clarity motivates closer examination of specific learning-based representational families. Some architectures impose additional structure on representation space or explicitly modulate relational interactions among observations, offering greater interpretability without sacrificing expressiveness. In particular, deep generative models organize variability through structured latent distributions, while

attention-based models selectively emphasize relational dependencies across space, time, or modality. These approaches are examined in the following sections as exemplars of learning-based representations that more closely align with the relational requirements of object inquiry in material dynamics.

2.3 Relational Requirements and Learning-Based Representations

Within the object-inquiry framework developed in this dissertation, representations are not evaluated solely by their capacity to transform observations into internal features, but by whether the relations encoded in representation space can be exposed, examined, and interpreted back in the observation space. Inquiry is not satisfied by internal organization alone; it requires that representational relations support interpretation of how observations relate to one another in ways that are meaningful for articulating object behavior. This distinction is essential for understanding why many deep learning models, despite their predictive success, are often regarded as black boxes in scientific contexts.

A large class of deep neural networks—including conventional convolutional neural networks [50, 66] used for classification or regression—operate primarily as predictive mappings,

$$\hat{y} = f_\theta(\mathbf{x}), \tag{2.1}$$

optimized to minimize task-specific loss functions []. While such models necessarily construct internal feature hierarchies, the relations among representations are not explicitly structured for interpretation, nor are they readily exposable as relations among observations. In these cases, representation space serves the purpose of prediction rather than inquiry: relations are instrumental, internal, and task-dependent, offering limited insight into how observations relate to one another beyond the output objective [43, 67, 68]. This opacity underlies the characterization of such models as black-box systems in scientific applications.

For learning-based representations to be suitable for object inquiry, stronger relational requirements must be met. First, relations encoded in representation space must correspond to interpretable relations among observations, such as similarity, variation, interaction, or progression. Second, these relations must be accessible for analysis, comparison, or visualization, allowing them to inform understanding of object behavior rather than merely supporting prediction.

At a foundational level, inquiry-capable representations often begin by stabilizing similarity relations among observations. Metric learning and

contrastive learning methods [69–71] explicitly regulate representational geometry so that proximity in representation space reflects consistent similarity in observation space [46, 56]. In these models, relations of the form

$$R_{ij}^{\text{sim}} = d(\mathbf{z}_i, \mathbf{z}_j) \tag{2.2}$$

are not incidental but intentionally constructed, enabling similarity relations to be interpreted and compared across experimental conditions. While such representations expose meaningful relations, they primarily support static comparison and do not, by themselves, articulate how observations may vary.

To address variability explicitly, generative models [43, 72, 73] construct representations in which relations among observations are organized through similarity and continuity. In these models, observations $\mathbf{x}$ are associated with latent variables $\mathbf{z}$ through a learned joint distribution,

$$p_\theta(\mathbf{x}, \mathbf{z}) = p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z}), \tag{2.3}$$

as realized in variational autoencoders, generative adversarial networks, normalizing flows, and diffusion models [43, 47, 48, 74]. Here, proximity and smooth trajectories in latent space correspond to similarity and gradual variation in observation space, allowing relations of variability to be examined explicitly. Importantly, the value of these models for inquiry lies not in their ability to generate samples, but in how the structure of latent space exposes relations among observed configurations that can be mapped back to observable change.

As inquiry moves beyond similarity and variability toward understanding how different components of observations interact, representations must expose relational structure at a finer granularity. Attention-based architectures address this requirement by modeling relations directly among tokens that represent components of observations, such as image patches, temporal segments, or measurement channels [49, 75, 76]. In attention mechanisms, relations are encoded through learned weights,

$$\alpha_{ij} = \text{softmax}\left(\frac{\mathbf{q}_i^\top \mathbf{k}_j}{\sqrt{d}}\right), \tag{2.4}$$

which determine how strongly token j contributes to the representation of token i [49]. These weights can be interpreted as relational maps that expose contextual interaction among observational components. Unlike implicit feature hierarchies, attention weights are directly inspectable, allowing relations in representation space to be interpreted as associations among observable elements.

Finally, when observations are indexed by time or ordered conditions, inquiry requires representations that preserve relations of progression [77,78]. Temporal models encode relations across successive observations,

$$\mathbf{z}_{t+1} = g_\theta(\mathbf{z}_t), \tag{2.5}$$

exposing continuity and directional structure that can be interpreted in terms of evolving object behavior. In such representations, relations among observations are not only exposable but ordered, supporting interpretation of dynamic processes.

Viewed through this lens, the distinction between black-box prediction and inquiry-capable representation becomes clear. Models that merely transform observations to outputs may achieve high predictive accuracy while obscuring relational structure. In contrast, representations designed to expose similarity, variability, interaction, and progression provide relational scaffolds that can be mapped back to observation space and interpreted in terms of object behavior. In the study of material dynamics, where understanding relations among evolving configurations is central, generative and attention-based representations satisfy complementary inquiry requirements and therefore form the primary focus of the methodological developments in subsequent chapters.

Chapter 3

Generative Inquiry Framework of Material Dynamics

This chapter develops and applies the generative modeling approach that forms the methodological core of this thesis. The preceding chapters established the conceptual foundation for understanding material dynamics as an object-inquiry process, in which objects are accessed through observations, organized through representations, and interpreted through relations that articulate dynamic behavior. Within this framework, material dynamics are only partially accessible through experimental observation. Owing to limitations of temporal resolution, measurement rate, and environmental control, experimental imaging techniques such as Coherent X-ray Diffraction Imaging (CXDI) and X-ray Absorption Fine Structure Computed Tomography (XAFS-CT) yield discrete, static representations of phenomena that are inherently continuous. The inability to directly capture intermediate transformations between measured states leads to structural sparsity in dynamic observation. In this thesis, such sparsity is treated not as a deficiency to be eliminated, but as a condition that motivates systematic construction of representations capable of organizing and relating observed configurations.

The generative modeling framework proposed in this chapter operationalizes this perspective within the object-inquiry process. Deep generative modeling is employed to construct a representation space in which experimentally observed material states are organized according to statistical similarity and continuity, realizing the first step of inquiry by embedding discrete observations into a coherent representational domain. Within this formulation, the latent space of a trained generative model functions as a structured space of material configurations implied by the experimental record. Local stochastic sampling within this space produces additional configurations that are consistent with the organization of observed data, extending the representational scope beyond individual measurements without asserting reconstruction of specific unobserved physical states. The second step of object inquiry is realized through systematic exploration of relations

among configurations within this representational space. By examining how generated configurations relate to observed ones—through proximity, continuity, and structured variation—the framework enables dynamic behavior to be articulated from relations among observations rather than from isolated states. In this way, generative modeling and stochastic sampling jointly support interpretation of material evolution as an organized pattern of variation shaped by both material processes and measurement conditions.

The chapter is structured to establish and examine this generative modeling paradigm across increasing levels of physical complexity. Section 3.1 introduces the formulation and implementation of the framework, detailing how representation construction and relational exploration are combined to support object inquiry under sparse observation. Sections 3.2–3.4 then demonstrate its application to three representative systems: (I) rigid-body motion in a tantalum (Ta) test chart, (II) stochastic diffusion of nanoparticles in an aqueous polymer solution, and (III) chemically coupled sulfidation in aging rubber–brass composites. Each case study examines how relations among observed and generated configurations organize into dominant modes of variation that characterize dynamic behavior. The chapter concludes with a discussion that generalizes these findings and situates generative modeling as a representational bridge between experimental observation and systematic characterization of material evolution.

3.1 Framework Overview

This section introduces the methodological foundation of the generative modeling framework developed to extract structural understanding from incomplete experimental observations of material evolution. In many time-resolved imaging experiments, what is recorded are discrete configurations rather than continuous trajectories, meaning that only partial information about the underlying transformations is available. The framework addresses this incompleteness by constructing a probabilistic generative representation of material dynamics through deep learning. A trained generative adversarial network (GAN) learns the statistical organization of experimentally observed configurations, encoding how material states are distributed under the combined influence of physical processes and measurement conditions.

By coupling this generative representation with Monte Carlo sampling, the framework systematically explores the learned latent space to construct plausible intermediate configurations that are consistent with the organization of the observed data. Through this exploration, regularities in how configurations vary and relate to one another become explicit, revealing the geometric and statistical structure of the representational space. The

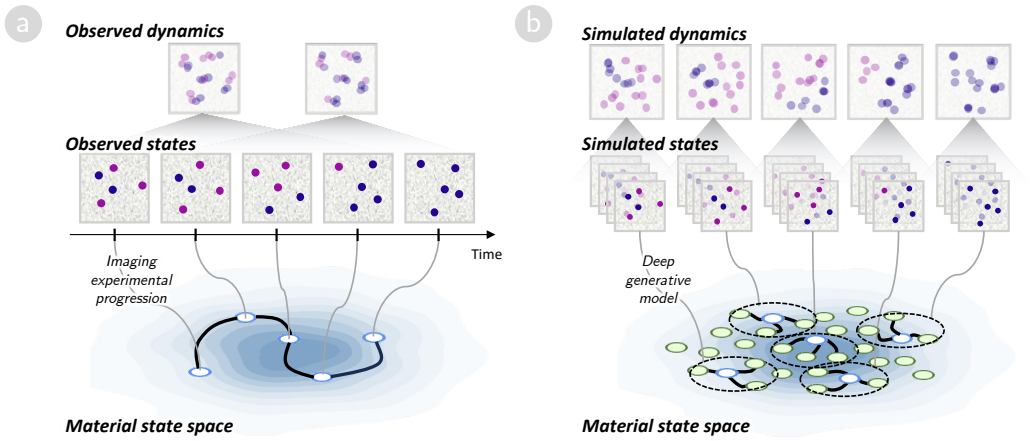

Figure 3.1: Overview of the two-stage analytical framework for extracting insights into material dynamics from experimental images. (**a**) Deep Generative Model Training: A GAN is trained on experimental images to construct latent space of material states. (**b**) Monte Carlo Simulation: The trained model generates material variations through latent space perturbations, which are then analyzed to identify distinct transformation behaviors and dynamic patterns.

integration of these two elements transforms the generative model into a practical tool for examining how materials may evolve within the range of variability supported by experimental observation, providing a statistical basis for analyzing the geometry of material transformations. The following subsections detail the conceptual formulation of this approach, its computational implementation, and the experimental settings used for validation.

3.1.1 From Generative Modeling to Monte Carlo Sampling

Time-resolved imaging techniques such as transmission electron microscopy (TEM) [31], scanning electron microscopy (SEM) [1], or coherent X-ray diffraction imaging (CXDI) [30,32,33] capture the evolution of materials as a sequence of discrete measurement outcomes. Each recorded configuration reflects not only the instantaneous physical state of the specimen but also the particular temporal and environmental conditions under which the observation occurs—the stage of the phenomenon's progression, the detector response, and the random fluctuations inherent to the measurement process [79–81]. Consequently, the variations among successive images embody both genuine physical transformation and stochastic modulation introduced

by observation itself. The empirical record is therefore not merely incomplete but intrinsically *distributional*: what is seen at any moment represents one realization drawn from a broader set of possible observable outcomes shaped jointly by material dynamics and experimental conditioning. This coupling of evolution and observation complicates direct interpretation of change, since apparent differences may arise as much from measurement variability as from the underlying transformation.

Deep generative modeling provides a means to represent this distributional landscape of observation [66, 82–84]. By learning the statistical regularities embedded in experimental data, a generative model constructs an internal representation of how observable configurations are distributed under the combined influence of material evolution and measurement conditions. Rather than treating variations across images as experimental noise to be suppressed, the model treats them as meaningful manifestations of variability—expressions of how the material may appear under slightly different conditions or stages of transformation [37, 85]. In this view, generative modeling functions as a representational mechanism that organizes observed variability into a coherent probability field over material configurations.

Formally, a generative model defines a joint probability distribution

$$p_\theta(\mathbf{x}, \mathbf{z}) = p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z}), \tag{3.1}$$

where $\mathbf{x}$ denotes the observable material configuration, $\mathbf{z}$ represents latent variables parameterizing variation in appearance and structure, and θ are the learned parameters of the model. Through optimization of θ, the network internalizes correlation structures that render the measured data self-consistent, effectively learning how observed configurations are distributed within the space of possible observations supported by the experiment. Sampling latent variables $\mathbf{z}' \sim p(\mathbf{z})$ and decoding them via G_θ produces synthetic configurations $\mathbf{x}' = G_\theta(\mathbf{z}')$. These generated realizations are not extrapolations of specific frames but *hypothetical material states*: configurations that are statistically consistent with the observed data and the conditions under which they were acquired.

This generative mechanism thereby extends the representational scope of time-resolved imaging. Each point in latent space corresponds to a configuration that could plausibly be observed given the learned organization of the data, and proximity within that space encodes similarity in observable structure. As illustrated schematically in Fig. 3.1, the latent space provides a continuous representational domain in which discrete experimentally observed configurations occupy sparse locations, while surrounding regions correspond to additional, unobserved but statistically consistent configurations. By sampling across the latent domain, the model constructs ensembles

of possible intermediate configurations, collectively describing the range of material appearances compatible with both the experimental record and its inherent variability. This representation acknowledges that change in time-resolved images cannot be interpreted deterministically; instead, it must be understood as sampling from a distribution shaped by evolving material structure and observational modulation.

Monte Carlo (MC) sampling [86] extends this representational strategy by providing a systematic means of exploring the latent space learned by the generative model. While the model defines a high-dimensional probability field over possible configurations, MC sampling enables controlled traversal of this field through stochastic perturbations of latent coordinates. Sequences of latent vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ and corresponding configurations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ reveal how configurations are distributed locally and how smoothly one configuration can be transformed into another within the learned representation. Regions of high sampling density indicate configurations that recur under the learned organization, while smoothly connected regions indicate continuous variation in observable structure.

The rationale for integrating Monte Carlo sampling with generative modeling is therefore representational and analytical. From a representational perspective, sampling converts an implicit probability field into explicit ensembles of configurations that can be visualized, compared, and quantified. Analytically, MC sampling provides a systematic approach to estimate observables—such as morphological descriptors or similarity metrics—across the latent space, enabling statistical characterization of how material configurations vary under the conditions encoded by the data. Through iterative sampling and aggregation, the framework exposes how the generative model organizes observed and hypothetical configurations into a coherent space of variation.

In this sense, the combination of deep generative learning and Monte Carlo exploration transforms the sparsity of time-resolved observations from a limitation into an analytical resource. Rather than attempting to recover a single unobserved trajectory, the framework constructs a representational ensemble that captures the range of material states consistent with experimental observation. What appears as stochastic fluctuation in individual images becomes, through this representation, a structured distribution that supports systematic analysis of material evolution.

3.1.2 Implementation of the Generative Inquiry Framework

Building upon the conceptual foundation established above, the proposed framework integrates deep generative modeling with stochastic sampling and statistical analysis to examine material transformations captured through time-resolved microscopy. In such experiments, each image represents a measurement realization conditioned by the temporal stage of the phenomenon and the stochastic character of observation. The recorded data thus convey both genuine structural evolution and probabilistic variability arising from experimental context. Within this setting, the implementation treats the experimental record as a measurement-conditioned probability distribution of material states and develops a computational procedure to generate, organize, and analyze the range of transformations represented within that distribution.

The principal objective of the framework is to enable systematic characterization of transformation behavior by constructing ensembles of synthetic yet statistically consistent material configurations. Deep generative modeling is used to learn a latent representation that encodes the coupled influence of material dynamics and measurement conditioning. Monte Carlo sampling then explores local probabilistic variation within this latent space, revealing how the learned distribution organizes possible pathways of variation between configurations. The resulting synthetic configurations serve as representative samples of the distribution implied by experimental observation, collectively delineating the probability field that characterizes material evolution under the given measurement conditions.

The implementation proceeds through two interdependent stages that together constitute the operational basis of the framework. The first stage involves training a deep generative model to establish a structured latent space representing observed material configurations and their statistical organization. The second stage applies Monte Carlo sampling within this space to perform stochastic exploration and quantitative analysis of variation across configurations. The integration of these two stages yields a coherent computational environment for generating, organizing, and quantifying the statistical structure of dynamic material systems as observed experimentally.

3.1.2.1 Generative Model Training

The first stage of the framework develops a deep generative model to construct a continuous latent representation of material configurations as observed under time-resolved microscopy. The model is trained to *reproduce the empirical distribution of experimental images* so as to statistically infer

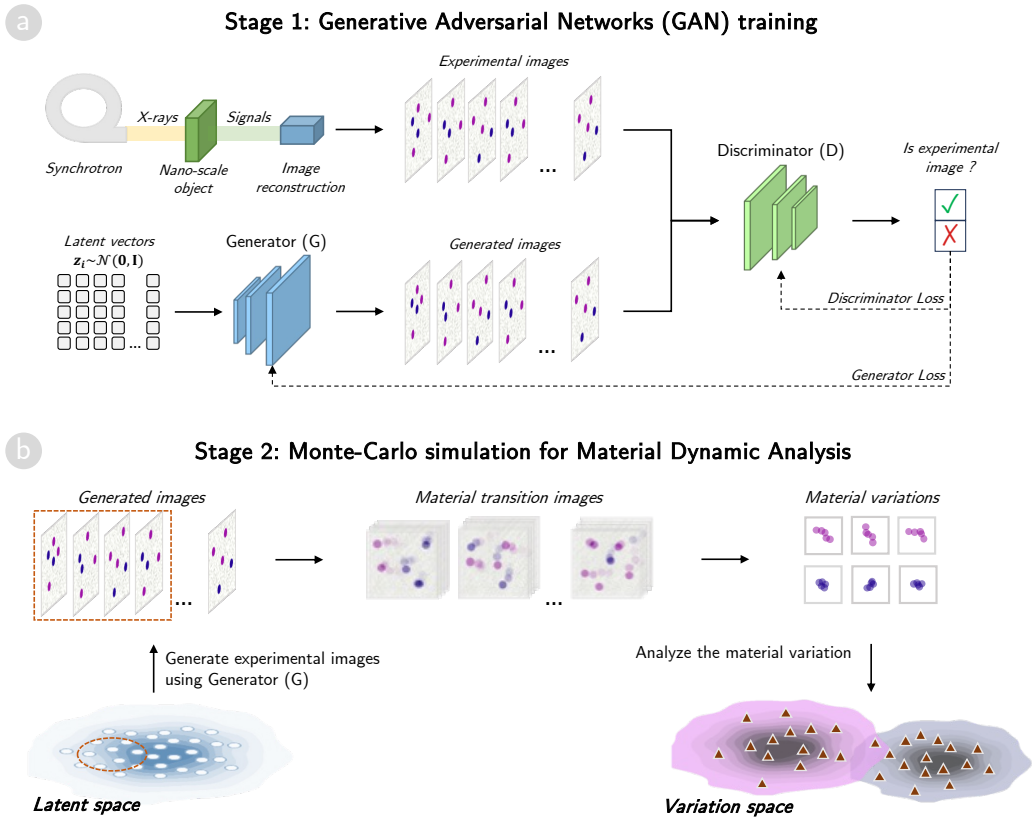

Figure 3.2: Overview of the two-stage analytical framework for extracting insights into material dynamics from experimental images. (**a**) Deep Generative Model Training: A GAN is trained on experimental images to construct latent space of material states. (**b**) Monte Carlo Simulation: The trained model generates material variations through latent space perturbations, which are then analyzed to identify distinct transformation behaviors and dynamic patterns.

the transformations that render those observations mutually consistent. Through this behavior, reproduction becomes an inferential process: by learning to recreate the variability present in the measurements, the model internalizes the probabilistic structure governing how material states change under both physical dynamics and observational conditioning. The resulting latent representation provides a structured space in which continuity corresponds to physically and experimentally admissible transformations. In this sense, the model does not merely memorize the data—it captures the probabilistic logic of their variation, enabling inference of change through statistical reproduction.

A generative adversarial network (GAN) [43] architecture is adopted for this purpose owing to its capacity to synthesize high-fidelity images from limited datasets while preserving the diversity of observed morphologies. Latent variables are sampled from a multivariate normal distribution $\mathcal{N}(0, I)$ to ensure isotropic representation of material configurations. The model comprises two competing neural networks: a generator $G : \mathbf{Z} \to \mathbf{X}$ that maps latent vectors $z \in \mathbf{Z}$ to synthetic material images $\hat{x} \in \mathbf{X}$, and a discriminator $D : \mathbf{X} \to \mathbb{R}$ that differentiates real experimental images from generated ones. Both networks employ convolutional architectures designed to capture hierarchical spatial correlations reflecting the multiscale organization of material microstructures. The generator progressively refines latent features into spatially resolved images, while the discriminator evaluates their physical plausibility by learning morphological, textural, and contrast-based attributes from empirical data.

Training is performed in the Wasserstein formulation (WGAN) [87], which measures the distance between real and generated distributions through the Earth Mover's metric. The optimization objective is expressed as

$$\begin{aligned} \min_{G} \max_{D} \; \mathcal{L}(D, G) = \mathbb{E}_{x \sim \mathbb{P}_{\mathcal{D}}}[D(x)] - \mathbb{E}_{z \sim \mathbb{P}_{\mathbf{Z}}}[D(G(z))] \\ + \lambda_{\text{gp}} \, \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} \left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \right], \end{aligned} \tag{3.2}$$

where $\mathbb{P}_{\mathcal{D}}$ denotes the empirical distribution of experimental data, $\mathbb{P}_{\mathbf{Z}}$ the prior distribution of latent vectors, and λ_{gp} the coefficient of the gradient penalty term [88]. This formulation enforces Lipschitz continuity in the discriminator and stabilizes adversarial optimization, thereby reducing mode collapse. The generator minimizes this metric by producing synthetic configurations whose distribution converges toward that of real observations, while the discriminator maximizes it to preserve discriminatory sensitivity. Optimization is performed using the Adam algorithm with a learning rate of 10^{-4}, $\beta_1 = 0.5$, and $\beta_2 = 0.9$, balancing gradient stability and convergence speed.

To enhance both fidelity and convergence, a *progressive growing* training strategy [89] is employed. This approach incrementally expands model capacity by introducing higher-resolution layers as training stabilizes. The process begins at a coarse resolution (e.g., 32×32 or 64×64 pixels for two-dimensional data, or 16^3–32^3 voxels for volumetric data), capturing global morphology and long-range structural relationships. As the adversarial loss stabilizes, new convolutional layers are appended to both generator and discriminator, doubling spatial resolution at each stage. During this transition, feature outputs from consecutive resolutions are linearly blended

through a fade-in parameter $\alpha \in [0, 1]$, such that

$$x_{\text{out}} = (1 - \alpha)\, x_{\text{low}} + \alpha\, x_{\text{high}}. \tag{3.3}$$

This blending ensures gradual adaptation to finer spatial detail and prevents instability from abrupt feature changes. Once the transition is complete, lower-resolution branches are fixed and training proceeds at the new scale until convergence.

Adaptive optimization control complements this progressive scheme. Learning rates are gradually decayed with increasing resolution to maintain stable gradients, and batch sizes are inversely scaled to offset the higher memory cost of large tensors. For volumetric datasets, cubic scaling of dimensions necessitates proportionally smaller batches to preserve statistical consistency. Instance normalization is applied in early layers to stabilize global feature magnitudes, while pixel-wise normalization in higher layers maintains local contrast and textural balance. Leaky-ReLU activations [90] ($\alpha = 0.2$) are employed throughout to sustain gradient flow, and spectral normalization is optionally applied to discriminator weights to further enforce the Lipschitz constraint [91].

The progressive growing strategy enhances both computational stability and the interpretability of learned representations. In early training stages, the model captures coarse morphological organization—phase boundaries, domain connectivity, or particle distributions—while subsequent stages refine localized features such as defect networks, strain fields, or compositional heterogeneity [66, 82–84]. This hierarchical learning sequence parallels the physical structure of materials, ensuring that the latent space encodes correlations across scales. The resulting manifold represents a continuous, measurement-conditioned probability field in which neighboring points correspond to incremental and physically coherent transformations. Sampling trajectories through this manifold thus yields statistically inferred pathways of change, derived from the model's capacity to reproduce and therefore to reason about the data distribution itself.

All available experimental observations are utilized in training to maximize the representational completeness of the learned manifold. The objective is not predictive accuracy but the establishment of a statistically and physically coherent space encompassing all observable configurations. Convergence is evaluated by monitoring stabilization of adversarial losses and by assessing the structural and statistical coherence of generated samples relative to empirical data. Upon completion of training, the generator defines a mapping between latent variables and admissible material states, enabling stochastic sampling and statistical interrogation of transformation behavior in the second stage of the framework.

3.1.2.2 Monte Carlo Simulation for Material Dynamic Analysis

The second stage of the framework employs Monte Carlo (MC) sampling within the latent space of the trained generative model to examine how statistically consistent variations of material configurations are organized around reference states. Within the object-inquiry framework, this stage corresponds to the second step of inquiry, in which dynamic behavior is articulated through relations among observations rather than through isolated configurations. Each latent perturbation is treated as a probabilistic realization of how an observed configuration may vary under the coupled influence of material evolution and measurement conditioning, enabling relations among neighboring configurations to be examined systematically. By accumulating ensembles of such perturbations, the framework exposes how variations are distributed locally and globally within the learned latent space. These distributions encode relational structure among configurations, such as continuity, clustering, and directional variation, which can be interpreted as signatures of dynamic behavior when mapped back to the observation space. The objective is therefore twofold: to represent transformations as structured variation around observed configurations, and to characterize how relations among these variations organize into coherent patterns that describe material behavior under experimental observation.

In conventional Markov Chain Monte Carlo (MCMC) schemes [92], samples are generated sequentially through a transition kernel that links successive states, producing an equilibrium distribution approximating a target probability density. In contrast, the present formulation employs a non-sequential sampling mechanism suited to the geometry of the generative latent space, where proximity encodes physical similarity rather than temporal adjacency. Each latent vector acts as an independent reference, from which nearby states are generated simultaneously by bounded stochastic perturbations of the latent coordinates. This approach retains the essence of Monte Carlo reasoning—stochastic exploration of a probability field—while adapting it to a static ensemble formulation that directly probes how the model organizes permissible variability. Here, locality in latent space embodies a Markovian notion of conditional variation: configurations close to each other correspond to short-range transformations that the model recognizes as physically coherent, even though no explicit sequential chain is constructed.

Let M latent vectors z_i be drawn from the prior distribution $\mathcal{N}(0, I)$. Each z_i corresponds to a reference configuration $\hat{x}_i = G(z_i)$ generated by the trained model. Around each reference point, a neighborhood of latent vectors is constructed by applying random perturbations ϵ_j sampled from a

uniform distribution $\mathcal{U}[-\gamma, \gamma]$:

$$z_{i,j} = z_i + \epsilon_j, \quad j = 1, \ldots, N. \tag{3.4}$$

These perturbed vectors are then decoded by the generator to yield synthetic configurations

$$\hat{x}_{i,j} = G(z_{i,j}), \tag{3.5}$$

which represent hypothetical transformations inferred from the model's learned distribution. Each $\hat{x}_{i,j}$ expresses a *variation* of the reference configuration $\hat{x}_i$, drawn from the range of changes that the model deems admissible according to the statistical relationships internalized during training. The parameter γ defines the perturbation radius in latent space and thus the scope of variability considered physically plausible. Conceptually, γ acts as a probabilistic boundary of admissibility, specifying the degree of variation that can occur without violating the structural coherence encoded in the learned manifold.

The choice of γ is calibrated to the scale of structural change observable in the experimental data. Its maximum permissible value is determined such that the average dissimilarity between generated configurations within this perturbation range does not exceed the empirical differences between consecutive experimental states recorded at the highest temporal resolution. This calibration ensures that MC sampling explores the same order of transformation magnitude as real physical processes, while maintaining statistical continuity with experimentally observed dynamics.

After this coarse constraint defined by γ, a refinement step imposes additional physical filters to identify configurations that remain consistent with measurable material properties. These filters enforce conservation laws and physical invariants such as constant mass or volume, compositional balance, and acceptable limits on derived observables (e.g., density, intensity correlation, or interfacial curvature). Variations violating these conditions are excluded, yielding a refined ensemble of synthetic configurations that represent both geometrically and physically admissible transformations. In this refinement, probabilistic inference of change transitions into inferential delineation of constraint: by filtering the stochastic outcomes, the sampling process reveals which variations the material system—and the measurement conditions—implicitly allow.

