Control and Prediction of Plasmonic Gold Nanoparticle Assembly

Muammer Yusuf Yaman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

David S. Ginger, Chair

David J. Masiello

Bo Zhang

Program Authorized to Offer Degree:

Chemistry : Data Science

University of Washington

Abstract

Control and Prediction of Plasmonic Gold Nanoparticle Assembly

Muammer Yusuf Yaman

Chair of the Supervisory Committee:
David S. Ginger
Department of Chemistry

Optically-active materials can be identified by their changeable optical signatures, allowing for noncontact insights into the nanoscale behavior of the material and predicting its functional use. This work discusses the optical characterization and plasmonic response of various metal nanoparticles, including those bio-templated. First, we explore the interaction between protein fiber and gold nanoparticles, emphasizing the effects of ionic strength and particle aspect ratio on the assembly process. Automated image analysis further sheds light on the unique behaviors of Au nanoparticles under varying assembly conditions. Secondly, we examine the structure-property relationship of plasmonic nanoparticle clusters, especially nanospheres. Advanced imaging techniques reveal detailed structural information, leading to a proposed non-invasive assembly approach under hyperspectral microscopy. We developed advanced machine learning tools to fast,

accurate predict the plasmonic response. Finally, we investigate the orientation of anisotropic plasmonic nanoparticles using polarized hyperspectral scattering, focusing on gold particles of differing aspect ratios. This investigation provides an in-depth understanding of the structure-property relationship for both nanospheres and nanorods. The culmination of this research highlights the potential of advanced machine learning and characterization techniques in predicting and controlling plasmonic gold nanoparticle assembly, offering invaluable insights into their optical properties.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I am deeply thankful to my advisor, Dr. David S. Ginger, for his unwavering guidance, mentorship, and encouragement throughout this research journey. His expertise, patience, and dedication have been instrumental in shaping the direction of this work and fostering my academic growth.

I extend my sincere appreciation to the members of my dissertation committee, Dr. Arka Majumdar, Dr. David Masiello, and Dr. Bo Zhang for their valuable insights, constructive feedback, and critical evaluations that have greatly enriched the quality of this research.

I am also indebted to my lab mates, Dr. Fangyuan Jiang, Dr. Rajiv Giridharagopal, Dr. Kathryn Guye, Dr. Jian Wang, Dr. Xudong Wang, Dr. Dean Waldow, Rahoul Ghosh, and all Ginger lab members for their camaraderie, stimulating discussions, and willingness to lend a helping hand whenever needed. Your support and collaborative spirit have made this research experience truly enjoyable and rewarding.

I would like to extend my gratitude to the Department of Energy Basic Energy Sciences Office for their generous funding support for much of the ensuing work via the Energy Frontier Research Center - the Center for the Science of Synthesis Across Scales (CSSAS). I'm also deeply thankful to the entire CSSAS community especially Dr. Sergei V. Kalinin, Dr. Maxim Ziatdinov, Dr. David Baker, Dr. Hao Shen, Dr. Zhe Li, and Dr. Shunzhi Wang. Beyond producing outstanding scientific work, they've fostered a distinct community and support system, leading to numerous professional relationships and lasting friendships.

I am profoundly grateful to my dear friends Fangyuan, Yasin&Laura, Murat, Tugrul, Jiajie, Farhad, Yangwei, Emerson, BurakX3, Mustafa, Alper, Emre, Oguzhan, Aykut, Hilmi and

many others, who have been by my side throughout this journey. Your presence has made a world of difference, and I am incredibly blessed to have you in my life.

Last but not least, I am immensely grateful to my family for their unwavering love, understanding, and encouragement throughout this challenging academic pursuit. Their constant support and belief in my abilities have been a source of strength and motivation. Without their unwavering encouragement, this accomplishment would not have been possible.

Chapter 1. INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

In this chapter, we briefly discuss the origin of the optical properties of plasmonic nanoparticles, macromolecular templates hierarchical assembly of plasmonic materials and finally, how to predict optical properties of plasmonic nanoparticles using simulation and advance machine learning techniques.