For each reference configuration $\hat{x}_i$, the associated ensemble of generated states is denoted

$$V_i = \{\hat{x}_i, \hat{x}_{i,1}, \hat{x}_{i,2}, \ldots, \hat{x}_{i,N}\}. \tag{3.6}$$

Each V_i thus defines a local probabilistic neighborhood that captures the admissible variability of the material state inferred by the model. These

ensembles are interpreted as unordered aggregates rather than time-ordered sequences, reflecting the non-deterministic nature of the reconstruction. Within this formulation, stochastic uncertainty is regarded not as random noise but as a quantitative expression of the multiplicity of equivalent transformation pathways consistent with physical and observational constraints.

Analytical evaluation of the ensembles $\{V_i\}$ enables investigation of how the generative model encodes the internal organization of constraint and variability. From each generated configuration, structural or compositional descriptors are extracted to quantify transformations relative to the reference state. These descriptors—defined through morphological metrics, texture statistics, or surrogate physical parameters—form a representation space in which unsupervised statistical techniques such as clustering, density mapping, or manifold projection are applied. The resulting organization of variations reveals dominant modes of change and the geometric structure of constraint within the learned manifold. In this manner, Monte Carlo sampling converts the statistical imagination of the generative model into an empirical basis for reasoning about what transformations are possible and what restrictions they obey.

Through this probabilistic exploration, the framework transforms the generative model from a static reproducer of configurations into a dynamic inferential instrument. The MC stage extends inference from the generation of individual admissible states to the analysis of how those states collectively express the lawful variability of material dynamics. By systematically generating, refining, and categorizing ensembles of plausible configurations, the framework quantifies both the diversity of inferred changes and the structure of constraints that delimit them, thereby providing an interpretable and data-grounded perspective on the probabilistic nature of material evolution.

3.1.3 Experimental Settings

To assess the effectiveness of the proposed generative modeling framework, a sequence of experimental evaluations was conducted using representative datasets that progressively increase in imaging complexity and physical process diversity. The evaluation is aligned with the framework's principal objective—organizing sparse experimental observations into coherent representational spaces that expose relations among material configurations and support articulation of dynamic behavior. Accordingly, the analysis examines how effectively the framework captures the statistical structure of observed variability and how coherently this structure is expressed through stochastic exploration of the learned representation space.

The evaluation focuses on two complementary representational aspects. The first concerns how the framework organizes variations among material

configurations, reflecting the range of observable transformations supported by experimental data. The second concerns how relations among these variations are structured within the learned representation space, enabling systematic interpretation of dynamic behavior from relations among observations. Together, these aspects characterize how effectively discrete experimental measurements are transformed into interpretable descriptions of material evolution.

Three evaluation criteria operationalize these representational objectives. The first criterion, **fidelity of generated configurations**, examines whether synthetic images produced by the generative model remain consistent with experimentally observed material morphology and structure. High fidelity indicates that the model has internalized the statistical organization of observed configurations. The second criterion, **smoothness of latent variation**, evaluates whether small perturbations in latent space produce coherent and gradual changes in the generated configurations that resemble realistic transitions between experimentally observed states. Together, these two criteria assess how effectively the generative representation captures and organizes variability in material configurations. The third criterion, **coherence of relational structure**, evaluates whether the distributions obtained through Monte Carlo (MC) sampling exhibit organized patterns—such as continuity, clustering, or directional variation—that align with known characteristics of material behavior. This criterion emphasizes interpretability of relational organization rather than recovery of governing rules.

Three case studies were designed to satisfy these criteria while progressively increasing the complexity of material dynamics under consideration. The first case involves two-dimensional (2D) phase images of a tantalum (Ta) test chart obtained using coherent X-ray diffraction imaging (CXDI) [68]. This dataset features deterministic translational and rotational motion without internal structural evolution, providing a controlled baseline for evaluating the framework's ability to organize geometrically coherent variation and to represent smooth trajectories within latent space. Agreement between generated and measured displacements provides a quantitative measure of representational consistency.

The second case employs CXDI phase images capturing the diffusion of gold nanoparticles (NPs) in a polyvinyl alcohol (PVA) matrix [33]. This dataset introduces stochastic, spatially heterogeneous motion characteristic of diffusive dynamics. It tests the framework's capacity to organize non-deterministic variation into structured distributions within representation space. MC sampling around reference configurations reveals how diffusive motion is expressed as structured variability, exposing correlations and spatial limits that are characteristic of diffusion processes.

The third case extends the analysis to three-dimensional (3D) multi-channel data using X-ray absorption fine structure computed tomography (XAFS-CT) of brass–rubber composites undergoing sulfidation and aging [23]. The dataset comprises volumetric reconstructions with three channels corresponding to copper valence states (Cu^0, Cu^{1+}, Cu^{2+}). This system presents coupled morphological and chemical variation, offering a rigorous test for multivariate representation learning. The analysis focuses on how ensembles derived from MC sampling capture correlated variation across structural and chemical channels, revealing how relations among modalities are organized within the learned representation space.

For each case, the generative model architecture is adapted to the dimensionality and structural characteristics of the data. Two-dimensional convolutional GANs are used for the CXDI datasets, while a three-dimensional convolutional GAN is implemented for the XAFS-CT dataset to preserve volumetric coherence. Network depth, kernel size, and resolution progression follow the protocol in Section 3.1.2.1, with parameter scaling adjusted to ensure stable optimization at each resolution level. Training is conducted on NVIDIA A100 GPUs (80 GB memory) to maintain convergence stability across modalities.

In all evaluations, a *material variation* is defined as an ensemble of generated configurations obtained by perturbing the latent representation of a reference state and sampling the surrounding representational neighborhood, as described in Section 3.1.2.2. Each ensemble represents the local organization of variability around an observed configuration, providing an empirical description of how material appearances are distributed under the experimental conditions. Statistical analysis of these ensembles through clustering and manifold projection characterizes the relational structure of variation, supporting interpretation of dynamic behavior from relations among observed and generated configurations. This consistent analytical formulation across all datasets provides a unified basis for evaluating how effectively the framework organizes material dynamics from experimentally sparse observations.

3.2 Case Study I: Proof-of-Concept with Ta Test Chart

This case study provides a proof-of-concept validation demonstrating that the proposed generative modeling framework can organize and express physically meaningful transformation behavior within its learned representation space. Specifically, it examines whether the latent representation learned

Table 3.1: Overview of datasets and imaging techniques (IDs shared with Table 3.2).

Case study	**Dataset**	**Imaging technique**
1	Ta test chart [68]	CXDI
2	NP diffusion [33]	CXDI
3	Aging brass clumps [23]	3D XAFS-CT

Table 3.2: Data specifications corresponding to the datasets in Table 3.1.

Case study	**# Images**	**Image size**	**Resolution**
1	10,530	178×178 pixels	20.72 nm pixel^{-1}
2	2,000	142×142 pixels	40.58 nm pixel^{-1}
3	4,956	$32 \times 32 \times 32$ voxels	0.65 μm voxel^{-1}

from experimental data can coherently organize static material configurations together with their associated modes of variation, such that dynamic behavior can be articulated through relations revealed by localized exploration of the latent space. The tantalum (Ta) test chart experiment provides a controlled and interpretable system in which all transformations are limited to rigid-body translation and rotation, without internal deformation of the object.

In this setting, variations expressed through local traversal of the latent space are expected to correspond to translations or rotations consistent with the deterministic motion imposed in the experiment. Because the material itself remains structurally unchanged, the representational organization should reflect rigid-body motion as smooth, low-dimensional variation rather than as heterogeneous or deformable change. Successful articulation of such behavior demonstrates that the framework is capable of organizing observed configurations and their relations in a manner that faithfully reflects known physical motion, establishing a baseline for subsequent case studies involving stochastic motion and coupled structural or chemical variation examined through Monte Carlo (MC) sampling.

The experimental dataset was obtained using coherent X-ray diffraction imaging (CXDI) [68], which recorded the horizontal translation of a Ta test chart moving at a constant velocity of 340 nm s^{-1} under triangular-aperture illumination. The experiment yielded 1,755 phase-reconstructed frames, each corresponding to a 7 ms exposure window. Because the transformation involved only rigid-body motion, the dataset provides a direct physical reference for evaluating whether the latent space can infer admissible geometric

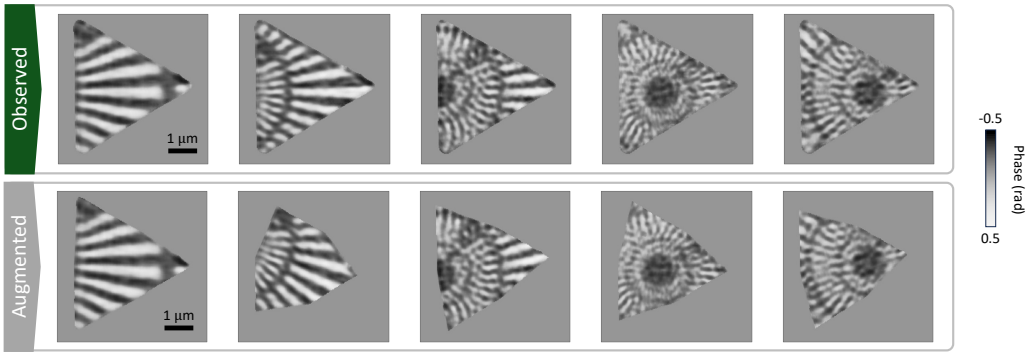

Figure 3.3: Phase images retrieved from Coherent X-ray Diffraction Imaging (CXDI) capturing horizontal translation of a Ta test chart, along with augmented rotation.

changes that correspond to the observed deterministic transformations.

To enhance the statistical robustness of the training data while maintaining physical realism, each frame was augmented by five random in-plane rotations within the range $[-30°, 30°]$, expanding the dataset to 10,530 images. This augmentation broadens the sampling of configuration space without introducing deformation, allowing the model to learn invariant geometric features and transformation symmetries.

The GAN was trained according to the progressive growing procedure outlined in Section 3.1.2.1, enabling hierarchical learning of both global morphology and local structural contrasts. Specifically, a Progressive Growing GAN (PG-GAN) [89] within the Wasserstein GAN–gradient penalty (WGAN-GP) framework [87] was employed. Training began with low-resolution images (e.g., 4×4 pixels) and progressively doubled the image size as new convolutional layers were appended to both the generator and discriminator. At each resolution stage, the generator applied 3×3 convolutions with upsampling, while the discriminator mirrored this structure with 3×3 convolutions followed by average-pooling downsampling. The number of convolutional filters was reduced proportionally as image resolution increased, ensuring that coarse-scale layers captured global structure whereas fine-scale layers refined local details. The architectural configuration used in this case study is illustrated in Figure 3.4 (see more details in Appendix A.1). Latent vectors were sampled from $\mathcal{N}(0, \mathbf{I})$ with dimensionality $d = 32$. For visualization, images were center-cropped to highlight regions unaffected by peripheral diffraction artifacts (Figure 3.5a), whereas full frames were retained during training to preserve complete structural information.

After model convergence, the latent space was explored using Monte

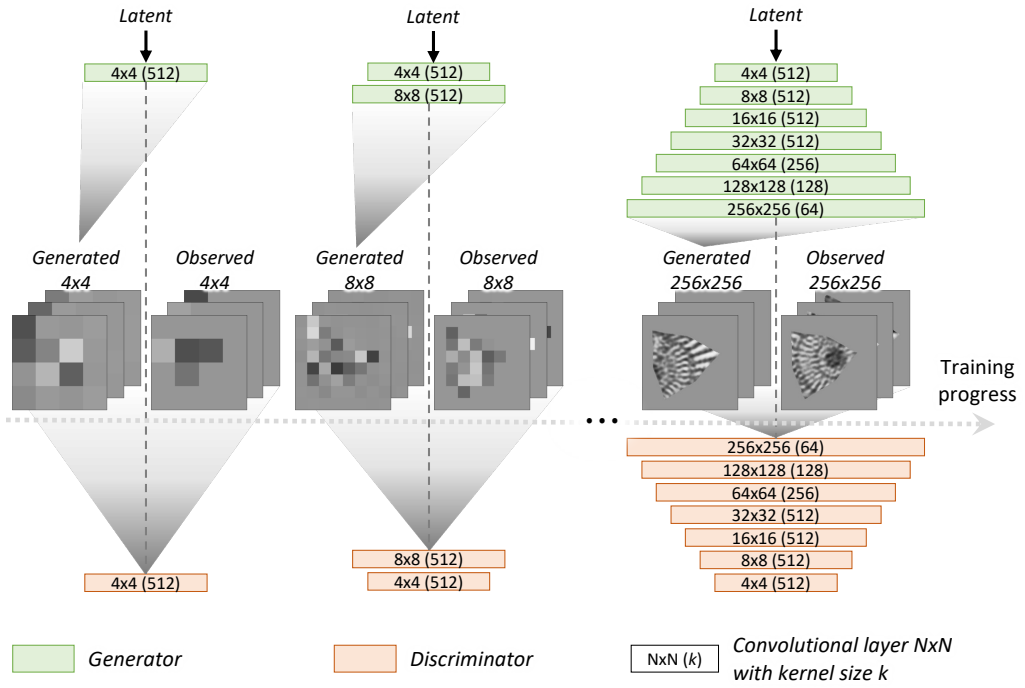

Figure 3.4: Schematic diagram illustrating the GAN architecture and progressive growing strategy used to synthesize CXDI phase images of the Ta test chart.

Carlo (MC) sampling as described in Section 3.1.2.2. A latent vector z corresponding to a synthesized reference image $\hat{x}_z = G(z)$ was selected, and 500 perturbed latent vectors were generated by adding random deviations $\epsilon_j \sim \mathcal{U}[-1, 1]$ to z. Although this perturbation range appears broad relative to the unit-variance prior, it corresponds to moderate local deviations in the high-dimensional latent geometry, ensuring that sampling remains within the manifold of admissible transformations (see details in Appendix A.2). The generator decoded these latent vectors into synthetic phase images,

$$\hat{x}_{z,j} = G(z + \epsilon_j), \tag{3.7}$$

forming an ensemble of probabilistic variations,

$$V_z = \{\hat{x}_z, \hat{x}_{z,1}, \hat{x}_{z,2}, \ldots, \hat{x}_{z,500}\}. \tag{3.8}$$

Each generated image represents an *inferred change*—a transformation plausibly admissible under the rigid-body constraint—relative to the reference configuration. Because the physical process in this case is fully deterministic, deviations from geometric coherence would indicate imperfections in

how the latent manifold encodes admissible transformations. To ensure that the inferred changes remained consistent with the deterministic constraints of the experiment, a structural filtering process was applied using template matching [93]. A distinctive probe line pattern on the test chart was defined as the reference template, and the intersection-over-union (IOU) metric [94] was computed as

$$\mathrm{IOU}_{2D} = \frac{|A_{\mathrm{sampled}} \cap A_{\mathrm{reference}}|}{|A_{\mathrm{sampled}} \cup A_{\mathrm{reference}}|}, \tag{3.9}$$

where $A_{\mathrm{reference}}$ and A_{sampled} denote the pixel regions of the reference and sampled configurations, respectively. Only generated images with $\mathrm{IOU}_{2D} \geq 0.8$ were retained, yielding 482 valid configurations that satisfied the geometric constraint of structural preservation. This filtering enforced the deterministic constraint of rigidity while isolating the probabilistic variability inferred by the model as admissible change.

To construct representative motion sequences, all valid generated images were reoriented to remove random offsets. Template matching was then applied in a sliding window across each image to identify the position of maximum similarity, from which translation and rotation magnitudes were computed. The images were arranged according to these displacement measures into continuous sequences of translation and rotation (Figures 3.5b–c). The resulting sequences display smooth, coherent progressions consistent with rigid-body motion, demonstrating that local perturbations in the latent space successfully infer physically admissible changes governed by the preserved rigidity constraint.

A quantitative baseline for evaluating transformation coherence was established using the experimental dataset. A subset of 641 frames with $\mathrm{IOU}_{2D} \geq 0.8$ was extracted from the original CXDI sequence, providing a benchmark for spatial continuity under controlled motion. Figure 3.6a compares the IOU distributions for experimental and generated sequences. The generated images exhibit slightly higher mean IOU values, indicating that transformations inferred through latent-space sampling correspond to smaller, smoother displacements—consistent with probabilistic interpolation within the learned constraint manifold.

To further confirm that the generative inference preserves deterministic structural constraints, two slit-like probe features, d_1 and d_2, were tracked across all generated configurations using a template-matching-based localization approach [93]. Each generated image was first reoriented to remove rotational offsets, and the reference template was horizontally slid across the aligned image to identify the region with the highest similarity, quantified by IOU_{2D} score. The dimensions of the matched slits were then measured to

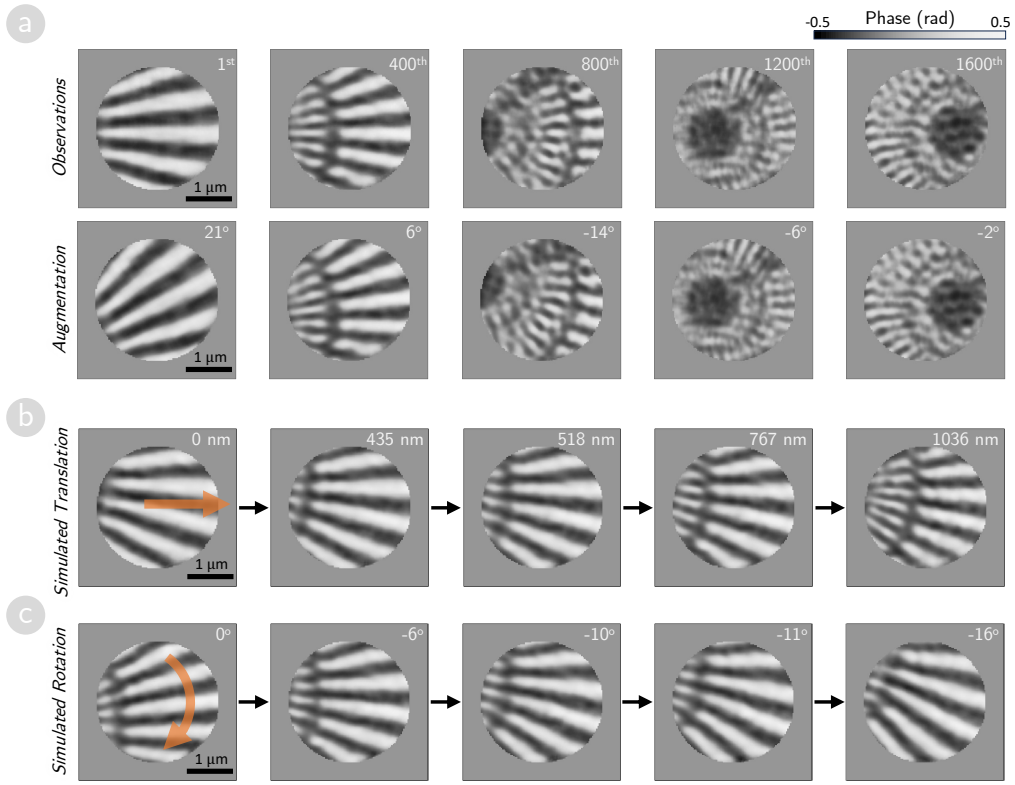

Figure 3.5: Sampled phase images depicting rigid-body motion progression of a Ta test chart. (**a**) Experimental CXDI phase images documenting controlled translation and rotation. Upper row: sequential images of translation (frames 1, 400, 800, 1200, 1600). Lower row: augmented in-plane rotations annotated by angle. (**b-c**) Representative GAN-generated sequences showing smooth progressions of translation (**b**) and rotation (**c**) extracted through MC sampling in latent space. Annotations indicate cumulative displacement or rotation relative to the initial frame, illustrating the model's ability to generate physically coherent transformation pathways.

assess geometric consistency. The measured slit widths were $d_1 = 309 \pm 11$ nm and $d_2 = 340 \pm 12$ nm, in close agreement with the experimental values $d_1 = 311 \pm 13$ nm and $d_2 = 324 \pm 11$ nm. The deviations correspond to less than two pixels (approximately 40 nm), confirming that the inferred variations remain geometrically consistent and free from deformation.

Collectively, these results verify that the generative model infers admissible changes—translation and rotation—consistent with the physical process while maintaining the deterministic constraint of rigidity. The explicit filtering ensured that the generated variations adhered to known physical

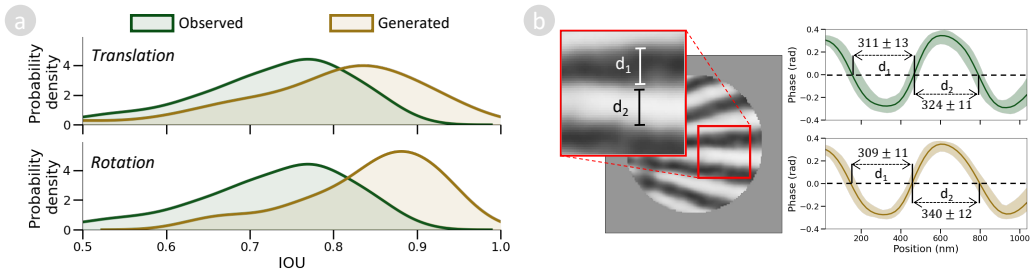

Figure 3.6: Quantitative validation of the Ta test chart case study. (**a**) Intersection-over-Union (IOU) distributions between consecutive frames for translation and rotation, comparing observed (green) and generated (yellow) sequences. (**b**) Comparison of slit widths d_1 and d_2 measured from observed and generated CXDI images. Solid lines represent mean widths, and shaded regions denote standard deviations, indicating preservation of geometric integrity.

laws, while the continuity and smoothness of transformations reflected the model's ability to infer constraint coherence within its latent geometry. This proof-of-concept thus establishes that the framework can reconstruct probabilistic changes under strict physical conditions, providing the rationale for extending the inference of constraints to more complex systems, where constraint structures are not predefined but emerge statistically through Monte Carlo sampling.

3.3 Case Study II: Nanoparticle Diffusion in Aqueous Media

Building upon the deterministic baseline established with the Ta test chart, this case study examines a system characterized by stochastic dynamics: the diffusion of gold nanoparticles (NPs) in an aqueous polyvinyl alcohol (PVA) matrix. In the context of object inquiry, the object of interest here is not a fixed configuration but a population of particle arrangements whose evolution is governed by thermally driven motion. The central question is whether the proposed framework can organize the observed variability of nanoparticle configurations into a coherent representational space and expose relations among these observations that articulate diffusive behavior.

In this setting, individual time-resolved images represent distinct realizations of particle arrangements sampled under similar experimental conditions. The generative representation learned from these observations embeds such configurations into a latent space whose geometry reflects statistical

similarity and variation among them. Local neighborhoods in this space correspond to configurations that differ by small collective rearrangements of particle positions, while broader structure reflects the range of spatial displacements accessible under diffusion. Relations among observations are therefore expressed through proximity, spread, and connectivity in latent space, rather than through deterministic trajectories.

The inquiry focuses on how dynamic behavior emerges from these relations. By examining ensembles of configurations generated through localized exploration of the latent space, the analysis characterizes diffusion as structured variability rather than as a single path of evolution. Quantities such as characteristic displacement scales and spatial correlation patterns arise from the distribution of generated configurations, while qualitative distinctions between diffusive regimes are reflected in how these distributions organize within representation space. In this way, the case study evaluates whether stochastic material dynamics can be articulated as relational patterns among observations, providing a representational description of diffusion grounded in the geometry of the learned latent space.

3.3.1 Representing Diffusion Configurations

The experimental dataset consists of 2,000 phase-reconstructed CXDI images capturing the real-time diffusion of gold NPs in a PVA matrix [33]. Each frame measures 142×142 pixels, corresponding to a spatial resolution of 40.58 nm per pixel. The sequence records evolving NP configurations governed by Brownian motion and medium heterogeneity, offering an ideal test for examining whether the model can infer physically admissible configurations and their stochastic transformations.

For this case study, we applied a preprocessing procedure adapted from prior work [33] to emphasize stochastic motion in CXDI phase images. The procedure involved (i) removing static components using a temporal window of 201 frames centered on each image and (ii) applying a Gaussian filter with a standard deviation of 1 pixel ($\approx$ 40.56 nm) to suppress spike noise. Subsequently, particle localizations were extracted using an Adaptive Thresholding method [95] with a neighborhood size of 11 pixels ($\approx$ 446 nm) and subtraction constant $C = 0.06$ radians. Pixels with positive local contrast were aggregated into connected regions, and only those with areas $\geq$ 11 pixels ($\approx 0.018\ \mu\text{m}^2$) were retained to match the expected NP size. These parameters follow those reported in [33], ensuring consistency with physical constraints and tracking fidelity. The schematic overview of this process is presented in Figure 3.7.

A convolutional GAN was employed with tailored architectural adaptations for this dataset (see more details in Appendix A.1). Specifically,

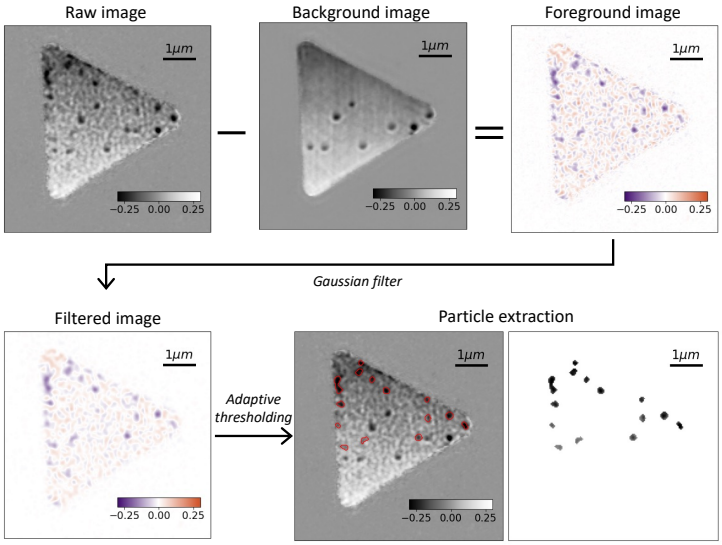

Figure 3.7: Extraction of diffusing nanoparticle regions from CXDI phase images. The preprocessing removes static background components and isolates local regions corresponding to moving particles.

the generator and discriminator each consisted of six convolutional layers in a progressive-growing schedule, starting from 4×4 latent feature maps and doubling spatial resolution at each stage until reaching the final image size of 128×128 pixels. The number of convolutional filters was gradually reduced as the resolution increased, allowing coarse layers to capture global spatial correlations among nanoparticles and finer layers to refine particle-level contrast. The schematic of this configuration is shown in Figure 3.8. Latent vectors were sampled from $\mathcal{N}(0, \mathbf{I})$ with dimensionality $d = 32$.

Representative experimental and generated CXDI images are shown in Figure 3.9a, with experimental frames (left) sampled at the 600^{th}, 900^{th}, and 1200^{th} time steps and corresponding synthetic images (right) selected from 1,000 GAN-generated samples decoded from latent vectors drawn i.i.d. from $\mathcal{N}(0, \mathbf{I})$. Using the validated image-processing workflow of [33], we extracted NP positions and morphologies in both sets and compared NP area distributions (Figure 3.9b), obtaining a low Wasserstein distance of 0.068. This close statistical agreement indicates that the learned representation does not merely reproduce isolated static configurations; it organizes the observed variability of nanoparticle coordination characteristic of diffusive motion. Within the object-inquiry framework, this fidelity demonstrates that relations among observed configurations are coherently embedded in representation space, establishing a reliable basis for articulating dynamic

behavior from their statistical organization.