1.2 PLASMONIC NANOPARTICLES AND THEIR OPTICAL PROPERTIES

The optical properties of plasmonic nanoparticles have captivated human curiosity for centuries, ever since they were first harnessed to imbue glass and ceramics with a captivating red hue. Over time, our comprehension of these fascinating particles and their remarkable attributes has matured, leading to the development of sophisticated technologies like plasmonic nanosensors[1], nanolasers[2], nanopixels,[3] and applications in the field of nanomedicine[4]. Central to these captivating optical properties is the phenomenon of localized surface plasmon resonance (LSPR),[5] a concept that arises from the oscillation of free electrons within metal nanoparticles, as illustrated in Figure Figure 1-1 below. Crucially, LSPR is only achievable when the nanoparticle's dimensions are significantly smaller than the incident wavelength of light, allowing the electron cloud within the metal to interact with the oscillating electric field of the incoming light wave.

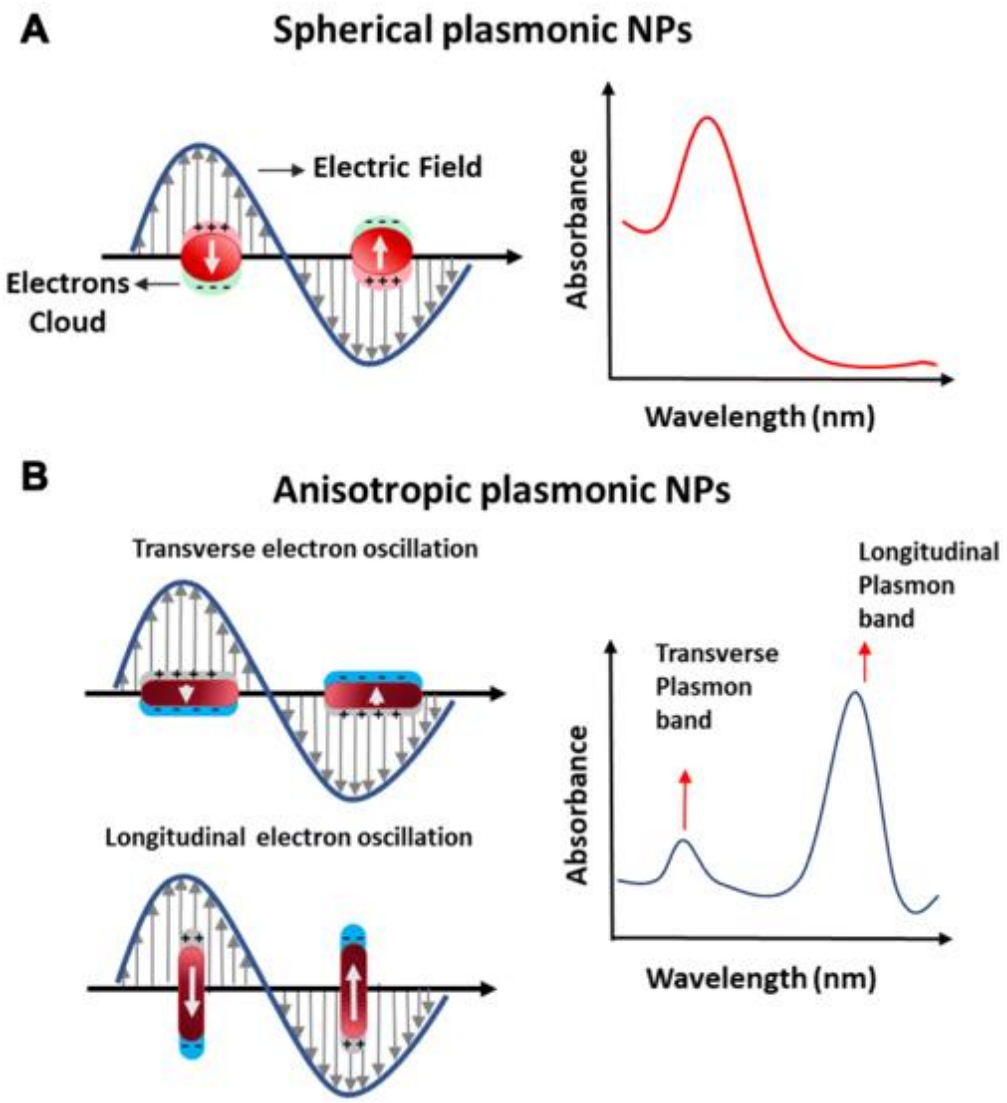

Figure 1-1 Schematic representation of surface plasmon resonance (SPR) band in (A) spherical and (B) rod-shaped plasmonic nanoparticles (NPs). Adapted from Ref.[141] with permission from the Springer Nature.