To examine whether nearby latent codes correspond to physically coherent nanoparticle rearrangements, we conducted a locality study in latent space. We sampled 1,000 reference vectors from $\mathcal{N}(0, \mathbf{I})$ and, around each reference, generated 100 variants by adding $\epsilon_j \sim \mathcal{U}_{[-\gamma,\gamma]}$ with $\gamma \in \{0.25, 0.50, \ldots, 2.00\}$. Because absolute distances in high-dimensional spaces can be misleading [96, 97], locality was quantified by the average Euclidean distance between a reference vector and its perturbed variants, normalized by the latent dimensionality (details in Appendix A.2). Under this normalization, $\gamma = 0.25$ and $\gamma = 2.0$ correspond to average per-dimension changes of approximately 0.025 and 0.20, respectively, providing an interpretable scale for probing progressively larger neighborhoods of the learned representation space.

The effect of these perturbations on generated images was quantified using two measures tailored to diffusion-driven variability: (1) the image-entropy difference ΔH [98], reflecting configurational complexity, and (2) the NP-area difference ΔA, indicating the magnitude of structural rearrangement. As shown in Figure 3.9c, both ΔH and ΔA increase smoothly with γ, evidencing a graded and coherent mapping between neighborhoods in latent space and variations in physical configuration space. Small perturbations ($\gamma \leq 0.5$) yield sub-pixel–level adjustments, whereas larger perturbations ($\gamma \geq 1.5$) induce broader reorganization of NP clusters.

Notably, $\gamma = 1$ reproduces frame-to-frame experimental variability, with $\Delta H \approx 0.26$ and $\Delta A \leq 200$ pixels, consistent with the observed depletion rate of approximately 10 particles s^{-1} in this system [33]. This correspondence indicates that local neighborhoods in representation space align quantitatively with relations observed between successive experimental frames. In terms of object inquiry, the geometry of the latent space encodes relations among observations in a manner that allows diffusive behavior to be articulated as structured variability rather than as deterministic trajectories. We therefore adopt $\gamma = 1$ as the standard sampling margin in subsequent Monte Carlo analyses. Collectively, these results demonstrate that the learned representation supports smooth, behavior-consistent variation aligned with experimental dynamics, providing a coherent representational basis for examining stochastic material behavior in more complex systems.

3.3.2 Extracting Diffusive Behaviors

Monte Carlo (MC) sampling was performed around 200 reference latent vectors using a sampling margin of $\gamma = 1$, identified in the previous analysis as corresponding to realistic diffusive variability observed between consecutive experimental frames. Around each reference state, 500 perturbed latent

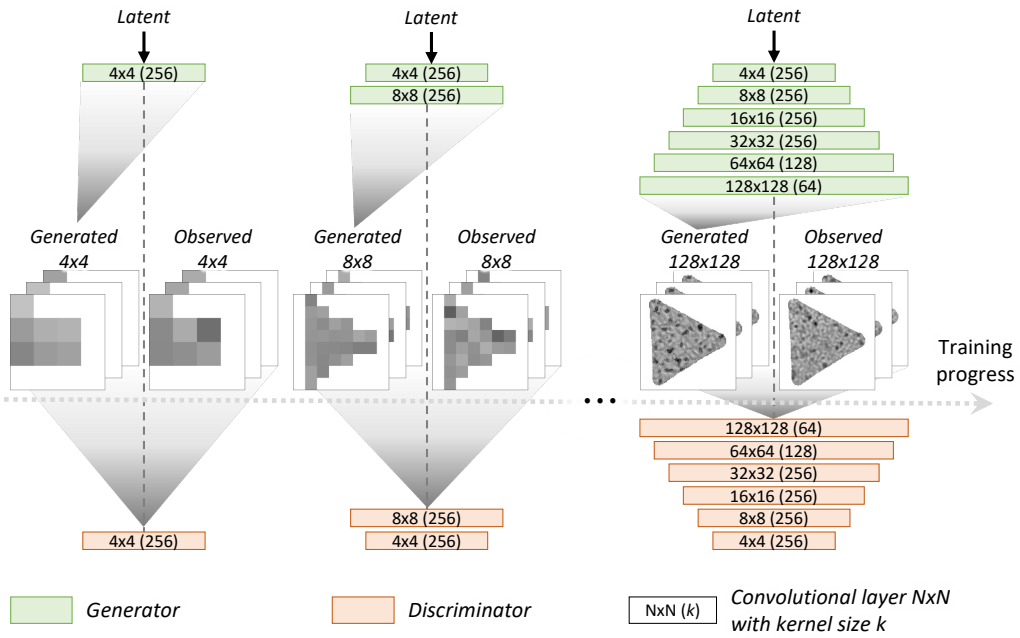

Figure 3.8: Schematic diagram illustrating the GAN architecture used to synthesize CXDI phase images of Au–NP diffusion. The generator and discriminator each consist of six convolutional stages, progressively doubling image resolution from 4×4 to 128×128 pixels while refining structural detail.

vectors were generated and decoded into synthetic configurations. To retain configurations consistent with experimentally observed variability, we applied an entropy-difference threshold of $\Delta H \leq 0.2$, which lies within the upper bound of 0.26 measured between successive CXDI frames. The retained samples were then pixel-wise averaged to produce images summarizing cumulative nanoparticle (NP) diffusion patterns over approximately one-second intervals. Each averaged image therefore represents a statistical condensation of locally sampled configuration ensembles, capturing how variations around an observed state are organized within the learned representation space.

From these averaged ensembles, we extracted individual NP diffusion trajectories using feature-based tracking [33]. Each trajectory corresponds to a locally inferred diffusive event and was characterized by two geometric descriptors: anisotropy and tortuosity [99,100]. Anisotropy quantifies the directional persistence of NP motion, while tortuosity measures the irregularity of diffusion paths. These descriptors translate the sampled structural variations into physically interpretable indicators of motion behavior [101, 102]. The distribution of diffusion areas derived from all extracted trajectories

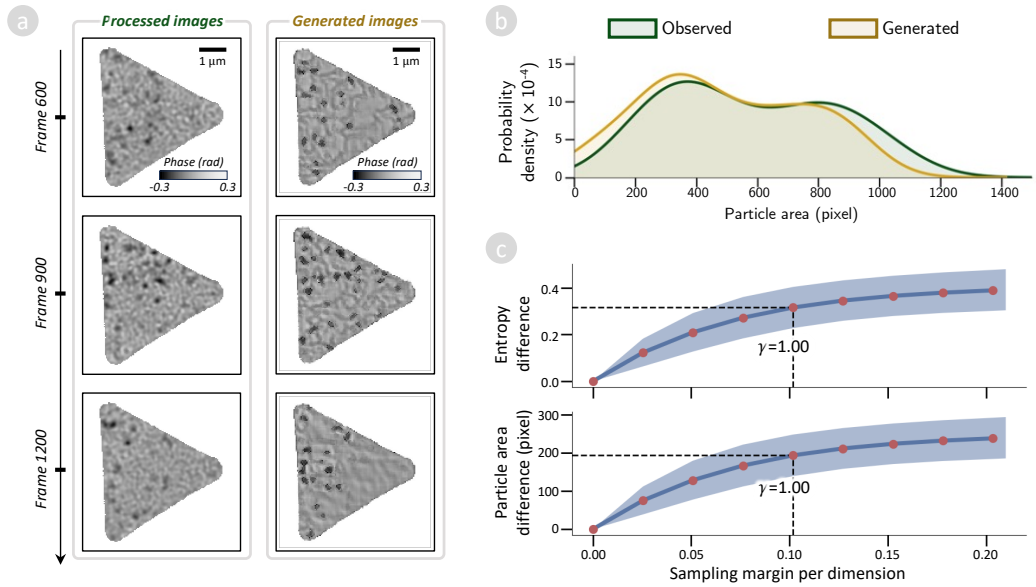

Figure 3.9: Evaluation of fidelity for generated NP configurations and diffusion dynamics. (**a**) Experimental and synthetic CXDI phase images showing representative NP dispersion. (**b**) Comparison of NP area distributions (Wasserstein distance = 0.068). (**c**) Entropy and NP area differences as functions of perturbation amplitude γ; $\gamma = 1$ reproduces experimental variability.

(Figure 3.10b) yields a mean of $7,712 \pm 2,862$ nm^2, closely matching the experimentally determined diffusion rate of $7,550 \pm 850$ nm^2 s^{-1} measured by X-ray Photon Correlation Spectroscopy (XPCS) [33]. This quantitative correspondence confirms that the inferred constraint magnitudes—defined by the rate and spatial extent of admissible changes—faithfully reproduce the known physical diffusion limits of the system.

To investigate the emergence of higher-order organization within these inferred constraints, we assembled a distance matrix from the standardized anisotropy and tortuosity descriptors and applied hierarchical agglomerative clustering (HAC) with average linkage [9, 103]. This unsupervised analysis groups trajectories according to geometric proximity in descriptor space, thereby revealing clusters of diffusive behavior as distinct *constraint regimes*. The dendrogram shown in Figure 3.10c identifies three major diffusion modes. The first exhibits directionally biased diffusion with low tortuosity, reflecting partially guided or channel-constrained motion possibly influenced by local polymer structure or field gradients [104]. The second cluster represents isotropic diffusion with high tortuosity and low anisotropy, consistent with

classic Brownian motion in a homogeneous medium [101]. The third cluster corresponds to confined diffusion characterized by minimal displacement and high curvature, indicating transient trapping or immobilization within local potential wells or heterogeneous polymer domains [102, 105]. The distributions of anisotropy and tortuosity within these clusters (Figure 3.10d) illustrate how the generative manifold organizes stochastic transformations into distinct behavioral patterns governed by physically meaningful constraints.

Overall, this analysis demonstrates how stochastic nanoparticle diffusion can be articulated as dynamic behavior emerging from relations among observed and generated configurations. Rather than relying on individual trajectories or deterministic descriptions, the framework organizes ensembles of configurations into structured patterns that reflect both known physical diffusion characteristics and higher-level behavioral organization. In this way, the generative modeling approach supports object inquiry by transforming observational variability into interpretable relational structure, providing a systematic basis for examining stochastic material dynamics in complex environments.

3.4 Case Study III: Sulfidation in Rubber–Brass Composites

Having established the generative modeling framework in deterministic and stochastic settings, this case study extends its application to a chemically evolving, volumetric process: the sulfidation of copper in aging rubber–brass composites. This process underpins durability and adhesion in steel-cord–reinforced rubber tires, where brass coatings on steel wires chemically interact with sulfur-containing rubber matrices. Despite its industrial relevance, the microscopic organization of copper redistribution during aging remains only partially characterized, with existing studies relying largely on empirical observation.

Within the object-inquiry framework, the object of interest is a chemically active material volume whose observable configurations evolve through coupled structural and compositional variation. Time-resolved X-ray absorption fine structure computed tomography (XAFS-CT) provides discrete volumetric observations of copper valence states (Cu, Cu_2S, CuS), capturing successive stages of chemical transformation within individual brass clumps. This case study examines whether the generative modeling framework can organize these observations into a coherent representation space and expose relations among configurations that articulate chemically driven dynamic

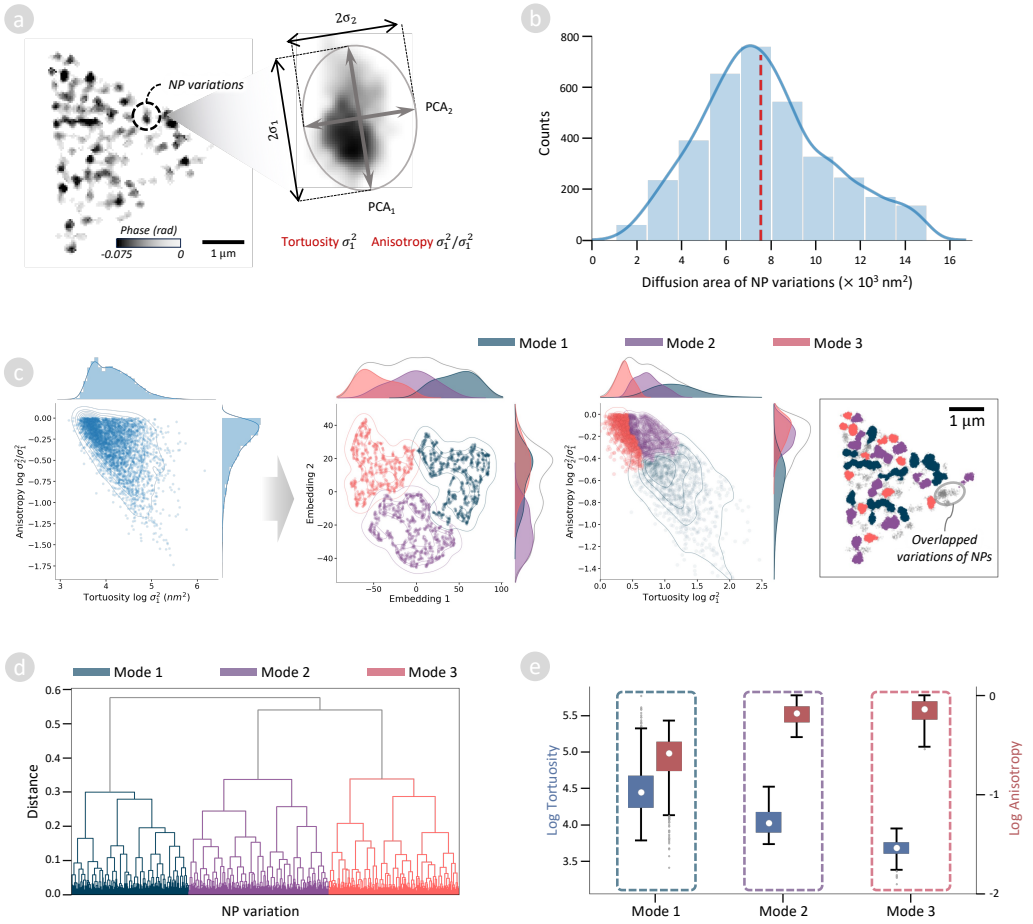

Figure 3.10: Analysis of NP diffusion in PVA solution. (**a**) Definition of motion descriptors (anisotropy, tortuosity). (**b**) Distribution of diffusion areas inferred from generative ensembles. (**c**) HAC dendrogram showing three distinct diffusion constraint regimes. (**d**) Box plots of anisotropy and tortuosity for each regime.

behavior.

In this setting, variation in the learned representation reflects spatial redistribution and compositional change of copper species during oxidation and sulfidation. Relations among configurations—expressed through continuity and organization in latent space—characterize how chemical evolution proceeds while preserving morphological coherence. By organizing observed and generated configurations into structured patterns of variation, the framework provides a representational description of sulfidation dynamics grounded in relations among volumetric observations.

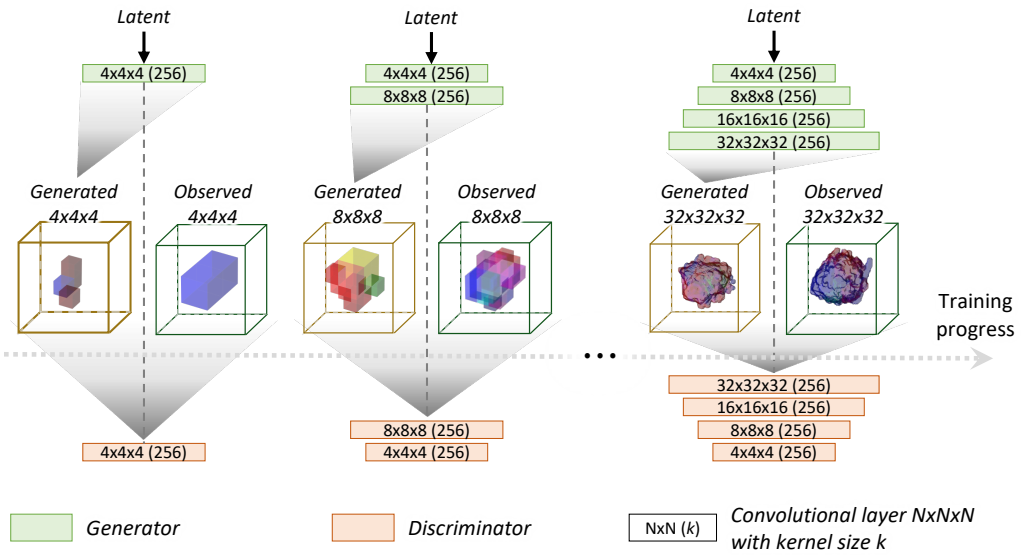

Figure 3.11: Schematic diagram illustrating the GAN architecture used to synthesize 3D XAFS-CT images of brass clumps under sulfidation. The generator and discriminator each consist of six convolutional stages, progressively doubling image resolution from $4 \times 4 \times 4$ to $32 \times 32 \times 32$ voxels while refining structural detail.

3.4.1 Representing Sulfidation Clumps

Three-dimensional datasets were acquired using X-ray Absorption Fine Structure Computed Tomography (XAFS-CT) [23] at the SPring-8 synchrotron facility [34]. This technique provides volumetric elemental mapping at sub-micrometer resolution, distinguishing copper oxidation states (Cu, Cu_2S, and CuS) through spectral contrast. Approximately 2,000 brass particles (Cu/Zn = 75/25) were embedded in a sulfur-containing rubber matrix, heated to 443 K for 10 min to form an adhesive interlayer, and subsequently aged under controlled environmental conditions for 0, 3, 14, and 28 days. These intervals capture the progression from initial surface sulfidation to deep internal oxidation within the brass–rubber composite.

To prepare the XAFS-CT data for GAN training, we employed a multi-step clump extraction and filtering pipeline adapted from our previous study [23]. After identifying distinct brass clumps via HDBSCAN clustering [106] on upsampled spatial coordinates, cubic subvolumes of $32 \times 32 \times 32$ voxels centered on each clump centroid were extracted. To ensure that the selected clumps were both spatially confined and morphologically meaningful, three filtering criteria were applied before cropping: (i) clumps exceeding 30

voxels in maximum extent along any axis, (ii) those smaller than 3 voxels in minimum extent, or (iii) having total volume below 1,000 voxels were excluded. These thresholds removed oversized regions likely to exceed the fixed input size, undersized or fragmented detections, and low-volume noise artifacts. Only clumps satisfying all criteria were retained, normalized into the 32^3 voxel format, and organized as multi-channel volumes representing the relative densities of Cu, Cu_2S, and CuS.

A 3D convolutional GAN was then trained following the progressive-growing strategy described in Section 3.1.2.1, adapted for multi-channel volumetric data. The model consisted of eight convolutional layers, with feature-map resolution doubling at each stage until the final 32^3 voxel resolution was reached. The number of filters was gradually reduced with increasing resolution, enabling coarse layers to capture global morphology and fine layers to refine local compositional gradients. Progressive up-scaling stabilized adversarial training, while adaptive batch-size reduction maintained gradient consistency at higher resolutions. The overall model architecture and progressive training schedule are illustrated in Figure 3.11 (see more details in Appendix A.1). All latent vectors were sampled from a standard normal distribution $\mathcal{N}(0, I)$ with latent dimensionality $d = 32$.

Representative experimental and GAN-generated volumes at matched aging intervals are shown in Figure 3.12a–b. The generated reconstructions reproduce the morphological and compositional heterogeneity observed experimentally, capturing both the Cu-rich matrix and Cu_2S and CuS subdomains. Quantitatively, voxel-intensity distributions of Cu, Cu_2S, and CuS exhibit close statistical agreement between experimental and generated data (Wasserstein distances: Cu = 0.010, Cu_2S = 0.017, CuS = 0.006; Figure 3.12c). This correspondence verifies that the learned generative manifold not only reproduces static compositional states but also embeds the statistical field of local variability that governs feasible transformations. In this sense, the manifold captures the underlying structure of *inferred changes*—the physically admissible redistributions of mass and morphology associated with sulfidation.

To evaluate the coherence and continuity of these inferred changes, we analyzed how local perturbations in latent space affect the generated volumes. A total of 1,000 latent vectors were drawn from $\mathcal{N}(0, I)$, each perturbed 100 times by adding uniform noise $\epsilon \sim \mathcal{U}_{[-\gamma,\gamma]}$ with γ ranging from 0.25 to 2.0 in increments of 0.25. Interpreting distances in high-dimensional latent spaces is nontrivial because Euclidean magnitudes lose direct physical meaning [96,97]; thus, we normalized the average Euclidean distance between each reference and perturbed latent vector by the latent dimensionality to provide an intuitive measure of locality (see more details in Appendix A.2).

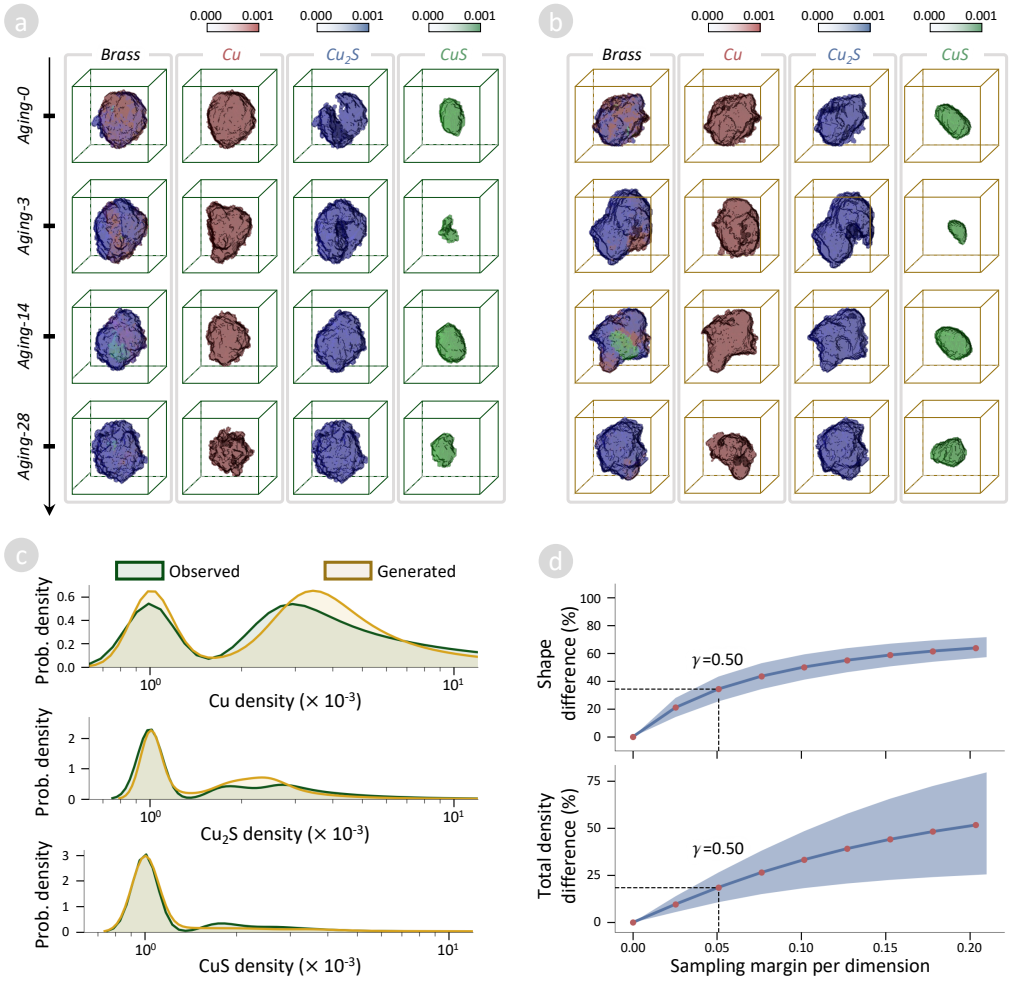

Figure 3.12: Evaluation of fidelity for generated brass clumps and transitions. (**a**) Experimental 3D XAFS-CT images showing brass clump evolution during aging (0-28 days), with separated visualizations of brass matrix and copper species (Cu, Cu_2S, CuS). (**b**) GAN-generated images representing plausible sulfidation states at corresponding aging stages. (**c**) Density distributions of copper species comparing observed (solid lines) and generated (shaded areas) images, demonstrating close statistical agreement. (**d**) Latent space continuity analysis: shape differences (%) and total density differences (%) as functions of sampling margin per dimension (γ normalized by latent space dimensionality for intuitive evaluation of perturbation magnitude), with γ=0.5 selected for subsequent dynamic analysis.

This normalization implies that perturbation margins of $\gamma = 0.25$ and $\gamma = 2.0$

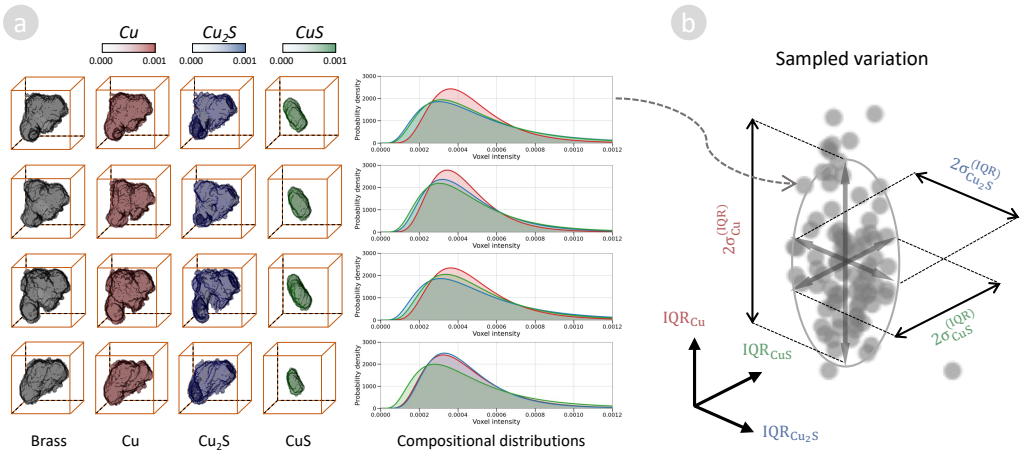

Figure 3.13: Characterizing clump variations in brass compositions of clumps depicted in generated 3D XAFS-CT images. **(a)** Sampled states of brass clumps and their constituent compositions by using a GAN model. (**b**) The schematic diagram of extracting variation descriptor through the Interquartile Range (IQR) derived from the compositional density distributions.

correspond to average per-dimension changes of approximately 0.025 and 0.20, respectively.

Each perturbed latent vector was decoded into a 3D multi-channel volume representing a hypothetical sulfidation configuration. Two quantitative measures were used to assess smoothness and physical plausibility: (1) the complement of the three-dimensional intersection-over-union (IOU_{3D}) to quantify morphological continuity, and (2) the total compositional density difference to quantify species redistribution. As shown in Figure 3.12d, both metrics increase smoothly with larger γ, indicating that the latent space encodes a continuous spectrum of transformation magnitudes. Small perturbations ($\gamma \leq 0.25$) result in near-identical reconstructions, corresponding to sub-voxel modifications in composition, whereas large perturbations ($\gamma \geq 1.5$) produce notable compositional rearrangements and morphological deviations. At $\gamma = 0.5$, the average IOU_{3D} remains around 0.7 and the compositional deviation below 20%, values consistent with the experimentally observed differences between consecutive aging intervals [23]. This sampling margin thus delineates the scale of physically meaningful local transitions and is adopted as the standard neighborhood for subsequent Monte Carlo sampling to further explore the manifold's representation of sulfidation dynamics.

3.4.2 Extracting Sulfidation Behaviors

Monte Carlo (MC) sampling was conducted using $\gamma = 0.5$ (corresponding to an average per-dimension distance of 0.05) to explore local neighborhoods within the learned latent space. From 3,000 reference latent vectors, 400 perturbed samples were generated per reference and decoded into synthetic volumetric configurations. To ensure consistency with experimentally observed variation, a two-stage filtering procedure was applied. First, the bounded sampling radius restricted exploration to locally coherent regions of the representation space. Second, samples were screened using compositional mass deviation ($\leq 20\%$) and three-dimensional intersection-over-union ($\mathrm{IOU_{3D}} \geq 0.5$) criteria, enforcing morphological continuity and compositional consistency with observed brass clumps. The retained ensembles—referred to here as *clump variations*—represent localized neighborhoods of variation around observed configurations, reflecting how chemically coupled transformations are organized within the learned representation.