According to Mie theory[6], in the case of gold nanoparticles, the electric dipole resulting from the oscillation of free electrons primarily drives the intense light scattering observed at the LSPR position. This light scattering is the reason behind the captivating, opalescent appearance of nanoparticle solutions—an aesthetic quality that has been leveraged for both artistic and practical purposes.[7] Moreover, it's crucial to note that the LSPR's extinction is exquisitely sensitive to fluctuations in the dielectric properties of the surrounding medium. Even slight variations in the

local refractive index can cause a noticeable redshift in the LSPR position. This remarkable sensitivity positions nanoparticles as exceptional candidates for applications in nanosensing, an area of great scientific and technological interest.

The versatility of plasmonic nanoparticles doesn't end with LSPR alone. By altering the geometry of the nanoparticle from the familiar isotropic spherical shape to various anisotropic configurations, such as nanorods, researchers can induce additional resonance modes. These modes include the longitudinal and transverse LSPRs, which are the result of variations in the nanoparticle's length and width, respectively.[5] When two spherical nanoparticles approach each other closely, their plasmonic modes couple, giving rise to a longitudinal resonance along the dimer axis, in addition to the conventional perpendicular transverse peak. During this coupling, electrons within the dimer structure oscillate collectively rather than independently. Furthermore, the electric field strength in the gap between the two particles amplifies significantly when the longitudinal mode is activated.

Given the unparalleled optoelectronic properties offered by plasmonic nanoparticles, researchers have devoted considerable effort to the controlled assembly of gold nanoparticles. This endeavor aims to harness their potential for a wide array of applications, including sensing, catalysis, laser technology, and optical transistors. Beyond static structures, Boykan and his team have also explored the exciting realm of stimulus-responsive organic functional materials. These materials have the intriguing capability to reversibly manipulate the interparticle distance, orientation, or aggregation state of inorganic nanoparticles. Consequently, they open up exciting avenues for dynamically tuning the optoelectronic characteristics of these systems, which will be discussed further in the following section.

1.3 MACROMOLECULAR TEMPLATES HIERARCHICAL ASSEMBLY

The quest for harnessing the fascinating properties of plasmonic nanoparticles has led to a plethora of research, aiming to understand, control, and utilize these unique materials. Among them, gold nanoparticles (AuNPs) have attracted significant attention due to their intriguing optical, catalytic, and electronic properties. Further interest has been shown in organizing these nanoparticles into functional assemblies, a concept known as hierarchical self-assembly. This chapter aims to provide a comprehensive literature review on the self-assembly of Au nanoparticles and the role of biological templates in facilitating this process.

The mechanisms of hierarchical self-assembly can be drive through Van der Waals interactions, electrostatic interactions, capillary forces. The main driving forces that contribute to self-assembly is the van der Waals interaction. Park and this team has shown how this force can lead to the organization of nanoparticles into well-defined structures. [8] The surface charge on nanoparticles can also play a crucial role in their self-assembly. Studies have been conducted to examine how varying surface charges can influence the final assembled structure. In liquid environments, capillary forces can also drive self-assembly. Research has shown that these forces can be used to arrange nanoparticles on both flat and curved surfaces [6].

The growth in the field of harnessing bio-derived and bio-inspired substances for the structured assembly of inorganic nanostructures can be attributed to the innate ability of organic structures to systematically arrange inorganic components, their atomic-level precision, and adaptable interactions that traditional ligand interactions might lack rapid. Biomolecular templates, including DNA[9], proteins[10], peptides, and other bio-inspired materials, serve as the scaffolds for orchestrating the assembly of AuNPs. These templates offer unique advantages such as atomic-scale control, programmable designs, and tunable interactions that are often beyond the reach of

traditional ligand-based approaches. The specificity of DNA base pairing allows for the precise arrangement of nanoparticles into predetermined geometries.[9] Mirkin and his team have extensively explored the potential of DNA to assemble nanoparticle structures[11]. Over a quarter of a century, they have innovatively used DNA to organize colloidal nanoparticles in carefully designed ionic crystal formations[9]. By expertly manipulating the binding strength of DNA attached to gold nanoparticles, they've demonstrated impressive control over nanoparticle crystallization. Their studies have highlighted the reversible nature of assembly and the creation of unique metallic behaviors in nanoparticles. DNA's versatility is further illustrated in its application in DNA origami, thin film assembly, and responsive nanoparticle dimers.[12]

Peptides and proteins, given their complex three-dimensional structures, offer another avenue for templating nanoparticle assembly. Rosi's team showcased how gold nanoparticles can be arranged into chiroptical helices using peptide templates[11], while Lee and colleagues revealed the growth of chiral helicoid nanoparticles with the introduction of chiral amino acids.[13] Yan's team, in 2018, spotlighted peptoids, synthetic peptide analogs, directing the formation of intricately branched gold nanoparticles, which achieved exceptional electric field enhancement compared to conventionally prepared particles. Tezcan's group innovated 2D lattices that are redox-responsive, mechanically adaptable, and can switch configurations.[12] Furthermore, recent strides in designing proteins from scratch have enabled the creation of 3D highly ordered crystals that are optically active and reversible photonic behavior as demonstrated by Baker and his team[14].

In essence, the convergence of bio-based and bio-inspired materials with inorganic nanoparticles not only presents a path to achieve specific functionalities but also heralds the dawn of novel optoelectronic traits (see Figure 1-2).

Hierarchically assembled plasmonic structures can be used for different optoelectronic applications ranging from sensing to drug deliveries. The assembled structures find application in sensing due to amplified surface plasmon resonance signals. Studies have shown their potential in detecting low concentrations of various analytes. Hierarchical assemblies of Au nanoparticles can encapsulate drugs for targeted delivery, thus minimizing side effects. Due to their enhanced optical properties, assembled structures can be used for photothermal therapies, offering new avenues in cancer treatment[8]. While there have been significant advancements in this field, gaps remain, especially concerning the long-term stability of these structures, scalability of the assembly process, and in vivo applications.