From each clump variation, voxel-based statistical descriptors were computed to quantify patterns of compositional redistribution among copper species. For each component $c \in \{\mathrm{Cu}, \mathrm{Cu_2S}, \mathrm{CuS}\}$, the interquartile range (IQR) of voxel intensities was calculated across all retained samples, and the variance of these IQR values was used as a measure of redistribution magnitude. Each clump variation was therefore summarized by a descriptor vector

$$\mathbf{d} = \left[\mathrm{Var}(\mathrm{IQR}_{\mathrm{Cu}}), \mathrm{Var}(\mathrm{IQR}_{\mathrm{Cu_2S}}), \mathrm{Var}(\mathrm{IQR}_{\mathrm{CuS}})\right], \tag{3.10}$$

which encodes how compositional variation is distributed within the local neighborhood of an observed configuration. Collectively, these descriptors define the geometric organization of variation across the representation space.

To visualize relationships among clump variations, we applied multidimensional scaling (MDS) to the descriptor vectors, yielding a two-dimensional embedding that reflects pairwise similarity in sulfidation behavior (Figure 3.14a). Three well-separated clusters, labeled A, B, and C, emerged, each representing a distinct constraint regime associated with a characteristic mode of sulfidation. Temporal mapping of these clusters onto the experimental aging intervals (Figure 3.14b) revealed a systematic evolution: Group A dominated at the initial stage (Aging-0), Group B appeared primarily at intermediate aging (3–14 days), and Group C became prevalent at advanced aging (14–28 days). This chronological trend mirrors the experimentally verified sequence of chemical conversion from metallic Cu to $\mathrm{Cu_2S}$ and subsequently to CuS [23]. Thus, the manifold's organization of transformation descriptors inherently reconstructs the temporal order of sulfidation, implying that these cluster boundaries reflect *inferred constraints*

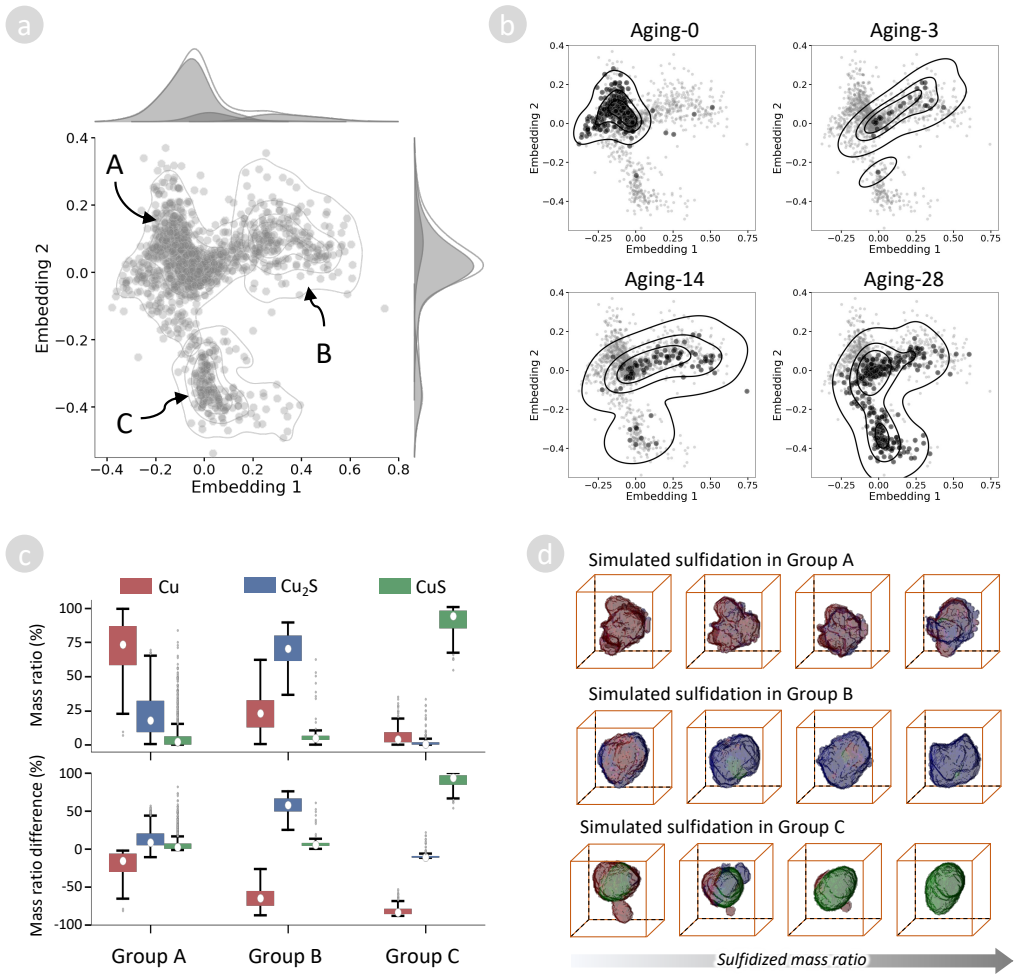

Figure 3.14: Analysis of sulfidation dynamics in aging rubber/brass composites. (**a**) Two-dimensional embedding space obtained via multidimensional scaling (MDS) of clump transformation descriptors, revealing three distinct sulfidation modes, denoted as groups A, B, and C. (**b**) Temporal mapping showing progression of sulfidation modes across aging stages (0-28 days), with contour lines indicating density distributions. For each stage, darker points denote variations associated with that stage, while lighter ones indicate variations involved with other stages. (**c**) Box plots summarizing the differences in compositional mass ratios (Cu, Cu_2S, CuS) across the identified sulfidation modes (the gray dots indicate outliers of the respective attributes). (**d**) Representative 3D visualizations of sulfidation progression for each mode, showing transformation from left to right with increasing sulfidized mass ratio.

consistent with known reaction pathways.

Compositional mass ratios of Cu, Cu_2S, and CuS within each cluster (Figure 3.14c) further clarify the nature of these behavioral patterns. Group A retains a high fraction of metallic Cu with minimal sulfidation products, corresponding to surface-limited chemical activity at early aging stages. Group B exhibits enrichment of Cu_2S, indicative of intermediate sulfidation concentrated near interfaces. Group C is dominated by CuS, reflecting extensive sulfidation and bulk oxidation associated with prolonged aging. These transitions describe an ordered progression of chemical behavior, from surface-localized reaction to volume-spanning transformation.

Representative transformations simulated from each group (Figure 3.14d) visualize these inferred constraint regimes. Clump variations in Groups A and B display Cu_2S accumulation localized near particle surfaces, consistent with surface sulfidation and interfacial bonding during vulcanization. In contrast, Group C exhibits CuS formation extending inward, reflecting the internal oxidation typical of long-term degradation. These generated transformations correspond well to experimental observations of chemical and structural aging, including the shift from adhesive interlayer formation to bulk CuS development associated with mechanical fatigue, cracking, and delamination [23, 107, 108].

Overall, this analysis demonstrates that the generative modeling framework does more than reproduce short-time configurational variability in nanoparticle coordination; it provides a structured representational space in which dynamic behavior can be articulated from relations among observed and generated configurations. Quantitative characteristics such as diffusion rate and displacement range emerge naturally from the statistical organization of sampled ensembles, while higher-level behavioral modes arise from relational structure revealed through geometric descriptors and clustering. In this way, deep generative representations act as effective mediators between discrete experimental observations and continuous descriptions of material behavior, without requiring explicit specification of governing mechanisms. Viewed through the lens of the object-inquiry framework, the learned latent space serves as a representation that organizes observations according to similarity and variation, while Monte Carlo sampling exposes relations among neighboring configurations. Dynamic behavior is then articulated by interpreting these relations—through continuity, dispersion, and clustering—rather than by tracing deterministic trajectories or isolating individual events. This correspondence highlights the significance of deep generative modeling not merely as a data-driven tool, but as a computational realization of object inquiry, transforming sparse observations into organized representations from which material behavior can be systematically derived.

3.5 Discussion

This chapter developed and examined a generative modeling framework for studying material dynamics from discretely sampled experimental observations. Imaging techniques such as CXDI and XAFS-CT provide increasingly detailed snapshots of material structure across spatial and chemical dimensions, yet they inevitably sample evolving processes at finite temporal and configurational resolution. Each recorded frame or tomographic reconstruction represents an observable configuration conditioned by experimental settings, while the transformations connecting successive observations remain unmeasured. As a result, material dynamics cannot be accessed directly as continuous trajectories, but must be approached through the relations that organize discrete observations. The framework presented here addresses this structural sparsity by constructing representations that organize observed configurations into continuous domains of variation, allowing dynamic behavior to be articulated from relational structure rather than from explicit temporal reconstruction.

Within the object-inquiry framework, deep generative modeling serves as the mechanism for the first step of inquiry: organizing observations into a coherent representation space. By training generative models on experimentally observed configurations, the learned latent space captures how material states are distributed under the combined influence of physical processes and measurement conditions. Proximity, continuity, and clustering within this space encode relations among observations that are not directly apparent in the raw data. Monte Carlo sampling then operationalizes the second step of inquiry by systematically exploring these relations, transforming the implicit organization of representation space into explicit ensembles of configurations that can be analyzed statistically. Dynamic behavior is thus articulated through patterns of variation, continuity, and organization revealed by relations among observed and generated configurations.

Across the three case studies, this representational strategy demonstrated consistent capability to articulate distinct forms of material behavior. In the tantalum test chart, rigid-body motion was expressed as smooth, low-dimensional variation in representation space, reflecting deterministic geometric behavior. In the nanoparticle diffusion system, stochastic motion emerged as structured variability organized into relational patterns that correspond to diffusive behavior, including isotropic, biased, and confined regimes. In the rubber–brass composite, chemically coupled sulfidation was articulated as coordinated spatial and compositional variation across volumetric observations, revealing ordered patterns of chemical aging. Despite the increasing complexity of physical processes involved, all three systems

were described within a unified representational framework that organizes observations and derives behavior from their relations.

These results illustrate that deep generative representations can function as scale-agnostic organizational spaces for dynamic phenomena. Rather than reconstructing missing trajectories or specifying governing mechanisms explicitly, the framework leverages the statistical organization of observed data to expose how material configurations relate and vary. In doing so, it mirrors long-standing practices of scientific inquiry, in which behavior is inferred from patterns and relations among observations, but implements them computationally through learned representations. The latent space does not replace physical description; instead, it provides a structured domain in which observational variability is organized in a manner amenable to systematic analysis.

The current implementation nevertheless exhibits important limitations. Because the representation is learned from finite experimental datasets, the scope of articulated behavior is bounded by the diversity of observed configurations. Patterns absent from the data cannot be expressed in representation space. In addition, Monte Carlo sampling treats local variations as memoryless, limiting the framework's ability to represent directional progression, path dependence, or hysteresis. These limitations reflect not conceptual shortcomings of the object-inquiry framework, but practical constraints of the present representational realization.

Future extensions could address these limitations by incorporating temporal conditioning, structured priors, or physics-informed architectures that enrich relational organization without departing from the inquiry logic. Time-conditioned generative models, latent dynamical systems, or diffusion-based architectures could allow ordered progression to be represented explicitly, while maintaining the emphasis on relations among configurations. Such developments would further strengthen the alignment between computational representation and scientific inquiry, enabling richer articulation of dynamic behavior from sparse observations.

More broadly, this chapter demonstrates that deep generative modeling can be understood not merely as a data-driven technique, but as a computational realization of object inquiry under incomplete observation. By transforming discrete experimental measurements into organized representational spaces and articulating behavior from relations within those spaces, the framework converts observational sparsity from a limitation into a resource. In this sense, generative representations provide a principled bridge between experimental observation and systematic understanding of material dynamics, grounded in the same logical structure that underlies scientific inquiry across domains.

Chapter 4

Attention-based Inquiry Framework of Material Dynamics

This chapter develops and applies an attention-based object-inquiry framework that forms the second methodological pillar of this thesis. Whereas the generative modeling approach in the previous chapter focused on organizing sparse experimental observations into representational spaces that articulate dynamic behavior through structured variation, the present chapter addresses a complementary situation frequently encountered in experimental practice. In many material systems, experimental data provide dense sequences of images and measurements across time and space, yet the structural variations responsible for property variation remain difficult to interpret within complex spatiotemporal patterns. The challenge in such settings is not the absence of data, but the difficulty of exposing and organizing relations within the observations that are relevant to the behavior of interest.

The framework introduced here is not merely an application of a pre-existing "second step" of inquiry; it is a concrete integration of attention-based transformer modeling into the object-inquiry process. In this integration, attention provides an explicit, token-level relational mechanism that links parts of an observation (e.g., spatial regions, temporal segments, or modality-specific channels) to one another and to measured responses. The resulting attention structure serves as a representation of relevance relations: it makes visible how observational components are selectively associated with behavior and how these associations persist, shift, or reorganize as the system evolves. In this way, the attention-based framework operationalizes object inquiry under dense observation by constructing relations that are directly inspectable and interpretable back in observation space.

Attention-based deep learning provides the operational mechanism for this integration. When trained to associate time-resolved structural observations with experimentally measured properties, spatiotemporal attention models learn weighted associations among data tokens conditioned on the response. Within this framework, attention weights are treated as relational indicators that expose which components of observation contribute

most strongly to the measured behavior under specific conditions and time intervals. By examining the spatial distribution, temporal persistence, and orientation content of attention patterns, the framework converts model-derived relations into descriptors of structured structural relevance. Rather than serving solely as a predictive device, attention thus functions as a relational tool that supports behavior articulation by making response-linked relations among observations explicit.

The chapter is organized to develop this framework from conceptual formulation to practical application. The opening section introduces the motivation for integrating attention into object inquiry, clarifying how relevance-based relations differ from similarity- or variation-based organization. The methodological section then presents the construction of spatiotemporal attention models, the extraction of attention-weighted representations, and the analytical procedures used to characterize their organization. These methods are applied to two contrasting case studies: deformation of gold nanocontacts, a nanoscale system with strong crystallographic regularity, and wet rubber–surface contact under sliding, a macroscale system governed by multiple interacting mechanisms without a fixed geometric reference. Through these examples, the chapter demonstrates how attention-based representations expose response-linked relations within dense observations and support systematic articulation of behavior across disparate physical regimes. The chapter concludes with a discussion that situates attention-based inquiry as complementary to the generative modeling approach, emphasizing how different representational mechanisms realize object inquiry under different observational regimes.

4.1 Framework Overview

This section introduces the methodological foundation of the attention-based object-inquiry framework developed to analyze fully observed yet structurally complex dynamical data. In many time-resolved experiments, imaging and property measurements are synchronized with sufficient temporal resolution to capture complete sequences of configurations. Nevertheless, the structural variations that drive the measured response often remain difficult to interpret. Property changes typically arise from localized and heterogeneous structural processes—such as deformation, slip activation, drainage patterning, or contact reorganization—that are not explicitly labeled in the data and frequently lack simple geometric regularity. As a result, the central challenge lies not in recovering missing configurations, but in exposing which aspects of the observed structural evolution are relevant to the behavior of interest and how these aspects are organized across space

and time.

The framework formulated here addresses this challenge by integrating spatiotemporal attention mechanisms into the object-inquiry process. When trained to associate sequences of structural observations with measured properties, an attention-based deep model learns weighted relations among spatial regions and temporal segments of the data. These attention weights are interpreted as relevance relations: they indicate how strongly different parts of the observed structure contribute to the modeled response under specific conditions. By examining the distribution, persistence, and organization of attention across space and time, the framework identifies structurally relevant variations within the evolving observations and characterizes how these variations are organized into coherent patterns linked to material behavior.

This formulation establishes the basis for the methodological developments that follow. The subsequent subsections describe how spatiotemporal attention is constructed within the model, how attention-weighted representations are formed in observation space, and how their organization is quantified to articulate behavior from relevance relations. Together, these components define the attention-based object-inquiry framework applied in the case studies presented later in this chapter.

4.1.1 Attention as Correlative Relations of Observations

Time-resolved experimental techniques increasingly provide datasets in which structural observations and material properties are synchronously recorded across space and time. Examples include *in situ* microscopy of nanoscale deformation accompanied by electrical or mechanical measurements, as well as macroscopic imaging of interfacial contact coupled to frictional or tribological response [26, 54, 55, 109]. In such settings, experimental observations are often complete in the sense that full sequences of images and measurements are available. Nevertheless, the structural origins of the measured response frequently remain difficult to interpret, as property variations are embedded within complex, high-dimensional spatiotemporal patterns :contentReference[oaicite:0]index=0.

From the perspective of object inquiry, this difficulty reflects a representational rather than an observational limitation. The object of interest is a material system whose behavior emerges from the coordinated evolution of multiple localized structural processes. Individual observations encode heterogeneous and overlapping phenomena—such as deformation, slip activation, drainage, or contact reorganization—distributed across space and time.

Measured properties, by contrast, represent aggregated outcomes of these processes rather than direct reflections of any single structural feature. As a result, inquiry cannot proceed by treating each observation as an indivisible entity; it requires representations that preserve relations between granular structural variations and the global response they collectively produce.

Conventional deep learning models trained to map entire image sequences directly to scalar properties can achieve high predictive accuracy [62–65]. This success indicates that such models internalize complex correlations across space and time. However, when these correlations remain embedded implicitly within latent activations, predictive performance alone does not clarify which components of the observation are relevant to the behavior of interest. The challenge is therefore not to increase model capacity, but to construct representations that expose how variability at finer structural scales contributes selectively and unevenly to the observed property.

Attention mechanisms provide a principled means to meet this representational requirement. In attention-based architectures [49], observations are decomposed into tokens corresponding to spatial regions, temporal segments, or feature channels, and relations among these tokens are encoded through attention weights. Rather than summarizing variability across an entire observation, attention enables preferential association: it emphasizes those components of the data whose variations are most strongly coupled to the measured response. In this sense, attention functions as a relational mechanism that preserves correlations between localized structural variation and aggregated material behavior.

Formally, attention weights quantify correlation-like associations between encoded elements of an observation conditioned on the modeling objective. For elements i and j, the attention coefficient α_{ij} is computed as

$$\alpha_{ij} = \mathrm{softmax}\left(\frac{q_i^\top k_j}{\sqrt{d_k}}\right), \tag{4.1}$$

where q_i and k_j are learned query and key representations, and d_k denotes their dimensionality. The inner product $q_i^\top k_j$ acts as a similarity or correlation score, measuring the degree to which variations in element j are statistically aligned with the contextual representation associated with element i. The subsequent normalization enforces a relative weighting, distributing attention across all elements such that $\sum_j \alpha_{ij} = 1$.

Within this formulation, attention does not represent absolute importance of individual components, but relative correlation structure among decomposed parts of the observation. Elements whose variations are consistently correlated with the response context receive higher attention weights, while

components exhibiting weak or inconsistent correlation are downweighted. In time-resolved experimental data, this correlation structure reflects how localized structural variations co-vary with measured properties across space and time.

When trained on synchronized image–measurement datasets, the resulting attention field therefore encodes a learned correlation map between granular structural components and aggregated material response. Importantly, these correlations are not imposed a priori; they emerge from repeated co-occurrence patterns present in the data. Attention thus provides a mechanism for preserving and exposing correlation relationships that are otherwise absorbed implicitly within high-dimensional latent representations.

From the perspective of object inquiry, this correlation-based view of attention clarifies its representational role. By decomposing observations into interacting components and encoding their correlation structure explicitly, attention-based representations make it possible to articulate dynamic behavior from how structural variability aligns with measured response. In this sense, attention functions as a preferential correlation operator, organizing dense observations into relational structures that reflect how material behavior emerges from the coordinated evolution of its constituent parts.

4.1.2 Implementation of the Attentive Inference Framework

The attention-based inquiry framework implements a two-stage analytical process that realizes object inquiry under conditions of dense but structurally complex observation. In contrast to the generative modeling approach, which emphasizes organizing variation among sparsely sampled configurations, the present framework is designed for settings in which time-resolved observations are complete, yet the relations that connect localized structural variation to measured behavior remain obscured. The methodological objective is therefore not to reconstruct missing states, but to construct representations that expose how distributed components of the observation contribute, in an uneven and organized manner, to the observed response.

The first stage of the framework corresponds to the representational stage of object inquiry. A spatiotemporal attention model is trained to associate sequences of structural observations with measured properties, treating images and measurements as coupled projections of the same evolving object. Through this training process, the model learns relevance relations that weight spatial regions, temporal segments, or feature channels according to how strongly their variations correlate with the measured response.

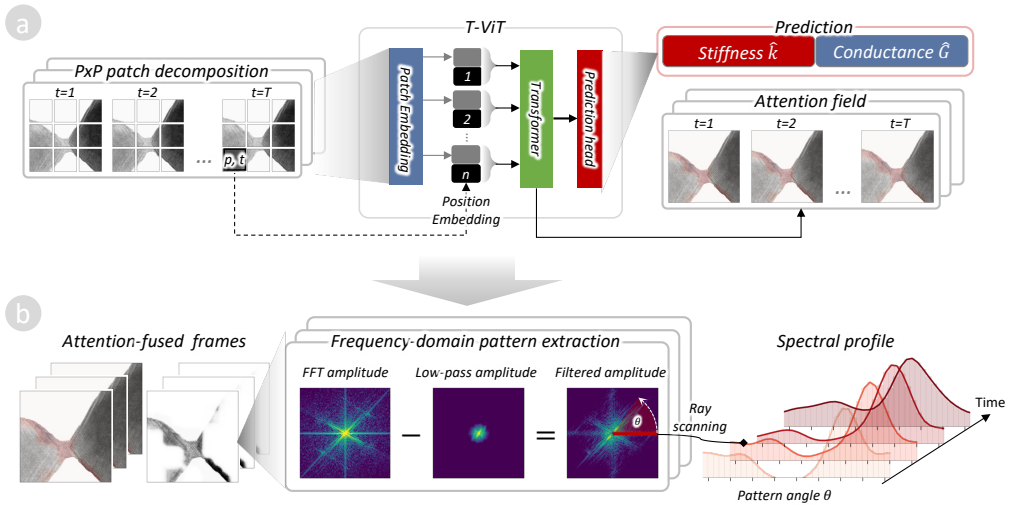

Figure 4.1: Overview of the attentive-inference framework for extracting insights into material dynamics from experimental image sequences. (**a**) Extracting spatiotemporal attention from structure–property dynamics: a transformer-based model learns how evolving structural configurations condition measured property changes, producing attention fields that encode the distribution of inferred structural contributions. (**b**) Deriving structural patterns in attention-highlighted regions: the attention fields are analyzed to extract descriptors of organization, revealing the inferred constraints that govern those changes.

These attention distributions form a representation in which observations are decomposed into interacting components, and relations among those components are explicitly encoded as relevance-weighted associations.

The second stage corresponds to the behavioral articulation stage of object inquiry. Rather than operating on raw observations, analysis is performed on the organization of attention-weighted representations. By examining the spatial arrangement, temporal persistence, orientation content, and recurrence of attention patterns, the framework characterizes how relevant structural variations are organized across space and time. Dynamic behavior is articulated through these relational patterns, which reveal how structural contributions accumulate, persist, or reorganize as the system evolves. In this way, attention is transformed from a purely computational weighting mechanism into a representational device that supports systematic interpretation of structure–property behavior.

4.1.2.1 Extracting spatiotemporal attention from structure–property dynamics

The first stage establishes the conditional relation between temporally evolving structural configurations and experimentally measured properties, yielding attention fields that highlight where and when local image variations exert the greatest statistical influence on measured responses. An experimental sequence $\{\mathbf{I}_t\}_{t=1}^{T}$, with each frame $\mathbf{I}_t \in \mathbb{R}^{M\times M}$, is paired with synchronized property records $\{\mathbf{y}_t\}_{t=1}^{T}$, where $\mathbf{y}_t \in \mathbb{R}^{J}$ represents one or more measured quantities (e.g., optical phase, strain, or conductivity). Each training sample consists of a fixed-length temporal window of T_w consecutive frames, $\{\mathbf{I}_{t-T_w/2}, \ldots, \mathbf{I}_{t+T_w/2}\}$, providing both backward and forward temporal context for predicting $\mathbf{y}_t$.

To capture dependencies across both spatial and temporal dimensions, a spatiotemporal Vision Transformer (TimeSformer; t-ViT) [75] is employed, following the self-attention formulation of Vaswani *et al.* [49]. Each frame is divided into non-overlapping patches of size $p \times p$, flattened, and linearly projected into patch embeddings. Spatial and temporal positional encodings are added to maintain order continuity, and the resulting sequence of tokens is processed by the transformer encoder. The model outputs a predicted property vector $\hat{\mathbf{y}}_t$ and a hierarchy of attention matrices $\{\mathbf{A}^{(l)}\}$ that represent the distribution of attention across spatial–temporal tokens at each layer l. The training objective minimizes a scale-weighted mean absolute error,

$$\mathcal{L} = \frac{1}{N}\sum_{t=1}^{N}\sum_{j=1}^{J}\lambda_j \left|\hat{y}_t^{(j)} - y_t^{(j)}\right|,$$

where λ_j rescales the property components according to their dynamic range, ensuring balanced optimization.

Because the attention mechanism operates on discrete patches, the native attention maps $\mathbf{A}^{(l)}$ are of coarse spatial resolution $(M/p) \times (M/p)$. To reconstruct a continuous, high-resolution representation suitable for physical interpretation, we employ an *ensemble accumulation* strategy that integrates model outputs across varying patch scales and shifted sampling positions. Multiple t-ViT models are trained independently using patch sizes $p \in \{5, 7, 11, 13\}$, capturing structural features across multiple spatial scales—from fine-grained local defects to extended deformation fronts. In addition, each input frame is subjected to controlled spatial translations of $\pm k$ pixels prior to tokenization. Because these translations alter the positions of patch boundaries, they effectively generate sub-patch offsets, allowing the ensemble of attention maps to interpolate finer structural details. The resulting ensemble of attention fields, $\{\mathbf{A}_t^{(p,k)}\}$, is first upsampled via bicubic

interpolation to the original resolution ($M \times M$) and then averaged element-wise across scales and translations:

$$\mathbf{A}_t = \frac{1}{N_{p,k}} \sum_{p,k} \mathbf{A}_t^{(p,k)}.$$

This accumulation procedure is equivalent to a Monte Carlo sampling over patch configurations, progressively reconstructing a high-resolution attention map that expresses the spatial continuity of structural relevance. Physically, $\mathbf{A}_t$ thus represents the model's empirical estimate of where and when local structural variations statistically condition the evolution of measured properties. These attention maps do not reveal direct causality but instead quantify the relative importance of structural regions whose variations collectively determine the system's macroscopic response [75, 110, 111].

To aggregate relevance across the hierarchical transformer layers, the attention roll-out method [76] is applied, recursively propagating layer-wise attention through residual connections:

$$\mathbf{R}^{(L)} = \mathbf{A}^{(L)}(\mathbf{I} + \mathbf{A}^{(L-1)}) \cdots (\mathbf{I} + \mathbf{A}^{(1)}),$$

yielding the cumulative relevance $\mathbf{R}^{(L)}$, which captures the total flow of attention from input tokens to the output prediction. The resulting attention field $\mathbf{A}_t = \mathbf{R}^{(L)}$ therefore visualizes how local structural configurations cooperatively modulate property evolution over time—a statistical reconstruction of change accumulation under correlation constraint.

4.1.2.2 Deriving structural patterns in attention-highlighted regions

The second stage translates the attention fields into physically interpretable descriptors that characterize the organization of structure underlying the inferred changes. Each attention field $\mathbf{A}_t$ serves as a statistical map of relevance but lacks direct physical meaning. To uncover the structural regularities that give rise to these patterns—periodicity, orientation, and coherence—we analyze the fused attention–image data in reciprocal space, where ordered organization manifests naturally as localized spectral energy distributions [112–114].

Since attention is learned in the model's latent representation space, its numerical values alone cannot be directly mapped to physical morphology. The first step therefore fuses the attention field with the corresponding experimental image to retain the original structural context:

$$\tilde{\mathbf{A}}_t = \sigma(\mathbf{A}_t), \quad \sigma(x) = \frac{1}{1 + e^{-x}}, \tag{4.2}$$

$$\mathbf{J}_t = \mathbf{I}_t \odot \tilde{\mathbf{A}}_t. \tag{4.3}$$

This fusion process weights the original image by its learned relevance, creating an attention-emphasized image $\mathbf{J}_t$ in which the spatial regions most influential to property evolution are locally enhanced. Physically, this operation integrates model-derived statistical evidence with observable morphology, ensuring that subsequent analysis refers to actual structural configurations rather than abstract attention values. The result may be viewed as a *selective observation*, retaining the experimentally resolved organization where correlation constraints are most pronounced.

Attention weighting introduces gradual intensity gradients that reflect the model's statistical confidence rather than true physical features. Likewise, illumination inhomogeneity or global contrast variations in the experimental images may dominate low-frequency spectral content, masking the subtle organization within high-frequency structural textures. To isolate the intrinsic organization, we estimate and remove the low-frequency component in Fourier space [115]:

$$\mathbf{J}_t^{\mathrm{LP}} = \mathcal{F}^{-1}(H_{\mathrm{LP}} \odot \mathcal{F}\{\mathbf{J}_t\}), \quad \mathbf{J}_t^{\Delta} = \mathbf{J}_t - \mathbf{J}_t^{\mathrm{LP}}, \tag{4.4}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are forward and inverse Fourier transforms, and H_{LP} is a circular low-pass mask that preserves the lowest 1–2% of the spectral radius. This filtering removes global variations while retaining the localized frequency content that encodes structural organization. The resulting $\mathbf{J}_t^{\Delta}$ represents a flattened contrast field emphasizing textural periodicities and directional patterns that arise from the internal constraints of the system.

The filtered image $\mathbf{J}_t^{\Delta}$ is then transformed into the amplitude spectrum,

$$\widetilde{\mathbf{M}}_t = \big|\, \mathcal{F}\{\mathbf{J}_t^{\Delta}\} \,\big|, \tag{4.5}$$

which reveals the distribution of spatial frequencies associated with repeating motifs or aligned features. Representing $\widetilde{\mathbf{M}}_t$ in polar coordinates (r, θ) enables quantification of orientation-dependent coherence using the variance-to-mean ratio statistic:

$$p_t(\theta_i) = \frac{\mathrm{Var}\Big(\{\widetilde{\mathbf{M}}_t(r,\theta_i)\}_r\Big)}{\mathrm{Mean}\Big(\{\widetilde{\mathbf{M}}_t(r,\theta_i)\}_r\Big)}, \quad \theta_i \in [0,\pi). \tag{4.6}$$

The resulting vector $\mathbf{p}_t = [p_t(\theta_0), \ldots, p_t(\theta_{n_\theta - 1})]^\top$ captures angular coherence, indicating whether the highlighted regions exhibit preferred alignments or

isotropic disorder. Peaks in $\mathbf{p}_t$ correspond to dominant orientations, while broad isotropic profiles signify uncorrelated or amorphous configurations.

Because the physical system may exhibit transient or overlapping modes of organization, the profiles $\{\mathbf{p}_t\}$ are compared across time to identify recurring organizational regimes. Pairwise dissimilarities are computed using Dynamic Time Warping (DTW) [116], which accommodates small angular offsets or symmetry-related shifts. Hierarchical clustering [46] is applied to the DTW matrix, and the optimal number of clusters is determined by silhouette analysis [117]. The clustered profiles summarize representative modes of structural organization emphasized by attention, each corresponding to a distinct type of correlation constraint that governs the system's evolution [17, 111].

This two-stage process establishes a coherent pathway from predictive modeling to inquiry-oriented interpretation. The first stage constructs attention fields that represent how localized structural variations across space and time are selectively associated with the measured response, providing a relevance-weighted representation of the observations. The second stage analyzes the organization of these attention-derived representations to articulate dynamic behavior, revealing how structurally relevant variations are arranged, persist, and recur as the system evolves. Through this design, attention functions as an analytical bridge between model-learned relevance and interpretable structure–property relations, enabling systematic examination of how heterogeneous structural organization gives rise to observable material behavior.

4.1.3 Experimental Settings

To evaluate the attention-based object-inquiry framework under experimental conditions that differ in physical determinacy and structural complexity, two representative systems were examined: deformation of gold nanocontacts (Au–NCs) and wet rubber–surface contact under sliding. These systems span a spectrum of material behavior, from nanoscale deformation governed by crystallographic regularity to macroscale interfacial dynamics shaped by fluid-mediated, heterogeneous interactions. This progression enables systematic assessment of how attention-based representations organize relations between structural observations and measured properties across regimes with increasing interpretive uncertainty.

The Au–NC system provides a setting in which structural evolution is partially constrained by known crystallographic symmetries. Deformation mechanisms such as slip, twinning, and necking occur within an ordered lattice framework, allowing localized structural variation to be interpreted against established physical regularities. In contrast, wet sliding involves a

highly disordered interface in which viscoelastic compliance, surface microtexture, and fluctuating water films jointly influence the emergent contact patterns. Structural organization in this regime is transient and spatially heterogeneous, lacking a fixed geometric reference. By considering both systems, the framework is evaluated on its ability to expose relations between localized structural variation and material response under markedly different organizational conditions.

Despite these differences, both datasets satisfy the essential requirements for attention-based object inquiry: time-resolved image sequences capturing structural evolution under external driving, synchronized with quantitative measurements of material response. These paired modalities enable examination of how distributed structural components contribute unevenly to observed behavior. Tables 4.1 and 4.2 summarize the principal dataset characteristics, including experiment counts, spatial resolution, acquisition rates, and the properties recorded for each system.

In the wet-contact system, high-speed optical imaging at 5000 Hz was synchronized with force-sensor measurements of the friction coefficient μ and slip rate S acquired at 1000 Hz. Each image sequence captures the transient evolution of the rubber–surface interface, where localized contact patches alternate between frictional engagement and fluid-supported lift. Within the object-inquiry framework, attention-based modeling is used to organize these observations by highlighting spatial regions and temporal intervals whose variations are most strongly associated with changes in μ and S. Analysis of the resulting attention-weighted representations focuses on how relevance is distributed and structured across the interface, revealing patterns such as directional streaking, hydrodynamic channeling, and transient clustering of microcontacts.

In the Au–NC system, structural evolution during elongation and compression was recorded using *in situ* transmission electron microscopy (TEM) at 30 Hz, synchronized with simultaneous conductance (G) and stiffness (k) measurements at 2.4 kHz. The image sequences encode the progression of atomic-scale rearrangements, including bond stretching, lattice slip, and neck thinning, that collectively influence electronic and mechanical response. Attention-based representations are constructed to organize relations between localized atomic domains and the measured properties, enabling examination of how crystallographic orientation, defect distribution, and structural coherence contribute to variations in G and k.

On this basis, the attention-based framework is evaluated according to two complementary criteria aligned with the stages of object inquiry:

Evaluating representational effectiveness. This criterion assesses whether the spatiotemporal attention model reliably captures the measured

Table 4.1: Measured properties and acquisition rates for the two datasets.

Case Study	Dataset	Measured Properties	Measuring Rate
1	Au-NC deformation	Stiffness k Conductance G	2.4 kHz
2	Wet rubber–surface contact	Slipping rate S Friction coefficient μ	100 Hz

Table 4.2: Experimental and imaging specifications for the two datasets.

Case Study	# Experiments	# Images	Image Size (px)	Imaging Rate
1	5	2,118	385 × 385	30 Hz
2	84	644,585	550 × 550	500 Hz

property evolution and whether the resulting attention distributions meaningfully organize the structural observations. Effectiveness is reflected not only in predictive agreement with experimental measurements, but also in the model's ability to concentrate relevance on structurally meaningful regions rather than diffuse or incidental patterns. This evaluation addresses whether the representation successfully encodes relations between localized structural variation and aggregated material response.

Evaluating behavioral articulation. This criterion examines whether analysis of attention-weighted representations reveals coherent patterns that articulate material behavior. In the wet-contact system, this includes identification of persistent orientation textures or recurrent spatial organizations within an otherwise fluctuating interface. In the Au–NC system, it involves recovery of lattice-aligned regions, defect-mediated anisotropy, or periodic structural organization consistent with known deformation modes. Successful behavioral articulation is indicated when attention-organized relations among observations can be interpreted in terms of physically meaningful structural organization.

Together, these criteria ensure that the attention-based framework is evaluated not merely as a predictive model, but as a representational mechanism for object inquiry. The first criterion establishes whether observations are effectively organized into relevance-weighted representations, while the second examines whether dynamic behavior can be articulated from the relations embedded within those representations. In this manner, predictive learning and interpretive analysis are integrated as successive stages of inquiry into structure–property behavior under complex experimental conditions.

4.2 Case Study I: Gold Nanocontact Deformation

Gold nanocontacts (Au–NCs) provide a natural foundation for introducing the attentive–inference framework because they constitute a physically constrained regime in which the governing mechanisms are partially understood and strongly shaped by crystalline geometry. Nano-contacts of this type arise routinely in scanning probe techniques such as scanning tunneling microscopy [118] and atomic force microscopy [119], where a metallic probe forms an atomic-scale junction with a surface. During elongation or compression, atoms within the constricted region reorganize through bond stretching, slip activation, lattice rotation, stacking-fault formation, and occasional twinning [120–122]. These events directly modulate electrical conductance and mechanical stiffness, two quantities widely used to characterize nanocontact behavior [26, 55, 123–125].

Despite extensive prior study, interpretation of nano-contact deformation still relies heavily on indirect or temporally sparse evidence. Conductance–elongation curves, for example, are typically interpreted through plateau structures that only coarsely encode the underlying atomistic rearrangements, while stiffness fluctuations are inferred as signatures of slip or bond transition events without direct visualization of the evolving geometry. In practice, these structural changes are highly localized: the dominant transformations occur within the narrow *nano-contact* region where atomic coordination is lowest, whereas the adjacent *surficial regions*—the tapered lattice segments leading toward the constriction—encode orientation-dependent signatures of slip direction, bond reorganization, and incipient defect formation. These regions act as intrinsic geometric references that reflect the constraints imposed by the crystal lattice.

Within this well-defined mechanical and crystallographic environment, the Au–NC system serves as the first validation setting for the attention-based object-inquiry framework. The combination of *in situ* TEM image sequences and synchronized electromechanical measurements provides a regime in which relations between localized structural variation and measured response are partially understood, offering a meaningful reference for evaluation. In this context, attention-based representations are used to organize spatial and temporal regions of the nanocontact according to their relevance to changes in conductance and stiffness, highlighting structurally significant domains such as surface-adjacent regions, slip-aligned planes, or zones of coherent lattice rotation. Because the lattice structure supplies a stable geometric baseline, relevance patterns exposed by attention can be directly

interpreted in terms of known deformation behavior. This makes the Au–NC case an essential foundation for assessing how attention-based representations articulate structure–property relations before extending the framework to systems with greater structural disorder and reduced physical regularity.

4.2.1 Modeling Spatiotemporal Attention

The Au–NC dataset provides a nanoscale test bed in which deformation is governed by atomic rearrangements, slip, and surface-mediated structural evolution—processes known to imprint directly on mechanical stiffness and electrical conductance [120–122]. Five complete approach–separation experiments were conducted to form and deform gold nanocontacts (Au–NCs), yielding time-resolved TEM image sequences synchronized with electromechanical measurements. Gold tips of 10 μm diameter were prepared by electrochemical etching and mounted on a piezo-actuated length-extension resonator, enabling controlled formation of a junction between tapered bases [26]. During each experiment, stiffness k and conductance G were sampled at 2.4 kHz using a four-probe configuration, while *in-situ* TEM images $\mathbf{I}_t$ were acquired at 30 Hz with a JEOL JEM-2000VF instrument operated at 200 kV under high vacuum ($< 2\times10^{-6}$ Pa). Temporal alignment was ensured by pairing each image with averaged electromechanical values computed over a sliding window of 80 samples. The complete dataset consisted of slightly over 2,000 frames, including 2,024 images containing formed nanocontacts and 94 frames depicting isolated tips, each associated with synchronized structural–property tuples.

Direct use of raw TEM frames is hindered by nonstationary background fluctuations, intensity drift, and low-contrast boundaries that obscure deformation-relevant regions [123, 124]. To isolate the metallic areas and suppress background dynamics, a U-Net [126] with a ResNet34 encoder [127] pretrained on ImageNet [128] was fine-tuned on 67 manually annotated frames. In comparison with a non-pretrained U-Net and Otsu thresholding [129], the pretrained model achieved the highest segmentation fidelity (IoU $= 0.975 \pm 0.014$), accurately delineating the neck region where lattice contrast is subtle but physically critical (Figure 4.2). Segmented images were cropped to 385×385 pixels, standardized in scale, and augmented through four rotations ($0°, 90°, 180°, 270°$) and dense temporal sampling (stride 1), yielding 8,016 training sequences from 2,004 originals.

Each augmented sequence of $T_w = 5$ consecutive frames was used to train a spatiotemporal Vision Transformer (t-ViT), tasked with predicting $\hat{\mathbf{y}}_t = [\hat{k}_t, \hat{G}_t]$ averaged across respective values of the frames (see more details

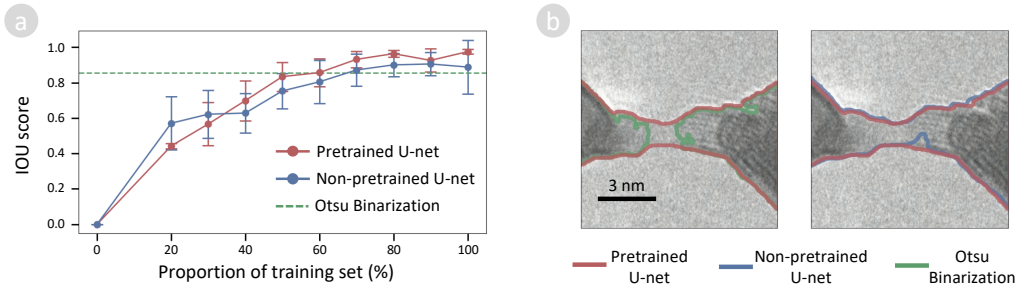

Figure 4.2: Comparison of TEM image segmentation obtained with pretrained and U-net model, non-pretrained U-net model, and Otsu Binarization method.

in Appendix B.1). Model optimization minimized a scale-balanced ℓ_1 loss,

$$\mathcal{L} = \frac{1}{N}\sum_{t=1}^{N}\left(\lambda_k|\hat{k}_t - k_t| + \lambda_G|\hat{G}_t - G_t|\right), \tag{4.7}$$

where λ_k and λ_G normalize stiffness and conductance to comparable numerical scales. In this dataset, the two properties share similar value ranges, and therefore $\lambda_k = \lambda_G = 1$ was used to maintain balanced contributions during training without additional rescaling. To capture multi-scale structural cues—from lattice fringes to mesoscale neck curvature—an ensemble was constructed across patch sizes $p \in \{5, 7, 11, 13\}$. Training each variant on a single NVIDIA A100 GPU required 3–7 hours with memory usage of 3–20 GB.

Inference robustness was examined using a perturbed testing set generated by controlled spatial translations (20–40 pixels along eight compass directions) and minor temporal distortions that replicate realistic acquisition irregularities. This procedure produced 48,096 testing instances [26], all evaluated under identical conditions across patch sizes. Ensemble predictions were computed by averaging across crops and model variants, achieving $R^2 = 0.97$ and MAE = 5.96 N/m for stiffness, and $R^2 = 0.99$ and MAE = 3.18 G_0 for conductance (Figure 4.3, Table 4.3). These results confirm stable generalization across nanoscale configurations not observed during training.

The primary goal of this case study is to determine whether the model-derived attention fields $\mathbf{A}_t$ encode meaningful information about Au-NC deformation guiding respective response in stiffness and conductance. In the Au–NC geometry, two physically active zones dominate the electromechanical response: (i) the *nanocontact region*, where the cross-section narrows

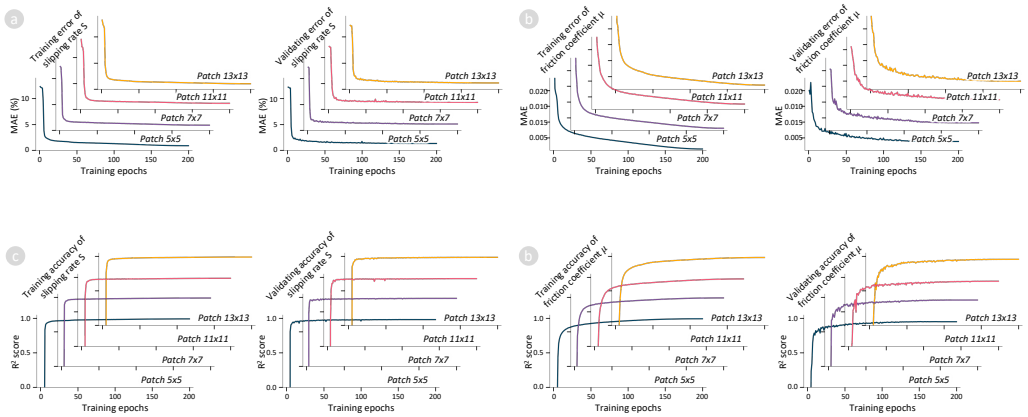

Figure 4.3: Training and validating curve of t-ViT models trained with Au-NC TEM images decomposed into 5×5, 7×7, 11×11, and 13×13 patches . (a) Training and (b) validating errors (MAE) of estimating stiffness k (N/m) and conductance G (G_0). (c) Training and (d) validating accuracy (R^2 score) of estimating stiffness k (N/m) and conductance G (G_0).

Table 4.3: Prediction accuracy of t-ViT models with different patch sizes $p \in \{5, 7, 11, 13\}$ evaluated on the testing dataset derived from the Au–NC deformation dataset, predicting stiffness k (N/m) and conductance G (G_0)

Patches ($p \times p$)	**R^2 (k / G)**	**MAE (k / G)**
5 × 5	0.95 / 0.96	8.27 / 5.72
7 × 7	0.97 / 0.99	5.83 / 3.15
11 × 11	0.97 / 0.99	5.99 / 2.99
13 × 13	0.97 / 0.99	6.00 / 2.34
Ensemble	0.97 / 0.99	5.96 / 3.18

and stresses concentrate, and (ii) the *surficial regions* comprising atomic terraces, steps, and low-coordination sites that accommodate slip, migration, and lattice rotation [26,54,130,131]. These regions regulate how deformation unfolds—through neck thinning, surface diffusion, or discrete rearrangement events—and thus impose recognizable constraints on the observed stiffness and conductance signals [132, 133]. If the learned representation accurately reflects this physics, high attention magnitudes should align with these specific geometric zones.

To assess this alignment, the metallic boundary $\mathcal{C}$ and constriction axis $\mathcal{L}$ were extracted from each segmented frame. Two distance fields were

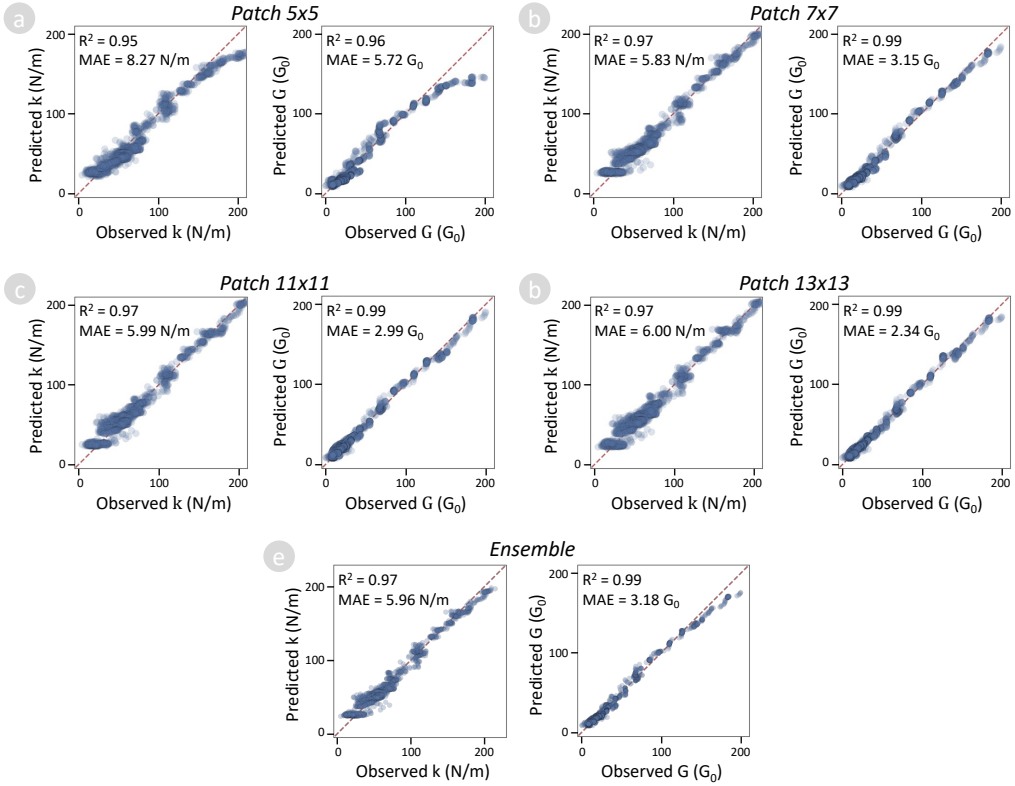

Figure 4.4: Comparison between predicted and observed values of stiffness and conductance using the t-ViT models trained with different decomposition of Au-NC TEM images. (a-d) Images are decomposed into: (a) 5×5 patches, (b) 7×7 patches, (c) 11×11 patches, (d) 13×13 patches.

computed for each pixel (x, y) within the metallic region:

$$d_{\text{cons}}(x, y) = \|(x, y) - \Pi_{\mathcal{L}}(x, y)\|_2, \tag{4.8a}$$

$$d_{\text{surf}}(x, y) = \min_{(x', y') \in \mathcal{C}} \|(x, y) - (x', y')\|_2, \tag{4.8b}$$

where $\Pi_{\mathcal{L}}(x, y)$ denotes orthogonal projection onto the constriction axis. Random pixel samples were binned by attention magnitude, and interval-averaged distances were computed. Both d_{cons} and d_{surf} decrease monotonically with increasing attention (Figure 4.5a), indicating that the model preferentially highlights the regions closest to the neck and surfaces—consistent with known sites of stress localization, atomic mobility, and conductance sensitivity [26, 124].

Together, these observations indicate that the attention fields recover physically meaningful structural deformation of Au-NCs: structural rear-

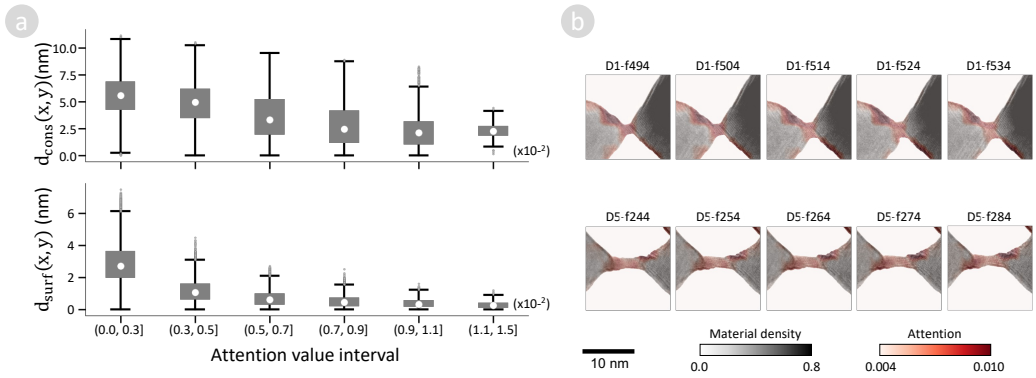

Figure 4.5: Evaluations of the deep-learning models used for processing pipeline. (a) Spatial correlation between the model-derived attention and geometric proximity to the constriction and surface. (b) Representative examples of model-derived attention maps overlaid on TEM images.

rangements originate at the neck and adjacent surfaces, their evolution controls both stiffness and conductance, and the model learns to assign relevance in accordance with these physically grounded mechanisms. The recovered spatial distributions thus provide a reliable basis for subsequent extraction of persistent orientation and deformation patterns, examined in the next subsection.

4.2.2 Interpreting Attentive Patterns

Having established that attention consistently concentrates on the constriction and surface terraces—the nanoscale regions most strongly associated with variations in stiffness and conductance—we now analyze how dynamic deformation behavior is expressed through the organization of these regions. In the Au–NC system, atomic rearrangements induced by external loading manifest as characteristic patterns of lattice reorientation rather than arbitrary fluctuations. Attention-based representations expose relations among these localized structural variations by highlighting regions whose orientation content and persistence co-vary with the measured response. By examining angular distributions, spatial localization, and temporal stability within the attentive regions, lattice reorientation behavior is articulated directly from the observed structural characteristics. This analysis therefore links measurable deformation behavior to relations implied by attention-weighted observations, providing a concrete basis for interpreting nanoscale mechanical response.

To extract these patterns, each of the five TEM sequences was segmented

into overlapping windows of 50 consecutive frames (stride 10), producing multiple partially independent samples along each deformation trajectory. For every frame within a window, the attention-weighted image $\mathbf{J}_t = \mathbf{I}_t \odot \sigma(\mathbf{A}_t)$ was transformed into a Fourier amplitude map $\widetilde{\mathbf{M}}_t(r, \theta)$, from which an angular coherence profile was computed,

$$p_t(\theta_i) = \frac{\mathrm{Var}\big(\{\widetilde{\mathbf{M}}_t(r, \theta_i)\}_r\big)}{\mathrm{Mean}\big(\{\widetilde{\mathbf{M}}_t(r, \theta_i)\}_r\big)}, \quad \theta_i \in [0, \pi), \tag{4.9}$$

quantifying directional persistence of lattice contrast. Averaging these profiles within each temporal window suppresses frame-specific fluctuations while emphasizing orientations that remain active throughout the interval. Pairwise dynamic time warping dissimilarities between the window-averaged profiles were then embedded with t-SNE, and hierarchical clustering identified two distinct categories of attentive organization (silhouette coefficient 0.68), indicating that the deformation process expresses two recurrent and clearly separable structural modes.

Figure 4.6c shows representative angular profiles extracted from the attention-weighted structural representations. Group 1 exhibits a single dominant peak near 50°, indicating a persistent lattice orientation maintained throughout the deformation interval. This narrow and coherent peak indicates that deformation proceeds predominantly through one crystallographic shear pathway, consistent with single-system slip in face-centered cubic metals [54, 122, 130]. In this regime, atomic planes glide cooperatively along a crystallographically compatible direction, resulting in minimal geometric incompatibility within the nanocontact. The sustained concentration of attention along a single orientation reflects that structural variation contributing to the measured response remains organized around one dominant deformation mode throughout the loading process.

Group 2, in contrast, displays multiple peaks spanning a wide angular range ($-50°$ to 50°), indicating that several lattice orientations appear and persist within the observation window. This broadened angular spectrum signals the participation of multiple deformation modes, such as activation of additional shear pathways or the emergence of twin-related domains, which are characteristic of nanoscale plasticity under geometric confinement [123, 131]. As the constriction narrows and continued single-system glide becomes unfavorable, the lattice reorganizes through cross-slip, local rotation, or twinning. These processes introduce intersecting domains and reoriented fringes, giving rise to a diverse set of angular signatures. Correspondingly, attention becomes distributed across multiple orientations, reflecting how deformation activity reorganizes spatially and crystallographically as the

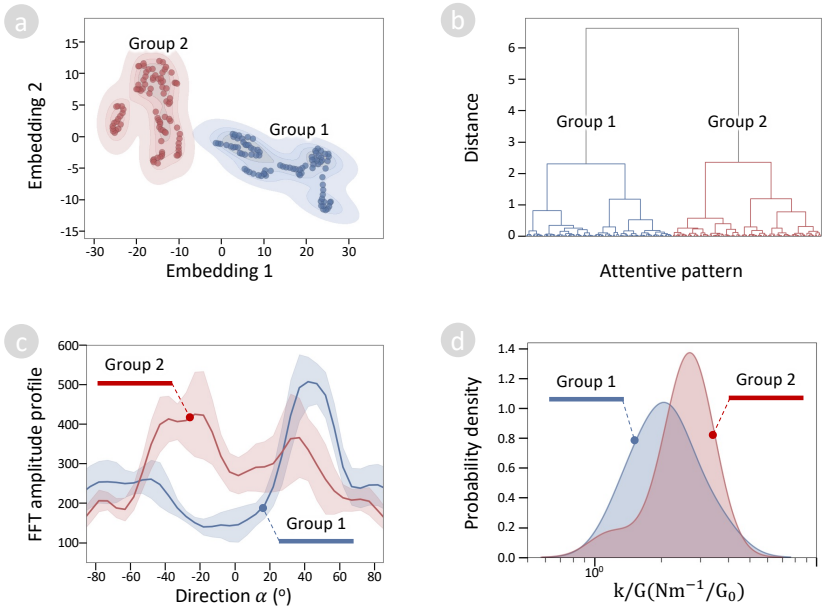

Figure 4.6: Categorization of lattice-oriented attentive patterns during Au–NC deformation. (**a**) t-SNE embedding of window-averaged angular profiles, with color indicating cluster assignment. (**b**) Dendrogram from hierarchical clustering applied to DTW dissimilarities. (**c**) Representative angular profiles for the two identified groups, expressed in real-space orientation α (converted from reciprocal-space angle θ). (**d**) Distributions of stiffness-to-conductance ratios (k/G) associated with windows in each group, illustrating systematic mechanical differences between the two deformation modes.

nanocontact adapts to the evolving geometric constraint.