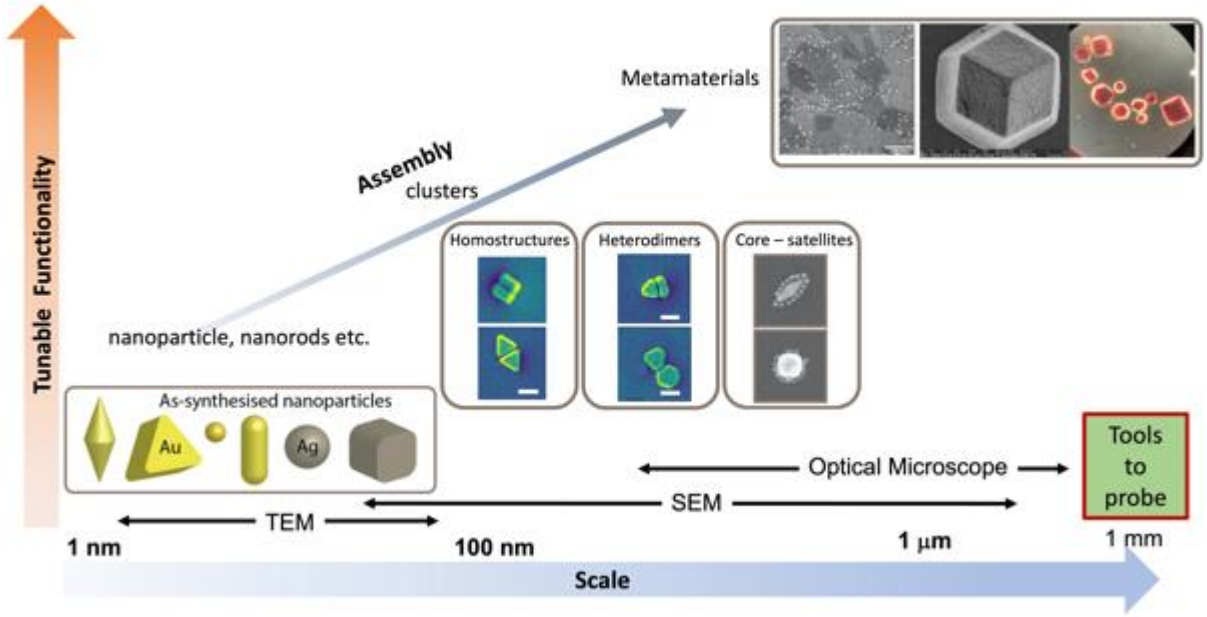

Figure 1-2 Assembly pathway of hierarchical metamaterials. The size of assembled plasmonic nanoparticles increases with their tunable functionalities. Different microscope can probe the assembled structure.

In conclusion, the structured assembly of Au nanoparticles presents an optimistic avenue for designing functional configurations with diverse applications. Existing literature suggests that biological frameworks offer unmatched benefits in producing consistent, enduring, and functional clusters. Nonetheless, there are ongoing challenges, highlighting the need for continued exploration.

1.4 PREDICTING PLASMONIC RESPONSES

The prediction of photonic responses in gold nanoparticles requires a synergistic approach, combining computational modeling, machine learning, and experimental validation. There are different computational modeling tools to predict plasmonic responses. Firstly, Finite Element Method (FEM) is a numerical technique widely employed in the study of various physical phenomena, including electromagnetism[15]. In the context of gold nanoparticles, it is especially useful for modeling the photonic responses under different conditions such as shape, size, and external electromagnetic fields. The complex geometric configurations can be discretized into smaller elements, and Maxwell's equations can be solved in these elements to derive the local electric field, magnetic field, and corresponding scattering and absorption properties.[15,16]

The primary advantage of FEM is its high accuracy and the ability to model irregular shapes and heterogeneous materials. However, the method can be computationally intensive, especially for larger systems or higher frequencies. Secondly, Discrete Dipole Approximation (DDA) is another numerical method for calculating scattering and absorption of electromagnetic waves by targets with arbitrary geometries.[17] This method treats the target as an array of polarizable points, solving for the resulting dipole moments under the influence of an external electromagnetic field. DDA is less computationally demanding than FEM but may be less accurate for certain shapes or

material inhomogeneities[18]. It is often more suited for particles that are much smaller than the wavelength of light and for systems where the geometry can be simplified into a grid of discrete points.

Machine learning, particularly neural networks, has gained traction in predicting photonic responses of gold nanoparticles. By training a neural network on a dataset generated through traditional computational methods (like FEM and DDA) or experimental data, researchers can make predictions on new, unexplored geometries and conditions much more quickly than by using computational simulations alone. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown promise in capturing the complex relationships between shape, size, material properties, and photonic responses, including scattering and absorption spectra.[19]

Big data approaches are also gaining momentum in the field of nanophotonic. By leveraging massive datasets from both experimental and computational sources, researchers can apply various data analytics techniques, including cluster analysis and principal component analysis, to identify underlying patterns and correlations.[20] This data-driven approach can not only speed up the design and optimization process but also uncover novel phenomena and design rules that are not readily apparent through traditional methods.

Computational models need to be validated against experimental data to ensure their accuracy and applicability. Methods such as spectroscopy (various speed and resolution, in Figure 1-3), electron microscopy, and interferometry are commonly employed to measure photonic responses of gold nanoparticles in Figure 1-2. These experimental results serve as the gold standard against which computational predictions can be compared.