These structural modes correlate directly with mechanical response (Figure 4.6d). Group 1, dominated by a single coherent orientation, exhibits systematically lower stiffness-to-conductance ratios (k/G), consistent with facile shear along a compatible slip system. Group 2, with its multi-orientation distributions, correlates with higher k/G, indicating increased rigidity when the lattice must accommodate deformation through constrained, intersecting mechanisms. This contrast illustrates how the attentive inference pipeline reveals not only where deformation occurs but how the underlying constraints shape atomic reorganization and, consequently, the observed electromechanical evolution.

To visualize these modes directly, Figure 4.7 presents representative image sequences from each group, showing the original TEM frames, their high-pass filtered versions, and attention-weighted overlays (see details of

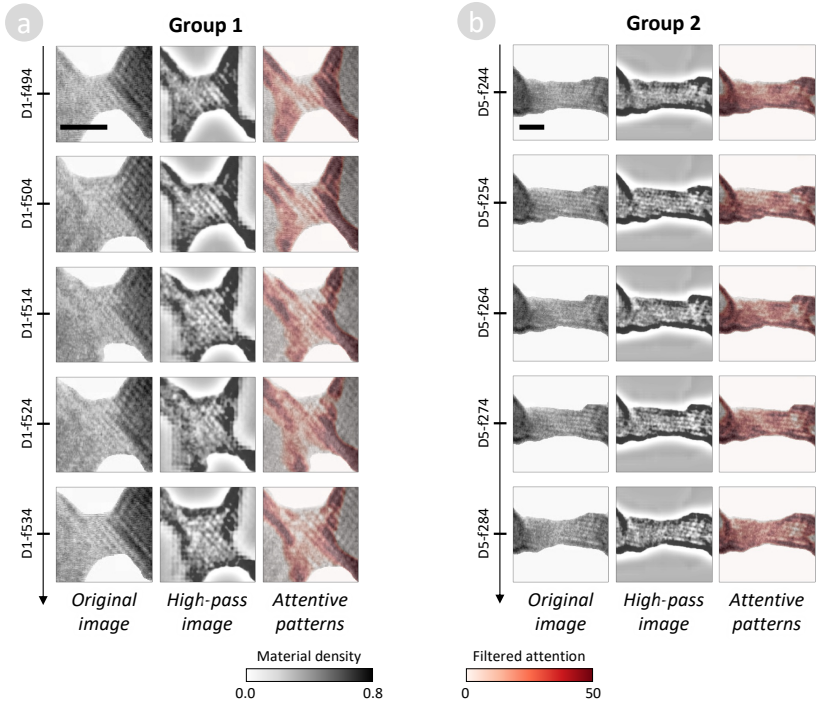

Figure 4.7: Representative deformation sequences for the two attentive-pattern groups. For each example, the original TEM frame, its high-pass counterpart, and the attention-filtered image are shown.

pattern filtering process in Appendix B.2). In Group 1, lattice fringes remain aligned across the constriction, producing uniform, elongated domains. In Group 2, the attentive regions highlight intersecting sets of fringes, rotated domains, and localized reorientation events–the structural signatures of twin-like or multi-slip accommodation. These visual differences reinforce the interpretation that the inferred constraints correspond to the lattice's available deformation pathways under nanoscale confinement.

In summary, analysis of the Au–NC system reveals two recurring forms of deformation organization articulated through attention-weighted structural representations. Group 1 corresponds to a mode dominated by single-system slip, in which lattice evolution proceeds along a crystallographically compatible pathway and structural variation remains narrowly organized. Group 2 corresponds to a mode in which geometric confinement and evolving contact shape promote reorganization through multiple deformation mechanisms, including cross-slip, local rotation, or twin-like reorientation. These modes delineate the characteristic ways in which the lattice accommodates applied loading and manifest as distinct structural patterns most strongly associated

with variations in conductance and stiffness. By isolating and comparing these recurring patterns, the framework establishes a direct connection between time-resolved microscopy observations and the dominant deformation behaviors that govern nanoscale mechanical and electronic response.

4.3 Case Study II: Wet Rubber–Surface Contact

Wet rubber–surface contact under sliding represents a highly variable and weakly constrained interfacial regime in which frictional performance emerges from the coupled influence of viscoelastic response, microtexture, and fluid-mediated separation. When a rubber sample slides over a water-covered surface, a thin lubricating film forms and breaks repeatedly as water escapes through the contact edge. This continual formation and disruption of microcontacts produces strong fluctuations in load support, leading to intermittent reductions in friction. The interfacial organization evolves over sub-millisecond timescales, and the spatial structures that appear within the contact region are transient, fragmented, and composition-dependent.

Within this study, the wet-contact system is investigated across multiple rubber formulations that differ in viscoelastic compliance, energy dissipation characteristics, and microstructural filler-induced texture. These compositional variations induce systematic modulation of wet-friction behavior: compliant formulations may conform more effectively yet generate weaker shear support, whereas stiffer or more structured formulations can stabilize broader, more coherent wet-contact textures that contribute to higher friction under lubrication. Because these effects arise from intrinsic material properties rather than engineered geometry, the wet-contact interface provides a direct view of how formulation-dependent mechanics shape frictional performance.

In this context, the wet-contact system serves as the second pillar of the attentive–inference framework, representing the macroscopic limit of interpretive uncertainty. Unlike lattice-governed systems, which provide a stable geometric reference for interpreting deformation, the wet-contact interface lacks any fixed structural baseline. Its apparent disorder conceals transient but functionally meaningful organization that must be inferred statistically. The relevant structures manifest as evolving *wet-contact textures*—coherent diagonal textures, fragmented traces, or bulk-like accumulations—that encode how the material engages the lubricated interface.

The analytical objectives in this regime therefore focus on organizing and interpreting these textures as expressions of dynamic contact behavior.

Localized variations in intensity and microcontact expression reflect immediate material responses to water-film breakup, viscoelastic deformation, and surface microtexture, while recurrent orientation patterns reveal how such variations are organized across space and time. These textures provide observable signatures of how frictional behavior emerges from distributed interfacial activity rather than from fixed geometric features.

By analyzing these compositionally varied, fluid-mediated contact sequences, this case study tests whether the attentive–inference framework can recover physically meaningful wet-contact textures and relate them to the frictional performance of each formulation. It embodies the framework's most challenging inferential setting: a system with minimal geometric determinacy, strong stochastic fluctuations, and friction governed by emergent material organization rather than fixed structural symmetry. The subsequent analysis examines whether spatiotemporal attention can localize meaningful interfacial changes and reveal the constraint structures that underpin wet-friction behavior across different material formulations.

4.3.1 Modeling Spatiotemporal Attention

To examine how evolving structural configurations govern macroscopic frictional behavior in wet rubber-surface contact, the attentive-inference framework was applied to paired sequences of contact images and synchronized property measurements. Each experiment yielded a time series $\{(\mathbf{I}_t, \mathbf{y}_t)\}_{t=1}^{T}$, where $\mathbf{I}_t \in \mathbb{R}^{550\times550}$ represents the instantaneous contact imprint, and $\mathbf{y}_t = [S_t, \mu_t]$ denotes the measured slip rate and friction coefficient. The experiments were conducted using a laboratory tribometer equipped with a transparent substrate, high-speed optical imaging (5000 Hz), and integrated force and displacement sensors recording at 1000 Hz. In total, 84 independent experiments were performed across rubber compositions, uniformly flat samples, and hydrodynamic conditions, generating approximately 6.4×10^5 frames that collectively describe the transient evolution of wetting process [134, 135].

Each raw image was cropped to 550×550 pixels to encompass the entire contact footprint, and sequential segments of $T_w = 5$ frames were used as model inputs, capturing short-term temporal dynamics over a millisecond. The dataset consisted of 128,916 original sequences, which were expanded through overlapping sampling and controlled spatial translations to increase diversity. No rotational augmentation was applied to preserve the directionality of tread and drainage features. These samples were used to train a spatiotemporal Vision Transformer (t-ViT) ensemble, designed to map evolving image sequences to property vectors $\mathbf{y}_t = [S_t, \mu_t]$.

Each t-ViT model divided image frames into non-overlapping patches

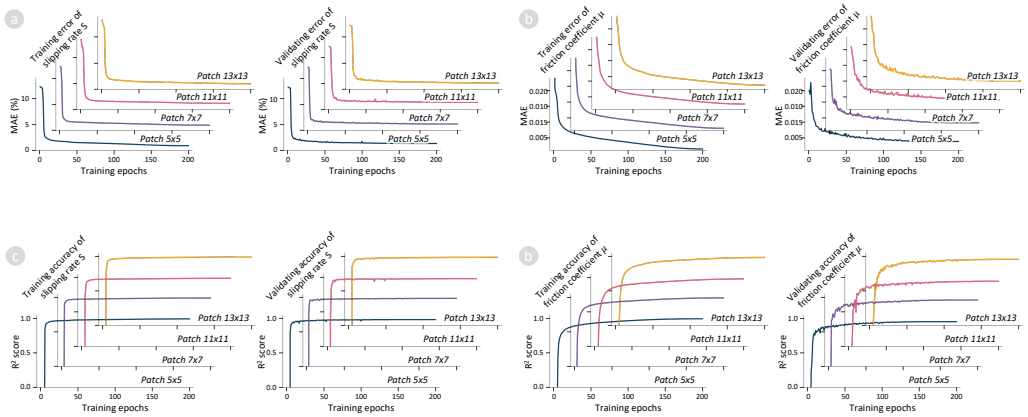

Figure 4.8: Training and validating curve of t-ViT models trained with rubber–surface contact images decomposed into 5×5, 7×7, 11×11, and 13×13 patches . (a) Training and (b) validating errors (MAE) of estimating slipping rate (%) and friction coefficient μ. (c) Training and (d) validating accuracy (R^2 score) of slipping rate (%) and friction coefficient μ.

and jointly processed spatial and temporal embeddings to learn structure–property correlations (see details in Appendix B.1). The training minimized a scale-weighted mean absolute error,

$$\mathcal{L} = \frac{1}{N} \sum_{t=1}^{N} \sum_{j \in \{S,\mu\}} \lambda_j \, |\hat{y}_t^{(j)} - y_t^{(j)}|, \tag{4.10}$$

where λ_j normalizes each property by its characteristic range. Models were trained across patch sizes $p \in \{5, 7, 11, 13\}$ to ensure sensitivity to both fine microcontact textures and mesoscale drainage geometries. Each configuration was trained on a single NVIDIA A100 GPU for approximately 2–6 days, with memory requirements ranging from 3 GB to 20 GB. The outputs of all trained models were subsequently averaged to obtain ensemble predictions and smoothed attention fields $\mathbf{A}_t$, reducing model-specific bias and emphasizing scale-consistent structures.

To evaluate generalization and to refine attention resolution, a large perturbed testing dataset was generated by spatially translating 500 × 500 crops within the 550×550 images by 50 pixels along eight compass directions. Combined with overlapping temporal sampling, this produced ≈ 2.5 millions test sequences that systematically covered variations in position and context. Each model evaluated these perturbed sequences independently, and the resulting outputs were aggregated by averaging across both spatial and patch-

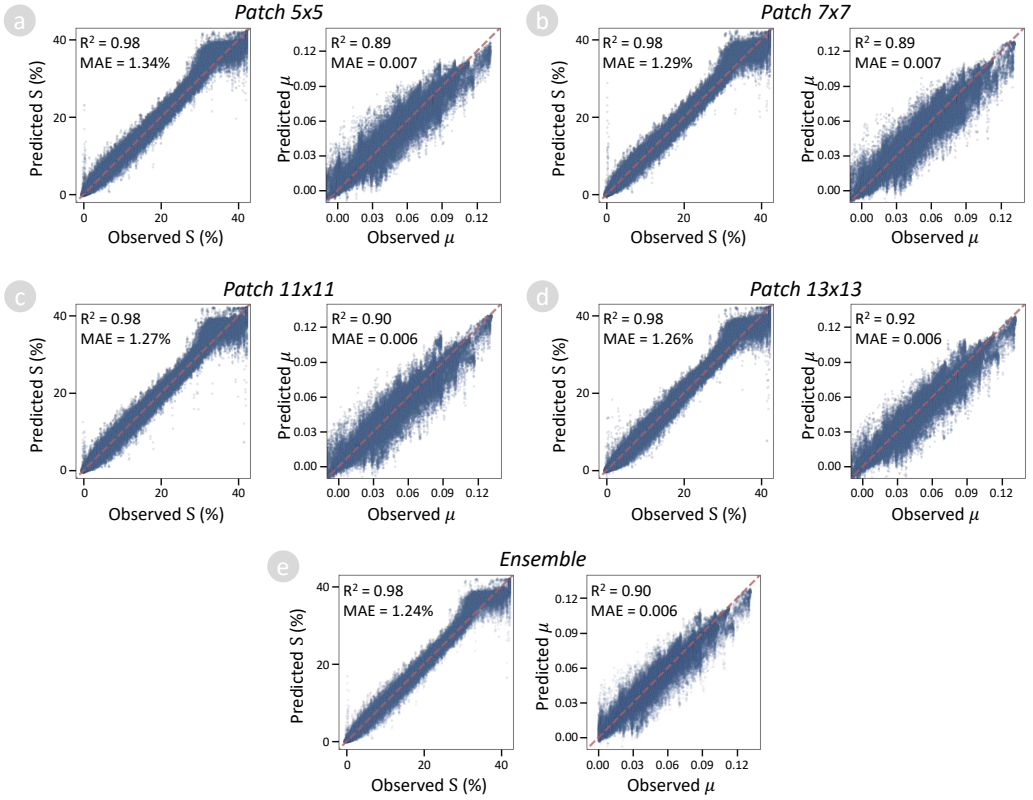

Figure 4.9: Comparison between predicted and observed values of slipping rate and friction coefficient using the t-ViT models trained with different decomposition of rubber–surface contact images. (a-d) Images are decomposed into: (a) 5×5 patches, (b) 7×7 patches, (c) 11×11 patches, (d) 13×13 patches.

scale ensembles. This process yielded high-resolution, translation-invariant attention maps that emphasize statistically consistent regions of relevance across the interface.

The ensemble achieved strong predictive performance, with $R^2 = 0.98$ (MAE = 1.24%) for the slip rate and $R^2 = 0.90$ (MAE = 0.006) for the friction coefficient (Figure 4.8 and Table 4.4). These results demonstrate that the model effectively learned the nonlinear relationship between evolving contact morphology and measured tribological responses. This predictive reliability provides the foundation for interpreting attention as a meaningful representation of structural causality—specifically, for identifying the regions that statistically drive variations in macroscopic frictional behavior.

To assess whether the model-derived attention maps $\mathbf{A}_t$ correctly localize

Table 4.4: Prediction accuracy of t-ViT models with different patch sizes $p \in \{5, 7, 11, 13\}$ evaluated on the testing dataset derived from the wet rubber–surface contact dataset, predicting slipping rate S (%) and friction coefficient μ.

Patches ($p \times p$)	$\mathbf{R}^2$ **(S / μ)**	**MAE (S / μ)**
5×5	0.98 / 0.88	1.33 / 0.007
7×7	0.98 / 0.89	1.29 / 0.006
11×11	0.98 / 0.90	1.27 / 0.006
13×13	0.98 / 0.92	1.26 / 0.006
Ensemble	0.98 / 0.90	1.24 / 0.006

the structural origins of the inferred changes in wet-contact sliding, we examined their correspondence with experimentally meaningful attributes of the rubber–surface interface. Two spatial attributes were considered. The first is the local contact intensity $\rho(x, y)$, which reflects the degree of load transfer and imprint strength at each point in the contact patch. The second is the distance to the contact edge, $d_{\text{edge}}(x, y)$, defined as the Euclidean distance from pixel coordinates (x, y) to the nearest boundary of the observed patch. Under wet sliding, this boundary acts as the primary pathway through which water escapes the interface, producing rapid film-breakup events and strong spatiotemporal fluctuations in the adjoining region. Although drainage-related dynamics play a role analogous to the shoulder-dominated behavior observed in treaded systems [134,135], the mechanism here originates entirely from geometric boundary effects in a flat sliding configuration.

As shown in Figure 4.10, the learned attention fields exhibit clear spatial organization with respect to physically relevant surface characteristics. Quantitative analysis in Figure 4.10a shows that the attention magnitude correlates positively with $\rho(x, y)$ and negatively with $d_{\text{edge}}(x, y)$. This indicates that the model systematically attributes higher relevance to (i) contact-rich interior regions that sustain shear load and (ii) boundary-adjacent zones where water-film breakup and reattachment occur. These allocations confirm that the attention mechanism responds simultaneously to load-bearing microcontacts and to drainage-driven fluctuations—two processes known to jointly shape wet-friction behavior across rubber formulations. Representative overlays of the attention distribution on surface images are displayed in Figure 4.10b, illustrating how regions identified by the model align with the spatial structures associated with these mechanisms.

Together, these behaviors demonstrate that the learned attention par-

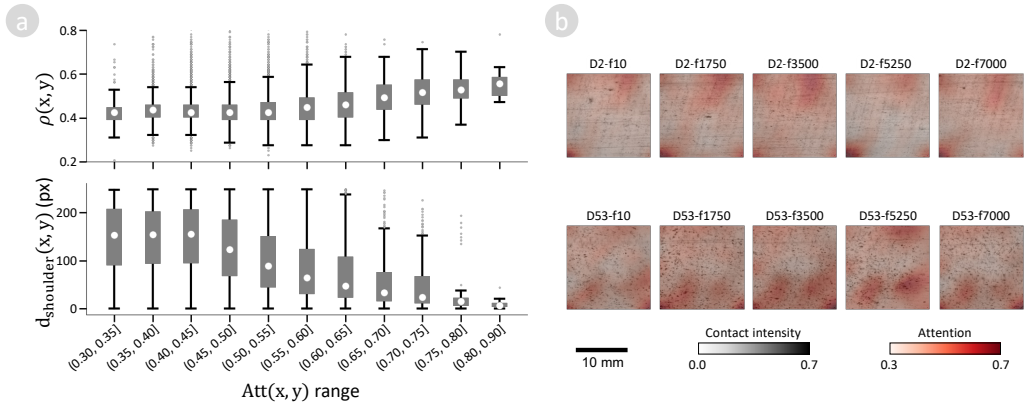

Figure 4.10: Efficiency evaluations of the deep-learning models applied to wet-contact images. (a) Comparison between predicted and observed values of slip rate S and friction coefficient μ using the t-ViT model ensemble. (b) Correlations between the model-derived attention and reference attributes of the contact interface, including contact intensity and distance to sample edge. (c) Representative examples of model-derived attention maps overlaid on contact images.

titions the wet-contact interface into physically interpretable subregions shaped by material properties: highly dynamic edge zones dominated by drainage-driven inferred changes, and interior regions expressing coherent wet-contact textures that act as constraint-bearing structures. This dual representation shows that the attentive–inference framework does more than predict macroscopic tribological quantities—it resolves how material-dependent interfacial organization regulates friction under wet sliding, translating statistical attention fields into interpretable evidence of texture stabilization and constraint formation in lubricated rubber–surface contact.

4.3.2 Interpreting Attentive Patterns

Following the construction of attention-based representations that organize relations among localized structural variations, this section addresses the second stage of object inquiry: articulating dynamic behavior from the organization of those relations. To this end, the attention-weighted images $\mathbf{J}_t = \mathbf{I}_t \odot \sigma(\mathbf{A}_t)$ are transformed into descriptors that characterize directional organization across each experimental sequence. These descriptors capture how interfacial regions fluctuate, reorganize, and recur as slip progresses, revealing both short-lived structural variability and persistent orientation tendencies over time. Wet rubber–surface contact provides a particularly informative setting for this analysis, as its interface exhibits extensive

fragmentation, intermittent reattachment, and recurring drainage pathways, producing a wide range of structural expressions that vary across time and experiments.

For each experiment, attention-fused images were transformed into amplitude spectra $\widetilde{\mathbf{M}}_t(r,\theta)$ using Fourier analysis, where repeating or preferentially aligned features give rise to angular concentrations of spectral energy. To summarize these frequency-domain expressions, time-resolved spectra were reduced to angular coherence profiles,

$$p_t(\theta_i) = \frac{\mathrm{Var}\{\widetilde{\mathbf{M}}_t(r,\theta_i)\}_r}{\mathrm{Mean}\{\widetilde{\mathbf{M}}_t(r,\theta_i)\}_r}, \qquad \theta_i \in [0,\pi), \tag{4.11}$$

which quantify the directional persistence of structural motifs emphasized by attention. Time-averaged profiles, computed for each experiment, provide a compact representation of how interfacial organization evolves and fluctuates over the full sliding cycle.

To identify dominant modes of dynamic organization across formulations and experiments, the angular coherence profiles extracted from the attention-weighted observations were compared using Dynamic Time Warping (DTW), embedded via t-SNE, and clustered using hierarchical linkage. The embedding shown in Figure 4.11a reveals two well-separated regions, and the linkage hierarchy in Figure 4.11b confirms a stable two-group structure with a silhouette score of 0.51. This separation reflects systematic differences in rubber composition and how these differences influence the organization of the wet-contact interface under lubrication. Rather than representing discrete material classes, the two groups capture recurring modes of interfacial organization that form the basis for articulating distinct frictional behaviors.

Representative angular coherence profiles from each group are shown in Figure 4.11c. Both groups exhibit dominant orientation components centered near $\pm 20^\circ$, corresponding to diagonal structural organization commonly associated with effective shear support in wet-contact conditions. The distinction lies in the broader angular structure surrounding these preferred orientations. Group 1 displays multiple strong peaks across a wide angular range—for example, near -40° and across bands extending between 20° and 60°. This diversity indicates dispersed directional repetition, suggesting that interfacial forces are distributed across competing structural pathways. In texture space, these profiles correspond to fragmented wet-contact textures, characterized by numerous narrow features, frequent reorientation, and rapidly reorganizing diagonal traces as the water film repeatedly breaks and reforms.

By contrast, Group 2 exhibits fewer and lower-magnitude peaks, indicating spatially broader and more coherent orientation accumulations.

These profiles correspond to coherent wet-contact textures, in which bulk-like features dominate and spectral repetition is muted. Such organization reflects stabilization of coarse-scale interfacial structures, shaped jointly by viscoelastic response and microstructural composition, and governs how load is supported and transmitted through the lubricated interface during sliding.

The functional implications of these texture modes are reflected in the frictional measurements. As shown in Figure 4.11d, experiments associated with Group 1 exhibit lower peak friction coefficients μ, consistent with force dissipation along unstable or inefficient interfacial pathways. The fragmented textures characteristic of this group provide only intermittent load-support channels, making the interface more susceptible to fluid-mediated separation. In contrast, Group 2 produces more substantial increases in μ, reaching higher peak friction values across formulations. These improvements align with the stabilizing influence of coherent, bulk-like textures, which furnish broader and more continuous pathways for shear transfer despite the presence of lubrication. Notably, although the preferred $\pm 20^\circ$ orientations resemble angular ranges sometimes used in angled-tread concepts for wet traction [136], the present experiments involve flat rubber samples without engineered drainage geometry. The observed orientations arise intrinsically from the material's viscoelastic compliance, microtexture, and composition-dependent interfacial mechanics. Their recurrence across formulations indicates a favorable mode of wet-contact organization that enhances shear support under sliding. The contrast between fragmented textures (Group 1) and consolidated textures (Group 2) therefore illustrates how the attentive–inference framework links material-dependent orientation structures to frictional performance, identifying the coherent wet-contact textures that act as constraint-bearing features in the lubricated interface.

To visualize how these organizational modes manifest in real space, representative examples from each group were processed using background suppression and attention-based fusion (see Appendix B.2 for the filtering procedure), as shown in Figure 4.3.2. In Group 1, the filtered images expose intersecting, fine-scale traces that reorganize rapidly during sliding. These fragmented wet-contact textures reflect unstable force pathways and frequent directional switching, consistent with the numerous, narrowly spaced peaks observed in their angular coherence profiles. By contrast, Group 2 exhibits large, coherently aligned accumulations oriented near the dominant diagonal directions. These bulk-like textures correspond to the broader, low-frequency peaks in the angular spectra and indicate a more consolidated mode of interfacial engagement in which load is transferred through spatially extended, stable pathways.

The emergence of these two regimes carries clear material implications.

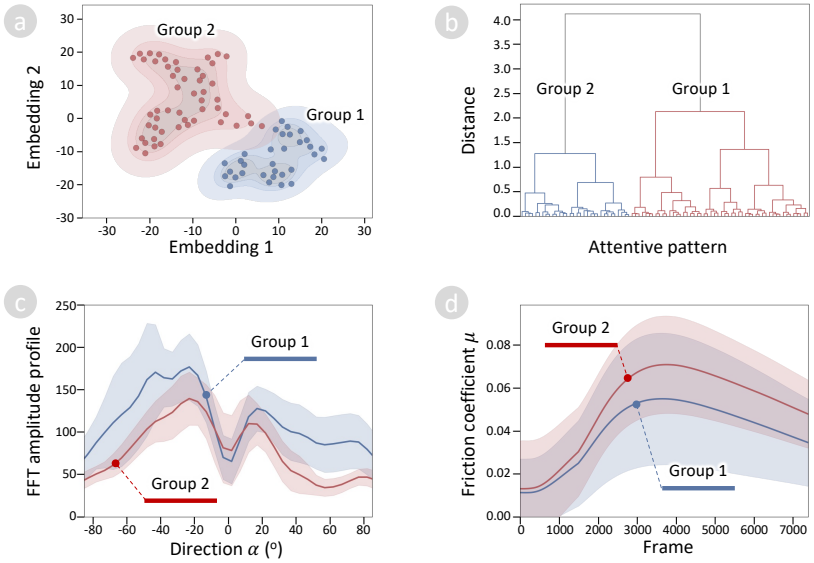

Figure 4.11: Clustering attentive organizational patterns in wet rubber–surface contact. (a) t-SNE embedding of angular coherence profiles across all experiments. (b) Hierarchical clustering dendrogram computed from DTW distances. (c) Representative angular coherence profiles of the two identified groups.(d) Temporal evolutions of friction coefficient μ for experiments in each group.

Since the rubber samples differ only in formulation, the distinct textures produced by each group indicate that composition-dependent viscoelastic and microstructural properties govern how wet contact organizes under lubrication. In Group 1, the fragmented textures suggest that the formulation dissipates forces across competing pathways, a behavior consistent with higher compliance, greater local loss, or microtextures that do not support sustained shear. In Group 2, the broad, coherently aligned accumulations imply that the formulation can conform effectively to the lubricated surface while maintaining sufficient stiffness and dissipation to preserve load-bearing contact. These characteristics promote the formation of bulk-like regions that carry shear more robustly and resist separation as the water film forms and breaks.