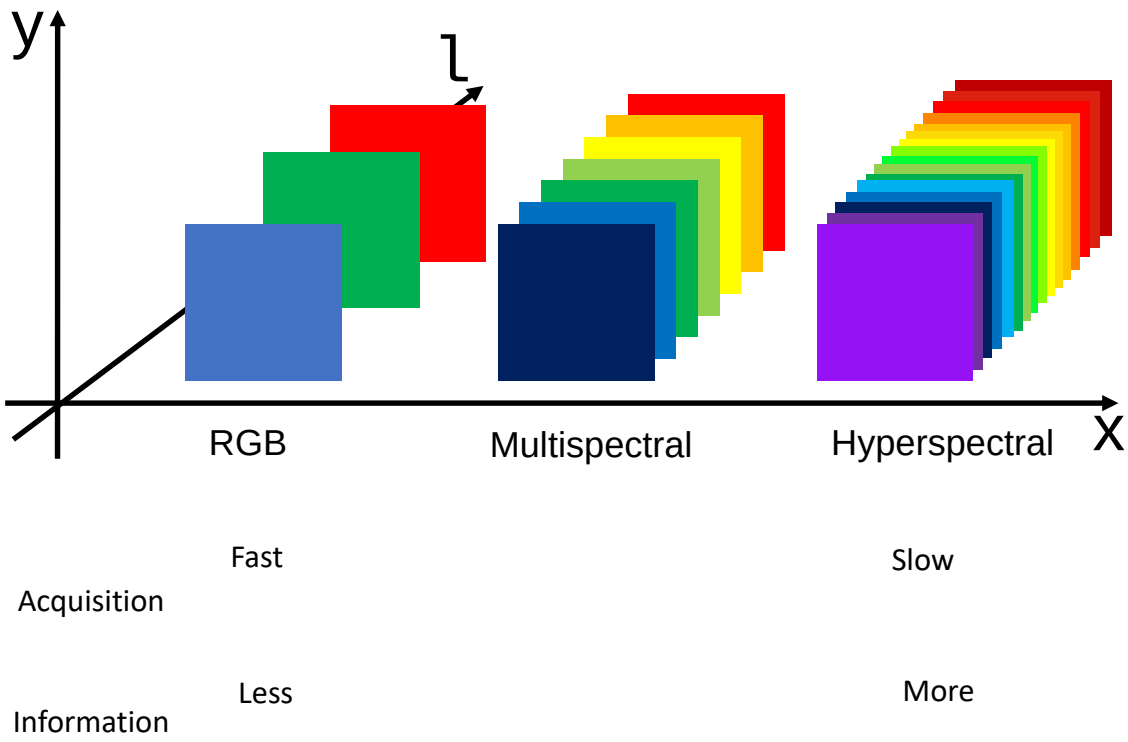

Figure 1-3 Different optical spectroscopy techniques. RGB camera acquires fast but has limited information whereas hyperspectral camera has full spectrum from 400 nm to 1000nm but slow acquisition time.

Apart from experimental verification, the accuracy of a new model can also be assessed by comparing its predictions with those from established models and existing data sets. This offers an additional layer of validation and can highlight any limitations or inaccuracies in the new approach. Such comparison is crucial in determining the reliability and applicability scope of the model in question. For instance, if a machine-learning model trained on spherical nanoparticles performs poorly when predicting the photonic response of rod-shaped nanoparticles, this limitation must be explicitly stated.

In summary, predicting the photonic responses of gold nanoparticles is a multi-faceted endeavor that benefits from a combination of computational modeling, machine learning, and experimental

validation. The synergistic use of these methods holds the promise of dramatically advancing our understanding and control of nanoparticle-based photonic devices and applications.

1.5 SUMMARY

This dissertation describes three studies which seek to understand the plasmonic assembly guidelines from induvial particles level to large area synthesis using protein fiber as a templates and to build structure – property relationship for plasmonic nanoparticles to predict their optical properties.

In Chapter 2, we delve into understanding the interaction between protein fiber and gold nanoparticles. crystalline changes occurring during the electrochemical doping process. Specifically, we focus on elucidating the effect of external factor like ionic strength, aspect ratio of the particles on the assembly process. Through automated image analysis tools, we investigate the unique behavior of Au nanoparticles after different assembly condition.

Chapter 3 takes a closer look at the structure – property relationship of plasmonic nanoparticle clusters especially nanospheres. With direct probing of plasmonic response of gold nanoparticle clusters, we uncover the structural information of these clusters using high resolution imaging. Subsequently, we propose to non-invasive approach to assembly of plasmonic materials under hyperspectral microscopy.