Taken together, these observations indicate that the mechanically favorable mode of interfacial organization under wet sliding is characterized by coherent, diagonally aligned wet-contact textures. Their ability to stabilize broad load-transfer pathways is directly associated with higher frictional performance across formulations. This relationship provides a

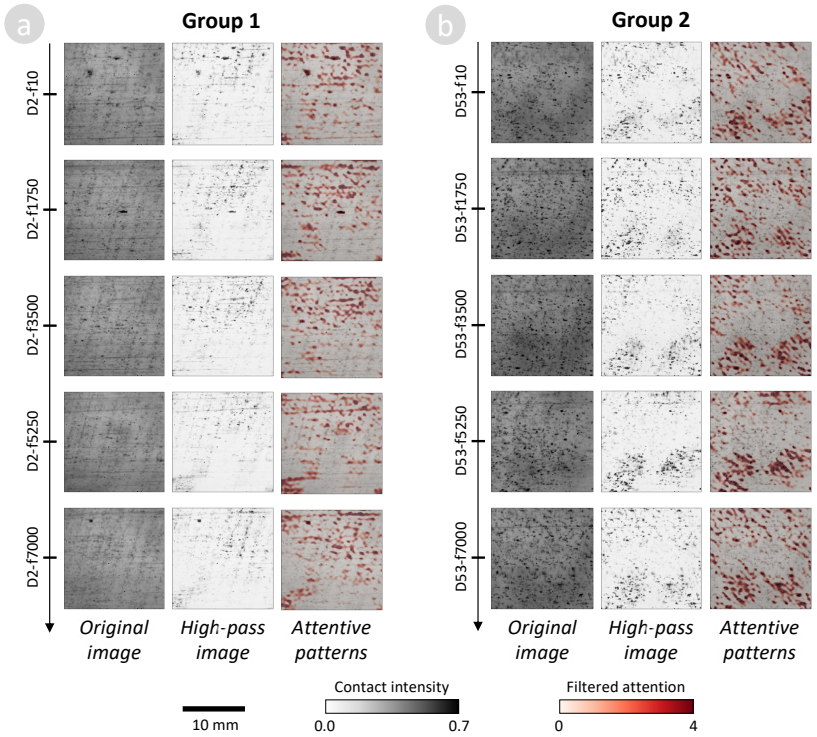

Figure 4.12: Representative examples showing the original contact images, high-pass filtered forms, and attention-fused overlays for sequences categorized in each group.

suggestive material-side implication: formulations that naturally support such coherently organized textures—through the interplay of viscoelastic compliance, energy dissipation at sliding frequencies, and microstructural surface characteristics—tend to exhibit stronger shear engagement under lubrication. While the present analysis does not prescribe specific compound designs, it highlights how the organization of wet-contact textures offers an informative lens for understanding how formulation-dependent properties relate to wet-friction performance.

4.4 Discussion

The results from both case studies demonstrate that the attention-based object-inquiry framework provides a systematic route for articulating dynamic behavior from time-resolved image sequences, even when the underlying processes involve heterogeneous mechanisms and complex structure–property couplings. Across the nanoscale Au–NC junction and the

macroscale wet rubber–surface contact, attention learned through supervised prediction consistently concentrates on structural regions known to play a dominant role in the evolution of measured properties. This consistency establishes attention not merely as a visualization artifact, but as a representational mechanism through which relations between localized structural variation and global response are organized and made accessible for analysis.

A central outcome of this work is the clarification of how dynamic behavior can be articulated from attention-weighted representations. In both systems, attention highlights localized structural variations that co-vary with changes in experimentally measured properties—atomic rearrangements within the nanocontact constriction during Au–NC deformation, and transient drainage- and microcontact-driven fluctuations near the edge of the wet-contact interface. These regions represent the parts of the observation whose variation is most strongly coupled to the response. Rather than treating these variations as isolated events, the framework organizes them into structured patterns across space and time, enabling behavior to be characterized in terms of recurring modes of organization.

Equally important is the identification of persistent organizational tendencies that recur across time, actuation conditions, and experimental realizations. These tendencies appear as stable orientation patterns, spatial coherence, or characteristic modes of structural arrangement emphasized by attention. In the Au–NC system, they manifest as dominant crystallographic deformation pathways, such as single-system slip, coherent lattice rotation, or multi-orientation rearrangement, each associated with distinct electromechanical responses. In the wet-contact system, they emerge as coherent diagonal textures or bulk-like accumulations that persist despite strong fluctuations arising from lubrication and microcontact dynamics. These organizational modes delimit how structural variation is expressed during evolution and provide a basis for interpreting how material behavior unfolds under external driving.

Through this perspective, the attention-based framework supports an interpretation of deep learning as a tool for organizing relations within complex observations rather than merely predicting outcomes. A model trained to associate images with measured properties must internalize how structural components are related to one another and to the response. Attention fields, together with their spectral and statistical characterization, provide a direct window into this internal organization. By examining how relevance is distributed, persists, and reorganizes, the framework enables interpretation of how structure, dynamics, and measurement are coupled within the data.

The two case studies highlight complementary aspects of this capabil-

ity. The Au–NC case demonstrates sensitivity to atomic-scale organization in a setting where crystallographic symmetry offers a clear reference for interpreting deformation behavior. The wet-contact case illustrates how meaningful organizational patterns can be extracted in a system lacking persistent geometric reference, showing that coherent interfacial textures can be articulated even when spatial organization appears transient or fragmented. Together, these examples underscore the generality of the approach across scales, material classes, and levels of structural determinacy.

More broadly, the results suggest that attention-based representations, when embedded within an object-inquiry framework, enable data-driven models to contribute directly to scientific understanding of dynamic phenomena. By organizing observations into relevance-weighted representations and articulating behavior from their relational structure, the framework transforms predictive models into analytical instruments. This approach complements traditional physical reasoning by revealing recurring organizational patterns, highlighting dominant modes of structural variation, and suggesting pathways through which material behavior emerges from complex observational data.

In summary, the attention-based object-inquiry framework provides a unified methodology for studying dynamic material and interfacial systems under dense but structurally complex observation. By leveraging attention to expose relations between localized structural variation and measured response, it extends deep learning beyond prediction toward systematic articulation of dynamic behavior. This foundation opens avenues for future integration with physics-informed modeling, simulation-based analysis, and experimental design aimed at probing the organizational principles governing material dynamics.

Chapter 5

Discussions and Conclusions

Understanding dynamic behavior through data-driven modeling requires more than reproducing observed transformations; it requires careful clarification of how representations organize observations, expose relations, and support the articulation of behavior. In revising the concluding chapter, the emphasis is therefore shifted from interpretive or inferential claims toward a synthesis of how representational structures function across the dissertation. Deep learning is treated consistently as a representational environment rather than a predictive or explanatory endpoint, and the revision aims to consolidate how different representational mechanisms contribute to object inquiry without introducing new methodological claims. The purpose of this chapter is thus refined to reflect, integrate, and contextualize the representational logic developed throughout the thesis.

The preceding chapters developed this perspective through two complementary representational realizations. Generative representations were examined as a means of organizing sparse, time-resolved observations into structured spaces of variation, enabling dynamic behavior to be articulated from proximity and continuity within those spaces. Attention-based representations were explored as a means of organizing dense observations by relevance, exposing correlations between localized structural variation and measured response and allowing behavior to be articulated from patterns of relational emphasis across space and time. Together, these approaches demonstrate how different observational regimes call for different representational emphases, while adhering to a common logic of object inquiry.

This final chapter reflects on what these developments collectively establish, the boundaries within which the current formulations operate, and directions for extending the representational framework. The first section, *Learning-Based Representations for Object Inquiry into Material Dynamics*, synthesizes how generative and attention-based representations support object inquiry by organizing observations and articulating behavior through relational structure. The second section, *Scope and Boundaries of Representation-Based Object Inquiry*, examines the conceptual and practical limits encountered in the present implementations, including the alignment

between representation geometry and physical dynamics and the temporal scope over which behavior can be articulated. The final section, *Future Directions for Extending Representations for Object-Inquiry*, outlines prospective developments aimed at richer representations of temporal organization, integration with physical structure, and broader applicability across dynamic systems.

Rather than serving as a summary of prior results, this chapter positions the dissertation within a broader perspective on representation-based scientific inquiry. It clarifies how learning-based representations can function as analytical instruments for organizing complex observations and articulating dynamic behavior, and it situates the object-inquiry framework as a disciplined approach for engaging with complexity in contemporary materials research and beyond.

5.1 Learning-Based Representations for Object Inquiry into Material Dynamics

This dissertation approaches the study of material dynamics through an object-inquiry perspective, framing the problem as one of organizing observations into representations that expose relations from which dynamic behavior can be articulated. Within this perspective, deep learning is treated not as a predictive mechanism or explanatory surrogate, but as a representational environment capable of structuring variation, relevance, and relational organization in complex data. This orientation also reflects a broader methodological difficulty in scientific inquiry: meaningful representations are required to organize observations and reveal relations that characterize physical behavior, yet constructing such representations often presupposes prior understanding of the phenomena under investigation. The representational strategies developed in this dissertation may therefore be viewed as providing a practical warm-start for inquiry in settings where such understanding remains incomplete. By learning representations directly from observational data, deep-learning models enable provisional organization of complex variability, allowing relations among observations to be examined even before comprehensive theoretical descriptions of the underlying processes are established. Within this framework, two distinct yet conceptually aligned modeling realizations—generative representations and attention-based representations—are developed not as competing methodologies, but as complementary means of examining how different representational structures organize observations and support the articulation of behavior in dynamic physical systems.

The generative representation framework operates at the level of organizing continuity and variation across configurations. By training deep generative models on static or sparsely sampled observations and examining their latent spaces through controlled sampling, this framework reveals how observed configurations are arranged according to similarity, proximity, and gradual variation. The resulting latent structures—smooth interpolations, directional spreads, and localized morphological adjustments—do not assert mechanistic explanations of physical change. Rather, they express how the model organizes possible configurations consistent with the observed data, enabling dynamic behavior to be articulated from relationships encoded in the geometry of the representation space. In this sense, generative representations provide a means of examining how short-range variation and continuity are structured across time-resolved material observations, offering an exploratory view of how possible configurations may evolve within the limits of the observed data distribution.

The attention-based representation framework, by contrast, operates at the level of relational emphasis within dense observational data. Transformer-based architectures compute attention fields that encode correlations among decomposed elements of the input, such as spatial regions or temporal segments. When applied to dynamic imaging data, these attention fields organize observations according to their relevance to measured responses, producing spatial and temporal patterns that highlight structurally significant regions. The resulting organizations—localized concentration along interfaces, distributed emphasis across deforming zones, or persistent orientation-aligned regions—do not reconstruct temporal transitions. Instead, they articulate how relations among localized variations are structured within the representation, enabling dynamic behavior to be examined through patterns of relevance and correlation across observations.

Viewed together, the two representational frameworks demonstrate that object inquiry does not rely on a single representational form, but can be realized through multiple modes of organization shaped by model architecture, representational capacity, and the nature of the observational data. Generative models articulate behavior through continuity and variation embedded in latent spaces, while attention-based models articulate behavior through relational weighting distributed across structural components. Both approaches adhere to the same inquiry logic: observations are organized into representations, relations are exposed through representational structure, and dynamic behavior is articulated from these relations. Their differences lie not in conceptual orientation, but in the kinds of organizational structures they make available—continuous manifolds in one case, relevance-weighted relational fields in the other.

Through this integrated perspective, deep learning emerges as a flexible representational environment for object inquiry into material dynamics. Rather than resolving the representation–understanding paradox inherent in scientific investigation, the approaches developed here demonstrate how learning-based representations can help navigate it by enabling representations and interpretations to co-evolve as inquiry progresses. By examining how representations organize observations and encode relations, the dissertation illustrates how learned models can serve as exploratory instruments for advancing understanding of dynamic material systems beyond what is directly visible in individual observations. The comparison between generative and attention-based representations further clarifies that learning-based models support multiple, compatible modes of inquiry, each offering a distinct lens on how material behavior unfolds across time, scale, and experimental context.

5.2 Scope and Boundaries of Representation-Based Object Inquiry

The representation-based formulation developed in this dissertation positions deep learning models as environments for organizing observations, exposing relations, and articulating dynamic behavior in material systems. While the generative and attention-based frameworks examined here provide concrete realizations of how such organization can be achieved, several conceptual and methodological boundaries remain. These boundaries do not reflect shortcomings in the empirical demonstrations presented, but rather delineate the scope within which the current representational approach has been developed and the points at which further theoretical and methodological refinement would be required.

A first boundary concerns the relationship between the geometry of learned representation spaces and the characteristics of the underlying physical dynamics. Different modeling architectures produce representations with distinct organizational properties: some emphasize smooth, continuous variation, others admit fragmented or highly anisotropic structure; some distribute variation across many dimensions, while others concentrate it along a small number of dominant axes. Physical systems, in turn, exhibit diverse modes of dynamic behavior, ranging from localized diffusion to extended interfacial propagation and mechanically driven deformation. The present work demonstrates how representational organization can meaningfully articulate short-range variation and relational structure in several concrete systems. However, the validity of these representations is typically established only in

regions of latent space that are well supported by observed data. Outside such regions, particularly in high-dimensional representations where extrapolated configurations may arise, there is generally no guarantee that the relations encoded in the representation correspond to physically meaningful states of the material system. In this sense, the meaningfulness of representation geometry may remain localized around observed configurations, while more distant regions of the latent space may admit mathematically plausible but physically invalid arrangements. Developing a principled account of how architectural bias, data composition, observational modality, and physical processes jointly constrain the geometry of representation spaces remains an open challenge. The absence of such a unifying account marks a conceptual boundary of the present study: while representational organization can be interpreted meaningfully within each examined setting, its broader alignment with classes of physical dynamics across high-dimensional representation spaces is not yet established.

A related boundary concerns the interpretability and physical grounding of relational structures identified by attention-based models. Transformer architectures organize observations through attention fields that encode correlations among decomposed elements of the input. These correlations highlight regions or components of observations that contribute strongly to measured responses, enabling localized patterns of structural relevance to be identified. However, the mechanisms by which attention weights are formed and propagated across layers remain largely opaque. As a result, although attention patterns may correlate with physically significant regions or structural changes, they do not in themselves guarantee that the identified relations correspond to physically meaningful mechanisms. The relational structures produced by attention-based models therefore provide suggestive but not definitive evidence regarding the physical processes governing the observed behavior. Bridging this gap between representational relevance and physical interpretability remains an important methodological challenge.

A second boundary concerns the temporal scope over which dynamic behavior is articulated. The object-inquiry framework itself is not inherently limited to short temporal ranges; in principle, behavior may be organized and examined across extended trajectories, hierarchical temporal scales, or cumulative structural evolution. The implementations explored in this dissertation, however, focus on foundational settings in which behavior can be articulated through short-range variation in generative representations or localized relevance patterns in attention-based representations. This focus reflects a deliberate prioritization of clarity over generality, establishing how behavior emerges from representational organization in its simplest forms before addressing more complex temporal dependencies. As a result, the

current modeling schemes do not explicitly encode long-range temporal structure, multi-step accumulation, or hierarchical temporal organization. Extending the framework to address such phenomena would require additional representational mechanisms designed to preserve and organize temporal dependencies across longer horizons.

Together, these boundaries identify the present limits of representation-based object inquiry as developed in this dissertation. They clarify where the current formulations are intentionally restricted and where future work is needed to broaden the scope of representational alignment with physical dynamics, interpretability, and temporal organization. By making these boundaries explicit, the discussion situates the present contributions as foundational steps toward a more comprehensive framework for articulating dynamic behavior from complex observational data.

5.3 Future Directions for Extending Representations for Object Inquiry

The representation-based perspective developed in this dissertation suggests a broader trajectory in which the study of dynamic physical systems is increasingly grounded in how computational models organize observations and expose relations among them. From the standpoint of object inquiry, the central requirement for future representational approaches is not adherence to a specific model class, but the ability to encode relations in forms that can be interpreted and examined. Whether realized through deep learning or other representation-learning paradigms, representations that support object inquiry must make relational structure explicit enough to articulate dynamic behavior from complex observations. The present work has demonstrated how generative and attention-based representations can provide an effective warm-start for such inquiry, enabling provisional organization of observational variability in settings where prior knowledge of the governing mechanisms remains incomplete. By structuring observations into latent environments where relations become examinable, learning-based models allow understanding to begin emerging even before fully grounded physical descriptions are available.

A natural continuation of this trajectory lies in reintegrating the knowledge extracted from learned representations back into the modeling process itself. As relations revealed by generative and attention-based representations are interpreted and associated with physical mechanisms, this knowledge may guide the construction of improved representations through physically informed initialization, architectural constraints, or hybrid mod-

eling strategies. In this sense, the role of deep learning representations evolves from exploratory organization toward knowledge-informed modeling, where representation spaces are shaped not only by observational data but also by accumulated scientific understanding. Such integration may allow future models to preserve the flexibility of data-driven representation learning while progressively embedding the structural principles governing the systems under investigation.

A related direction concerns the construction of representation spaces whose geometry more faithfully reflects physically meaningful variation. As discussed in the previous section, learned representations may organize observations coherently in regions corresponding to observed material states, yet the validity of relations across the entire high-dimensional representation space cannot be assumed. Future work may therefore focus on developing representation-learning approaches in which physically admissible configurations, conservation relations, or structural constraints are incorporated directly into the representation construction process. Rather than treating latent spaces as purely statistical embeddings, such approaches would aim to ensure that trajectories through representation space correspond more closely to physically realizable states and transformations.

Another important direction concerns the representation of temporal organization. While the present work focuses on short-range variation and localized relational emphasis, future representations for object inquiry may encode evolution at the level of trajectories rather than isolated observations or short segments. Representation-learning models that organize sequences as continuous paths in a structured space—rather than as collections of frames—would enable behavior to be articulated from how trajectories unfold, branch, or recur. Importantly, this does not require explicit prediction of future states, but rather representations that preserve temporal relations in a way that supports interpretive examination. Such models would differ from conventional sequence-generation approaches by prioritizing relational coherence and interpretability over perceptual realism.

Beyond deep generative and attention-based models, a wide range of representation-learning approaches may contribute to object inquiry, provided they expose interpretable relations among observations. Metric learning, contrastive representation learning, manifold learning, and graph-based representations all offer mechanisms for organizing similarity, dependence, or interaction structure in ways that may be amenable to inquiry. Future work may also explore hybrid or domain-adapted representations that integrate physical priors, experimental context, or multi-modal observations while maintaining interpretability of the resulting relations. In particular, improving the interpretability of attention-based representations—so that relations

highlighted by attention mechanisms can be more directly associated with physical interactions—remains an important methodological goal.

More broadly, the extension of object-inquiry representations invites reflection on how computational representations relate to scientific reasoning itself. By emphasizing the organization of observations and the articulation of behavior from relational structure, the framework aligns with longstanding practices in science in which understanding emerges from identifying patterns, regularities, and modes of organization rather than from direct reconstruction of underlying mechanisms. As representation-learning models continue to evolve, they may increasingly serve as environments in which such organizational structures can be explored systematically. In this sense, the future development of representations for object inquiry is not limited to materials research, but contributes to a wider effort to integrate data-driven modeling with interpretive scientific practices.

Appendix A

Supplementary Materials of Generative Framework

A.1 PG-GAN Architecture Specifications

To model the distribution of material configurations across datasets with varying resolution and dimensionality, we employed a Progressive Growing GAN (PG-GAN) architecture, following the WGAN-GP formulation. The PG-GAN approach enables stable training by incrementally expanding both the generator and discriminator networks, starting from low-resolution representations and growing toward the target resolution. Each progressive stage introduces additional layers: the generator performs upsampling followed by 3×3 convolutions, while the discriminator applies 3×3 convolutions followed by average pooling for downsampling. This symmetric design ensures consistency in feature hierarchy across network depths.

Our implementation customized the number of convolutional filters at each resolution stage to accommodate the complexity of the target datasets. For two-dimensional data such as the tantalum test chart and gold nanoparticle diffusion, higher filter counts were assigned at coarser levels to capture global structural trends, with a gradual reduction as the resolution increased. For three-dimensional data—specifically, the aging brass clump dataset—this reduction was omitted; a consistently high filter count was retained across stages due to the increased representational demands and limited number of progressive stages feasible for small-volume inputs.

Latent vectors were sampled from a standard Gaussian distribution $\mathcal{N}(0, \mathbf{I})$ with a fixed dimensionality $d = 32$ for all datasets. Supplementary Table A.1 summarizes the generator configuration in terms of convolutional layers and filter counts, which are mirrored in the discriminator design.

Training followed the Wasserstein GAN loss with gradient penalty ($\lambda = 10.0$), ensuring Lipschitz continuity. Optimization used the Adam optimizer with learning rate 0.001, $\beta_1 = 0.0$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. All convolutional layers were followed by LeakyReLU activations with negative slope 0.2.

The progressive training strategy advanced through resolution stages,

Table A.1: Architecture configuration for progressive growing GAN model

Dataset	**# Conv**	**# Filters**
Ta test chart	7	(512, 512, 512, 512, 256, 128, 64)
NP diffusion	6	(256, 256, 256, 256, 128, 64)
Aging brass clumps	3	(256, 256, 256)

Table A.2: Training configuration of progressive training strategy for GAN model

Dataset	**# Epochs**	**# Batch size**
Ta test chart	(32, 16, 16, 16, 16, 32, 64)	(32, 32, 32, 32, 32, 16, 8)
NP diffusion	(32, 16, 16, 16, 32, 64)	(16, 16, 16, 16, 8, 4)
Aging brass clumps	(16, 8, 16, 32)	(8, 8, 4, 2)

each associated with a fixed number of epochs. For example, the tantalum dataset was trained over seven stages with decreasing epoch counts in intermediate layers and increased epochs at both extremes to stabilize feature learning. Batch sizes were adjusted per stage to balance GPU memory constraints and training stability, particularly in the 3D case where volumetric scaling imposed stricter memory limits. The full configuration of training epochs and batch sizes across datasets is provided in Supplementary Table A.2.

A.2 Standardization of Sampling Margin

Following GAN model training, we evaluated whether proximity in the latent space corresponds to physical similarity between material states. However, the interpretation of proximity in high-dimensional latent spaces is inherently nontrivial. In such spaces, the concept of distance can become unintuitive due to phenomena related to the "concentration of measure" [137,138]. As dimensionality increases, Euclidean distances between randomly sampled points tend to concentrate around a narrow range, diminishing the contrast between "near" and "far." Consequently, a fixed perturbation magnitude (e.g., γ) does not have an interpretable or consistent effect across latent spaces of different dimensions.

To mitigate this ambiguity and permit a controlled exploration of latent space, we introduced a standardized notion of sampling margin. The latent space is modeled as a d-dimensional isotropic Gaussian $\mathcal{N}(0, \mathbf{I})$ with $d = 32$. Around any reference point x_i, perturbed samples $x_{i,j}$ are generated

via additive noise vectors ϵ_j, each drawn component-wise from a uniform distribution $\mathcal{U}_{[-\gamma,\gamma]}$. To assess the effective scale of perturbation, we compute the empirical average Euclidean distance between perturbed and original vectors, denoted $\mathrm{E}[\|x_i - x_{i,j}\|]$.

To make this measure more interpretable across dimensionalities, we normalize the displacement by the number of latent dimensions:

$$\text{Sampling Margin per Dimension} = \frac{\mathrm{E}[\|x_i - x_{i,j}\|]}{d} \tag{A.1}$$

This value provides a dimension-agnostic metric that approximates the average displacement applied to each latent coordinate. It offers a standardized interpretation of perturbation intensity, facilitating direct comparisons across latent spaces and aiding in the calibration of local sampling procedures. In particular, this is critical for Monte Carlo-based trajectory generation, where controlled perturbations must preserve local coherence while enabling exploration. Throughout the experiments, both raw γ values and normalized margins were reported to ensure interpretability, robustness, and reproducibility of the inferred transformations.

Appendix B

Supplementary Materials of Attentive Framework

B.1 Transformer-ViT Model Architecture and Training Process

The correlation between temporal image sequences and physical measurements was modeled using an adapted TimeSformer architecture (t-ViT) [75], which was modified for regression from grayscale microscopy data. Each input sample consisted of $T_w = 5$ consecutive grayscale frames. Images were first decomposed into non-overlapping spatial patches and linearly projected into a 240-dimensional feature space. To capture multiscale spatial structure, four decomposition scales were used: 5×5, 7×7, 11×11, and 13×13 patches per image.

The model employed divided space-time attention, whereby spatial attention is computed independently within each image and temporal attention is applied across matching patch positions over time. Each transformer block includes a feed-forward MLP with hidden dimension 960 (four times the embedding size), GELU activation, residual connections, and LayerNorm with $\epsilon = 10^{-6}$. The class token output is mapped through a linear head to two regression targets.

In the Au–NC deformation task, the model predicted stiffness ($\hat{k}$) and conductance ($\hat{G}$). Input frames were cropped to 385×385 pixels; to satisfy patch divisibility for 13×13 decomposition, zero-padding extended the size to 390×390 pixels. Padding was used selectively to preserve original resolution and avoid interpolation effects. Because the Au–NC is distinctly segmented from the background, padding introduced negligible representational bias.

For tire hydroplaning, the model predicted slip rate ($\hat{S}$) and friction coefficient ($\hat{\mu}$). Contact images were cropped to 500×500 pixels and resized according to patch configuration: 275×275 for 5×5, 280×280 for 7×7, 275×275 for 11×11, and 286×286 for 13×13. Unlike the Au–NC task, padding was avoided in favor of direct resizing to prevent edge artifacts,

which could distort attention weights in complex textured domains.

Training sought both accurate regression performance and stable attention map formation. All models were trained for 200 epochs using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) with a weight decay of 10^{-4} [139]. Learning rates were set to 10^{-3} for 5×5 patches and 5×10^{-4} for all others. The Au–NC dataset used a batch size of 64; the hydroplaning dataset used 16 due to its larger image scale and volume.

GPU memory consumption ranged from 2.68 GB for 5×5 to 20.20 GB for 13×13 models. For the Au–NC case, training times ranged from 2.5 to 7 hours, depending on patch size; for hydroplaning, training ranged from 2 to 6 days per model. Inference time per instance (five-frame sequence) ranged from 0.6–1.4 seconds for Au–NC and 2.5–5.6 seconds for hydroplaning, depending on input size and model configuration.

B.2 Pattern Filtering and Fusion

To extract spatial orientations emphasized by attention profiles, a Fourier-based filtering approach was applied and fused with the original images. This technique preserved structural fidelity while accentuating the orientations identified as significant by the attention mechanism.

For each attention group g, we defined a group-level descriptor by averaging the angular profiles of all members:

$$\bar{\mathbf{p}}_g = \frac{1}{N_g} \sum_{j=1}^{N_g} \mathbf{p}_j, \tag{B.1}$$

where N_g is the number of samples in group g. The descriptor $\bar{\mathbf{p}}_g$ thus captures dominant orientations consistently emphasized across the group.

Given an image $I_k(x, y)$ corresponding to group g, its Fourier representation in polar coordinates (r, θ) is:

$$\mathcal{F}[I_k](r, \theta) = \widetilde{\mathbf{M}}_k(r, \theta) e^{i\Phi_k(r,\theta)}, \tag{B.2}$$

with $\widetilde{\mathbf{M}}_k$ and Φ_k representing amplitude and phase, respectively. Orientation-specific filtering was performed by reweighting the amplitude using the group descriptor:

$$\widetilde{\mathbf{M}}_{g,k}(r, \theta) = \bar{p}_g(\theta) \widetilde{\mathbf{M}}_k(r, \theta), \tag{B.3}$$

$$F_{g,k}(r, \theta) = \widetilde{\mathbf{M}}_{g,k}(r, \theta) e^{i\Phi_k(r,\theta)}. \tag{B.4}$$

The filtered image in spatial domain is obtained by inverse transform:

$$I_{g,k}(x, y) = \mathcal{F}^{-1}\left[F_{g,k}(r, \theta)\right]. \tag{B.5}$$

This result was normalized using a sigmoid function:

$$W_{g,k}(x,y) = \frac{1}{1+\exp[-k(I_{g,k}(x,y)-x_0)]}, \tag{B.6}$$

where x_0 is the mean of $I_{g,k}$ and k is scaled relative to its variance. This adaptive normalization enhances contrast between emphasized and suppressed regions.

The final filtered image is computed via element-wise multiplication:

$$I_{g,k}^{\text{filtered}}(x,y) = I_k(x,y) \odot W_{g,k}(x,y), \tag{B.7}$$

where $I_{g,k}^{\text{filtered}}$ reveals patterns highlighted by the group descriptor while retaining base image fidelity.

Publications

1. Journal Papers (Refereed)

- **Duc-Anh Dao**, Minh-Quyet Ha, Tien-Sinh Vu, Shuntaro Takazawa, Nozomu Ishiguro, Yukio Takahashi, Masato Suzuki, Takashi Kakubo, Naoya Amino, Hirosuke Matsui, *et al.* (2025). *Material dynamics analysis with deep generative model. Digital Discovery.* Royal Society of Chemistry.
- Shuntaro Takazawa, **Duc-Anh Dao**, Masaki Abe, Hideshi Uematsu, Nozomu Ishiguro, Taiki Hoshino, Hieu Chi Dam, & Yukio Takahashi (2023). *Coupling x-ray photon correlation spectroscopy and dynamic coherent x-ray diffraction imaging: Particle motion analysis from nano-to-micrometer scale. Physical Review Research*, **5**(4), L042019. APS.
- Tien-Sinh Vu, Minh-Quyet Ha, Adam Mukharil Bachtiar, **Duc-Anh Dao**, Truyen Tran, Hiori Kino, Shuntaro Takazawa, Nozomu Ishiguro, Yuhei Sasaki, Masaki Abe, *et al.* (2025). *PID3Net: a deep learning approach for single-shot coherent X-ray diffraction imaging of dynamic phenomena. npj Computational Materials*, **11**(1), 66. Nature Publishing Group.

2. International Conferences (Oral Presentations)

- **Duc-Anh Dao**, Shuntaro Takazawa, Masaki Abe, Hideshi Uematsu, Nozomu Ishiguro, Taiki Hoshino, Yukio Takahashi, & Hieu-Chi Dam. *A framework to investigate motion behaviors of particles in heterogeneous media. International Conference on Materials for Advanced Technologies (ICMAT)*, June 2023.
- **Duc-Anh Dao**, Shuntaro Takazawa, Masaki Abe, Hideshi Uematsu, Nozomu Ishiguro, Taiki Hoshino, Yukio Takahashi, & Hieu-Chi Dam. *A framework to investigate motion behaviors of particles in heterogeneous media. MRM Conference*, December 2023.

References

[1] A. Mohammed and A. Abdullah, "Scanning electron microscopy (sem): A review," in *Proceedings of the 2018 International Conference on Hydraulics and Pneumatics—HERVEX, Băile Govora, Romania*, vol. 2018, 2018, pp. 7–9.

[2] S. Takazawa, J. Kang, M. Abe, H. Uematsu, N. Ishiguro, and Y. Takahashi, "Demonstration of single-frame coherent x-ray diffraction imaging using triangular aperture: Towards dynamic nanoimaging of extended objects," *Optics Express*, vol. 29, no. 10, pp. 14 394–14 402, 2021.

[3] W. L. Bragg, "The diffraction of short electromagnetic waves by a crystal," *Scientia*, vol. 23, no. 45, 1929.

[4] A. Damascelli, Z. Hussain, and Z.-X. Shen, "Angle-resolved photoemission studies of the cuprate superconductors," *Reviews of modern physics*, vol. 75, no. 2, p. 473, 2003.

[5] A.-A. Liu, K. Li, and T. Kanade, "A semi-markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations," *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 359–369, 2011.

[6] W. J. Godinez, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences," *Medical image analysis*, vol. 13, no. 2, pp. 325–342, 2009.

[7] H.-v. Hoegen *et al.*, "Structural dynamics at surfaces by ultrafast reflection high-energy electron diffraction," *Structural Dynamics*, vol. 11, no. 2, 2024.

[8] X. Chen, X. Zhou, F. De Geuser, A. K. da Silva, H. Zhao, E. Woods, C. Liu, D. Ponge, B. Gault, and D. Raabe, "Atom probe tomography-assisted kinetic assessment of spinodal decomposition in an al-12.5 at.% zn alloy," *Acta Materialia*, vol. 268, p. 119757, 2024.

[9] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011.

[10] M. J. Cherukara, Y. S. Nashed, and R. J. Harder, "Real-time coherent diffraction inversion using deep generative networks," *Scientific reports*, vol. 8, no. 1, p. 16520, 2018.

[11] S. Takazawa, K. Ninomiya, M.-Q. Ha, T.-S. Vu, Y. Sasaki, M. Abe, H. Uematsu, N. Okawa, N. Ishiguro, K. Ozaki *et al.*, "Spatiotemporal mapping of alloy mesostructure dynamics via multimodal coherent x-ray diffraction imaging," *Proceedings of the National Academy of Sciences*, vol. 122, no. 38, p. e2513369122, 2025.

[12] V. Oommen, K. Shukla, S. Goswami, R. Dingreville, and G. E. Karniadakis, "Learning two-phase microstructure evolution using neural operators and autoencoder architectures," *npj Computational Materials*, vol. 8, no. 1, p. 190, 2022.

[13] S. Kondo, T. Mitsuma, N. Shibata, and Y. Ikuhara, "Direct observation of individual dislocation interaction processes with grain boundaries," *Science advances*, vol. 2, no. 11, p. e1501926, 2016.

[14] R. Brunelli, *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[15] S. V. Kalinin, A. R. Lupini, O. Dyck, S. Jesse, M. Ziatdinov, and R. K. Vasudevan, "Lab on a beam—big data and artificial intelligence in scanning transmission electron microscopy," *MRS Bulletin*, vol. 44, no. 7, pp. 565–575, 2019.

[16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[17] S. Lee, H. Park, J. Y. Kim, J. Kim, M.-J. Choi, S. Han, S. Kim, W. Kim, H. W. Jang, J. Park *et al.*, "Unveiling crystal orientation-dependent interface property in composite cathodes for solid-state batteries by in situ microscopic probe," *Nature Communications*, vol. 15, no. 1, p. 7947, 2024.

[18] T. Liu, L. Liang, D. Raabe, and L. Dai, "The martensitic transition pathway in steel," *Journal of Materials Science & Technology*, vol. 134, pp. 244–253, 2023.

[19] M. Karami, Z. Zhu, Z. Zeng, N. Tamura, Y. Yang, and X. Chen, "Two-tier compatibility of superelastic bicrystal micropillar at grain boundary," *Nano letters*, vol. 20, no. 11, pp. 8332–8338, 2020.

[20] C. Yang, B. Zhang, L. Fu, Z. Wang, J. Teng, R. Shao, Z. Wu, X. Chang, J. Ding, L. Wang *et al.*, "Chemical inhomogeneity–induced profuse nanotwinning and phase transformation in aucu nanowires," *Nature Communications*, vol. 14, no. 1, p. 5705, 2023.

[21] X. Li, K. Thornton, Q. Nie, P. Voorhees, and J. S. Lowengrub, "Two- and three-dimensional equilibrium morphology of a misfitting particle and the gibbs–thomson effect," *Acta materialia*, vol. 52, no. 20, pp. 5829–5843, 2004.

[22] S. Li, P. J. Withers, W. Chen, and K. Yan, "Atomic-scale investigation of the mechanisms of deformation-induced martensitic transformation at ultra-cryogenic temperatures," *Journal of Materials Science & Technology*, vol. 210, pp. 138–150, 2025.

[23] H. Matsui, Y. Muramoto, R. Niwa, T. Kakubo, N. Amino, T. Uruga, M.-Q. Ha, D.-T. Dinh, H.-C. Dam, and M. Tada, "Machine learning-derived reaction statistics for 3d spectroimaging of copper sulfidation in heterogeneous rubber/brass composites," *Communications Materials*, vol. 4, no. 1, p. 88, 2023.

[24] S. Wang, R. Du, Y. Guo, S. Sun, and X. Ke, "Structural phase transitions, mechanical and electronic properties of zrse2 under high pressures via the first-principles calculations," *Computational Materials Science*, vol. 226, p. 112214, 2023.

[25] T. Ishida, F. Cleri, K. Kakushima, M. Mita, T. Sato, M. Miyata, N. Itamura, J. Endo, H. Toshiyoshi, N. Sasaki *et al.*, "Exceptional plasticity of silicon nanobridges," *Nanotechnology*, vol. 22, no. 35, p. 355704, 2011.

[26] K. Ishizuka, M. Tomitori, T. Arai, and Y. Oshima, "Mechanical analysis of gold nanocontacts during stretching using an in-situ transmission electron microscope equipped with a force sensor," *Applied Physics Express*, vol. 13, no. 2, p. 025001, 2020.

[27] H. Matsui, N. Ishiguro, Y. Tan, N. Maejima, Y. Muramoto, T. Uruga, K. Higashi, D.-N. Nguyen, H.-C. Dam, G. Samjeské *et al.*, "Variation

of local structure and reactivity of pt/c catalyst for accelerated degradation test of polymer electrolyte fuel cell visualized by operando 3d ct-xafs imaging," *ChemNanoMat*, vol. 8, no. 4, p. e202200008, 2022.

[28] Y. Wang, H. Zhao, C. Liu, Y. Ootani, N. Ozawa, and M. Kubo, "Mechanisms of chemical-reaction-induced tensile deformation of an fe/ni/cr alloy revealed by reactive atomistic simulations," *RSC advances*, vol. 13, no. 10, pp. 6630–6636, 2023.

[29] R. Noyori, "Chiral metal complexes as discriminating molecular catalysts," *Science*, vol. 248, no. 4960, pp. 1194–1199, 1990.

[30] A. Barty, S. Boutet, M. J. Bogan, S. Hau-Riege, S. Marchesini, K. Sokolowski-Tinten, N. Stojanovic, R. Tobey, H. Ehrke, A. Cavalleri *et al.*, "Ultrafast single-shot diffraction imaging of nanoscale dynamics," *Nature photonics*, vol. 2, no. 7, pp. 415–419, 2008.

[31] D. B. Williams, C. B. Carter, D. B. Williams, and C. B. Carter, *The transmission electron microscope*. Springer, 1996.

[32] J. Kang, S. Takazawa, N. Ishiguro, and Y. Takahashi, "Single-frame coherent diffraction imaging of extended objects using triangular aperture," *Optics Express*, vol. 29, no. 2, pp. 1441–1453, 2021.

[33] S. Takazawa, D.-A. Dao, M. Abe, H. Uematsu, N. Ishiguro, T. Hoshino, H. C. Dam, and Y. Takahashi, "Coupling x-ray photon correlation spectroscopy and dynamic coherent x-ray diffraction imaging: Particle motion analysis from nano-to-micrometer scale," *Physical Review Research*, vol. 5, no. 4, p. L042019, 2023.

[34] T. Uruga, M. Tada, O. Sekizawa, Y. Takagi, T. Yokoyama, and Y. Iwasawa, "Spring-8 bl36xu: Synchrotron radiation x-ray-based multi-analytical beamline for polymer electrolyte fuel cells under operating conditions," *The Chemical Record*, vol. 19, no. 7, pp. 1444–1456, 2019.

[35] L. Huang, C. Wong, and E. Grumstrup, "Time-resolved microscopy: a new frontier in physical chemistry," pp. 5997–5998, 2020.

[36] S. V. Kalinin, E. Strelcov, A. Belianinov, S. Somnath, R. K. Vasudevan, E. J. Lingerfelt, R. K. Archibald, C. Chen, R. Proksch, N. Laanait *et al.*, "Big, deep, and smart data in scanning probe microscopy," 2016.

[37] S. V. Kalinin, O. Dyck, S. Jesse, and M. Ziatdinov, "Exploring order parameters and dynamic processes in disordered systems via variational autoencoders," *Science Advances*, vol. 7, no. 17, p. eabd5084, 2021.

[38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[39] T. L. Pham, H. Kino, K. Terakura, T. Miyake, K. Tsuda, I. Takigawa, and H. C. Dam, "Machine learning reveals orbital interaction in materials," *Science and technology of advanced materials*, vol. 18, no. 1, p. 756, 2017.

[40] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, no. 7715, pp. 547–555, 2018.

[41] J. Schmidt, M. R. Marques, S. Botti, and M. A. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj computational materials*, vol. 5, no. 1, p. 83, 2019.

[42] D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge, and P. W. Chung, "Applying machine learning techniques to predict the properties of energetic materials," *Scientific reports*, vol. 8, no. 1, p. 9059, 2018.

[43] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, p. 139–144, Oct. 2020.

[44] I. Goodfellow, "Deep learning," 2016.

[45] A. R. Durmaz, M. Müller, B. Lei, A. Thomas, D. Britz, E. A. Holm, C. Eberl, F. Mücklich, and P. Gumbsch, "A deep learning approach for complex microstructure inference," *Nature communications*, vol. 12, no. 1, p. 6272, 2021.

[46] K. P. Murphy, *Probabilistic machine learning: an introduction*. MIT press, 2022.

[47] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, p. 307–392, Nov. 2019.

[48] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[50] "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015.

[51] K. Yang, Y. Cao, Y. Zhang, S. Fan, M. Tang, D. Aberg, B. Sadigh, and F. Zhou, "Self-supervised learning and prediction of microstructure evolution with convolutional recurrent neural networks," *Patterns*, vol. 2, no. 5, 2021.

[52] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, 2016.

[53] M. Ge, F. Su, Z. Zhao, and D. Su, "Deep learning analysis on microscopic imaging in materials science," *Materials Today Nano*, vol. 11, p. 100087, 2020.

[54] T. Kizuka, "Atomic configuration and mechanical and electrical properties of stable gold wires of single-atom width," *Physical Review B—Condensed Matter and Materials Physics*, vol. 77, no. 15, p. 155401, 2008.

[55] J. Zhang, K. Ishizuka, M. Tomitori, T. Arai, K. Hongo, R. Maezono, E. Tosatti, and Y. Oshima, "Peculiar atomic bond nature in platinum monatomic chains," *Nano Letters*, vol. 21, no. 9, pp. 3922–3928, 2021.

[56] D. Morgan and R. Jacobs, "Opportunities and challenges for machine learning in materials science," *Annual Review of Materials Research*, vol. 50, no. 1, pp. 71–103, 2020.

[57] N. Creange, O. Dyck, R. K. Vasudevan, M. Ziatdinov, and S. V. Kalinin, "Towards automating structural discovery in scanning transmission electron microscopy," *Machine Learning: Science and Technology*, vol. 3, no. 1, p. 015024, 2022.

[58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[59] J. M. Ede, "Deep learning in electron microscopy," *Machine Learning: Science and Technology*, vol. 2, no. 1, p. 011004, 2021.

[60] R. Sainju, W.-Y. Chen, S. Schaefer, Q. Yang, C. Ding, M. Li, and Y. Zhu, "Defecttrack: a deep learning-based multi-object tracking algorithm for quantitative defect analysis of in-situ tem videos in real-time," *Scientific reports*, vol. 12, no. 1, p. 15705, 2022.

[61] Y. Hirabayashi, H. Iga, H. Ogawa, S. Tokuta, Y. Shimada, and A. Yamamoto, "Deep learning for three-dimensional segmentation of electron microscopy images of complex ceramic materials," *npj Computational Materials*, vol. 10, no. 1, p. 46, 2024.

[62] R. E. Goodall and A. A. Lee, "Predicting materials properties without crystal structure: deep representation learning from stoichiometry," *Nature communications*, vol. 11, no. 1, p. 6280, 2020.

[63] K. Faraz, T. Grenier, C. Ducottet, and T. Epicier, "Deep learning detection of nanoparticles and multiple object tracking of their dynamic evolution during in situ etem studies," *Scientific reports*, vol. 12, no. 1, p. 2484, 2022.

[64] R. Jiang, J. Smith, Y.-T. Yi, T. Sun, B. J. Simonds, and A. D. Rollett, "Deep learning approaches for instantaneous laser absorptance prediction in additive manufacturing," *npj Computational Materials*, vol. 10, no. 1, p. 6, 2024.

[65] T.-S. Vu, M.-Q. Ha, A. M. Bachtiar, D.-A. Dao, T. Tran, H. Kino, S. Takazawa, N. Ishiguro, Y. Sasaki, M. Abe *et al.*, "Pid3net: a deep learning approach for single-shot coherent x-ray diffraction imaging of dynamic phenomena," *npj Computational Materials*, vol. 11, no. 1, p. 66, 2025.

[66] A. Khan, C.-H. Lee, P. Y. Huang, and B. K. Clark, "Leveraging generative adversarial networks to create realistic scanning transmission electron microscopy images," *npj Computational Materials*, vol. 9, no. 1, p. 85, 2023.

[67] N. Botteghi, M. Guo, and C. Brune, "Deep kernel learning of dynamical models from high-dimensional noisy data," *Scientific reports*, vol. 12, no. 1, p. 21530, 2022.

[68] T.-S. Vu, M.-Q. Ha, A. M. Bachtiar, D.-A. Dao, T. Tran, H. Kino, S. Takazawa, N. Ishiguro, Y. Sasaki, M. Abe, H. Uematsu, N. Okawa,

K. Ozaki, K. Kobayashi, Y. Honjo, H. Nishino, Y. Joti, T. Hatsui, Y. Takahashi, and H.-C. Dam, "Pid3net: a deep learning approach for single-shot coherent x-ray diffraction imaging of dynamic phenomena," *npj Computational Materials*, vol. 11, no. 1, p. 66, 2025.

[69] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1735–1742.

[70] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[71] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.

[72] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[73] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 214–223. [Online]. Available: https://proceedings.mlr.press/v70/arjovsky17a.html

[74] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International conference on machine learning*. PMLR, 2015, pp. 1530–1538.

[75] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *Proceedings of the International Conference on Machine Learning (ICML)*, July 2021.

[76] S. Abnar and W. Zuidema, "Quantifying attention flow in transformers," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 4190–4197.

[77] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[78] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised scalable representation learning for multivariate time series," *Advances in neural information processing systems*, vol. 32, 2019.

[79] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan, "Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design," *npj Computational Materials*, vol. 5, no. 1, p. 21, 2019.

[80] P. Lyngby and K. S. Thygesen, "Data-driven discovery of 2d materials by deep generative models," *npj Computational Materials*, vol. 8, no. 1, p. 232, 2022.

[81] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," *Nature*, vol. 624, no. 7990, pp. 80–85, 2023.

[82] S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen, D. Podareanu, R. Ruiz de Austri, and R. Verheyen, "Event generation and statistical sampling for physics with deep generative models and a density information buffer," *Nature Communications*, vol. 12, no. 1, p. 2985, 2021.

[83] X. Lyu and X. Ren, "Microstructure reconstruction of 2d/3d random materials via diffusion-based deep generative models," *Scientific Reports*, vol. 14, no. 1, p. 5041, 2024.

[84] B. Murgas, J. Stickel, and S. Ghosh, "Generative adversarial network (gan) enabled statistically equivalent virtual microstructures (sevm) for modeling cold spray formed bimodal polycrystals," *npj Computational Materials*, vol. 10, no. 1, p. 32, 2024.

[85] C. X. Hernández, H. K. Wayment-Steele, M. M. Sultan, B. E. Husic, and V. S. Pande, "Variational encoding of complex dynamics," *Physical Review E*, vol. 97, no. 6, p. 062412, 2018.

[86] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American statistical association*, vol. 44, no. 247, pp. 335–341, 1949.

[87] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[88] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[89] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[90] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022.

[91] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *International Conference on Learning Representations*, 2018.

[92] C. Robert and G. Casella, "A short history of markov chain monte carlo: Subjective recollections from incomplete data," 2011.

[93] R. Brunelli, *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[94] P. Jaccard, "The distribution of the flora in the alpine zone. 1," *New phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[95] W. Niblack, *An introduction to digital image processing*. Strandberg Publishing Company, 1985.

[96] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *International conference on database theory*. Springer, 2001, pp. 420–434.

[97] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in *Database Theory—ICDT'99: 7th International Conference Jerusalem, Israel, January 10–12, 1999 Proceedings 7*. Springer, 1999, pp. 217–235.

[98] R. K. Pathria, *Statistical mechanics*. Elsevier, 2016.

[99] J. Lerner, P. A. Gomez-Garcia, R. L. McCarthy, Z. Liu, M. Lakadamyali, and K. S. Zaret, "Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin," *Molecular Cell*, vol. 79, no. 4, pp. 677–688, 2020.

[100] J. E. Maris, F. T. Rabouw, B. M. Weckhuysen, and F. Meirer, "Classification-based motion analysis of single-molecule trajectories using diffusionlab," *Scientific Reports*, vol. 12, no. 1, p. 9595, 2022.

[101] E. A. Codling, M. J. Plank, and S. Benhamou, "Random walk models in biology," *Journal of the Royal society interface*, vol. 5, no. 25, pp. 813–834, 2008.

[102] F. Höfling and T. Franosch, "Anomalous transport in the crowded world of biological cells," *Reports on Progress in Physics*, vol. 76, no. 4, p. 046602, 2013.

[103]

[104] S. Kang, J.-H. Kim, M. Lee, J. W. Yu, J. Kim, D. Kang, H. Baek, Y. Bae, B. H. Kim, S. Kang *et al.*, "Real-space imaging of nanoparticle transport and interaction dynamics by graphene liquid cell tem," *Science Advances*, vol. 7, no. 49, p. eabi5419, 2021.

[105] M. J. Saxton, "Anomalous diffusion due to obstacles: a monte carlo study," *Biophysical journal*, vol. 66, no. 2, pp. 394–401, 1994.

[106] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.

[107] L. Liu, R. Chen, W. Liu, Y. Zhang, X. Shi, and Q. Pan, "Fabrication of superhydrophobic copper sulfide film for corrosion protection of copper," *Surface and Coatings Technology*, vol. 272, pp. 221–228, 2015.

[108] K. Ozawa and K. Mase, "Evidence for chemical bond formation at rubber–brass interface: Photoelectron spectroscopy study of bonding interaction between copper sulfide and model molecules of natural rubber," *Surface Science*, vol. 654, pp. 14–19, 2016.

[109] Z. Yao, J. Rogalinski, E. M. Asimakopoulou, Y. Zhang, K. Gordeyeva, Z. Atoufi, H. Dierks, S. McDonald, S. Hall, J. Wallentin *et al.*, "New opportunities for time-resolved imaging using diffraction-limited storage rings," *Synchrotron Radiation*, vol. 31, no. 5, 2024.

[110] T.-S. Vu, M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, Y. Abe, T. Tran, H. Tran, H. Kino, T. Miyake, K. Tsuda *et al.*, "Towards understanding structure–property relations in materials with interpretable deep learning," *npj Computational Materials*, vol. 9, no. 1, p. 215, 2023.

[111] M. Gidding, T. Janssen, C. Davies, and A. Kirilyuk, "Dynamic self-organisation and pattern formation by magnon-polarons," *Nature communications*, vol. 14, no. 1, p. 2208, 2023.

[112] B. Josso, D. R. Burton, and M. J. Lalor, "Texture orientation and anisotropy calculation by fourier transform and principal component analysis," *Mechanical Systems and Signal Processing*, vol. 19, no. 5, pp. 1152–1161, 2005.

[113] J. P. Marquez, "Fourier analysis and automated measurement of cell and fiber angular orientation distributions," *International journal of solids and structures*, vol. 43, no. 21, pp. 6413–6423, 2006.

[114] R. Alberini, A. Spagnoli, M. J. Sadeghinia, B. Skallerud, M. Terzano, and G. A. Holzapfel, "Fourier transform-based method for quantifying the three-dimensional orientation distribution of fibrous units," *Scientific Reports*, vol. 14, no. 1, p. 1999, 2024.

[115] M. Hüpfel, A. Yu. Kobitski, W. Zhang, and G. U. Nienhaus, "Wavelet-based background and noise subtraction for fluorescence microscopy images," *Biomedical Optics Express*, vol. 12, no. 2, pp. 969–980, 2021.

[116] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the 3rd international conference on knowledge discovery and data mining*, 1994, pp. 359–370.

[117] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[118] G. Binnig and H. Rohrer, "Scanning tunneling microscopy—from birth to adolescence," *reviews of modern physics*, vol. 59, no. 3, p. 615, 1987.

[119] F. J. Giessibl, "Advances in atomic force microscopy," *Reviews of modern physics*, vol. 75, no. 3, p. 949, 2003.

[120] U. Landman, W. Luedtke, N. A. Burnham, and R. J. Colton, "Atomistic mechanisms and dynamics of adhesion, nanoindentation, and fracture," *Science*, vol. 248, no. 4954, pp. 454–461, 1990.

[121] G. Rubio-Bollinger, S. R. Bahn, N. Agraït, K. W. Jacobsen, and S. Vieira, "Mechanical properties and formation mechanisms of a wire of single gold atoms," *Physical Review Letters*, vol. 87, no. 2, p. 026101, 2001.

[122] N. Agraıt, A. L. Yeyati, and J. M. Van Ruitenbeek, "Quantum properties of atomic-sized conductors," *Physics Reports*, vol. 377, no. 2-3, pp. 81–279, 2003.

[123] T. Kizuka, "Atomic process of point contact in gold studied by time-resolved high-resolution transmission electron microscopy," *Physical Review Letters*, vol. 81, no. 20, p. 4448, 1998.

[124] T. Kizuka and K. Monna, "Atomic configuration, conductance, and tensile force of platinum wires of single-atom width," *Physical Review B—Condensed Matter and Materials Physics*, vol. 80, no. 20, p. 205406, 2009.

[125] S. Kawai, F. F. Canova, T. Glatzel, A. S. Foster, and E. Meyer, "Atomic-scale dissipation processes in dynamic force spectroscopy," *Physical Review B—Condensed Matter and Materials Physics*, vol. 84, no. 11, p. 115415, 2011.

[126] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[127] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[128] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[129] M. Sezgin and B. l. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging*, vol. 13, no. 1, pp. 146–168, 2004.

[130] J. Liu, J. Zhang, T. Arai, M. Tomitori, and Y. Oshima, "Critical shear stress of gold nanocontacts estimated by in situ transmission electron microscopy equipped with a quartz length-extension resonator," *Applied Physics Express*, vol. 14, no. 7, p. 075006, 2021.

[131] J. Zhang, M. Tomitori, T. Arai, and Y. Oshima, "Surface effect on young's modulus of sub-two-nanometer gold [111] nanocontacts," *Physical Review Letters*, vol. 128, no. 14, p. 146101, 2022.

[132] A. P. Sutton and T. N. Todorov, "Mechanical and electrical properties of metallic contacts at the nanometer scale," *Journal of Physics and Chemistry of Solids*, vol. 55, no. 10, pp. 1169–1174, 1994.

[133] A. Khosravi, A. Lainé, A. Vanossi, J. Wang, A. Siria, and E. Tosatti, "Understanding the rheology of nanocontacts," *Nature Communications*, vol. 13, no. 1, p. 2428, 2022.

[134] D. Cabut, M. Michard, S. Simoens, L. Mees, V. Todoroff, C. Hermange, and Y. Le Chenadec, "Analysis of the water flow inside tire grooves of a rolling car using refraction particle image velocimetry," *Physics of Fluids*, vol. 33, no. 3, 2021.

[135] R. Nakanishi, M. Matsubara, T. Ishibashi, S. Kawasaki, H. Suzuki, H. Kawabata, S. Kawamura, and D. Tajiri, "Tire mechanical model for cornering simulation with friction coefficient calculated from viscoelasticity of rubber by multiscale friction theory," *Vehicle System Dynamics*, vol. 62, no. 9, pp. 2401–2422, 2024.

[136] T. Fwa, S. S. Kumar, K. Anupam, and G. P. Ong, "Effectiveness of tire-tread patterns in reducing the risk of hydroplaning," *Transportation research record*, vol. 2094, no. 1, pp. 91–102, 2009.

[137] A. Blum, J. Hopcroft, and R. Kannan, *Foundations of data science*. Cambridge University Press, 2020.

[138] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47.

[139] D. Kinga, J. B. Adam *et al.*, "A method for stochastic optimization," in *International conference on learning representations (ICLR)*, vol. 5, no. 6. California;, 2015.