Finally, in Chapter 4, we study the orientation of anisotropic plasmonic nanoparticles. Employing polarized hyperspectral scattering techniques, we investigate the plasmonic response of gold particles with different aspect ratios. By gaining a comprehensive understanding of these structure – property relationship of plasmonic particles both nanospheres and nanorods, we aim to propose in-situ characterization of the macromolecular templated hierarchical assembly of plasmonic materials.

Overall, this dissertation showcases the remarkable capabilities of control and prediction of plasmonic gold nanoparticle assembly using advanced machine learning tools and characterization techniques. By examining different experimental processes at different scale, we contribute valuable knowledge to the assembly of plasmonic nanoparticles and predicting their optical properties.

Chapter 2. Alignment of Au Nanorods Along de novo Designed Protein Nanofibers Studied with Automated Image Analysis

Adapted with permission from Muammer Y. Yaman, Kathryn N. Guye, Maxim Ziatdinov, Hao Shen, David Baker, Sergei V. Kalinin, David Ginger. "Alignment of Au Nanorods Along de novo Designed Protein Nanofibers Studied with Automated Image Analysis" Soft Matter 2021,17, 6109-6115 DOI: 10.1039/D1SM00645B Copyright 2021, Royal Society of Chemistry.

2.1 Overview

In this study, we focus on exploring the directional assembly of anisotropic Au nanorods along de novo-designed 1D protein nanofiber templates. Using machine learning and automated image processing, we analyze scanning electron microscopy (SEM) images to study how the attachment density and alignment fidelity are influenced by variables such as the length and aspect ratio of the Au nanorods, as well as the salt concentration of the solution. We find that the Au nanorods prefer to align parallel to the protein nanofibers. This preference decreases with increasing salt concentration, but is only weakly sensitive to the nanorod aspect ratio. While the overall specific Au nanorod attachment density to the protein fibers increases with increasing solution ionic strength, this increase is dominated primarily by non-specific binding to the substrate background, and we find that most specific attachment (highest ratio of nanorods attached to the nanofiber template as compared to the substrates) occurs at the lower studied salt concentrations, with the maximum ratio of specific to non-specific binding occurring when the protein fiber solution are prepared in 75 mM NaCl concentration.

2.2 INTRODUCTION

Hierarchically-organized hybrid organic/inorganic structures are interesting candidates for developing new advanced materials. Using programmable macromolecular building blocks such as proteins,[21–23] peptoids,[24–27] and polymers,[28–36] as templating agents to scaffold the assembly of functional inorganic building blocks is one approach to achieve this goal. It can be advantageous, for example, to combine the programmable atomic-precision afforded by biomolecular self-assembly, with the optical and electronic properties of inorganic materials. These materials are important in fields ranging from plasmonics,[31,37,38] to quantum optics,[39,40] and biosensing,[41–43]. Kotov and co-workers recently used protein[44] and peptide[45] templates to direct the assembly of gold nanorods to produce chiroptical structures with record chiral dichroism and optical asymmetry g-factors[45].

In pursuing such approaches, a key challenge is to understand and control the physical interactions between the inorganic building blocks and the biological templates, especially at the interfaces. Understanding these processes can provide guide rules for future design and rational synthesis of hierarchical materials. Electrostatic interactions are classic and widely used, yet still interesting area to explore. For example, by controlling salt concentration, which affects the electrostatic force at the protein–mineral interfaces, Pyles et al. showed the orientation proteins prefer to attach to an inorganic substrate can be tailored by varying the ionic strength of the solution.[46]

Here, we explore the converse problem: what factors control the fidelity of alignment of metal nanorods with protein fibers acting as the assembly templates. We use a *de novo* designed protein fiber[47] engineered to have a very high density of negatively-charged surface residues, under neutral condition, as a template to drive the electrostatic assembly of positively-charged gold nanorods

along the nanofiber axis. We conduct a series of experiments varying the external parameters such as the aspect ratio of the Au nanorods and the salt concentration (ionic strength) to assess how they affect the assembly of the Au nanorods on the protein fibers relative to an aminosilanized ITO substrate. We develop and apply an automated image analysis tool to facilitate analysis of the experimental data. We find that the fidelity of the electrostatically-driven self-assembly of Au nanorods onto protein nanofiber templates, depends on the aspect ratio of the nanorods, as well as the ionic strength of the solution.

2.3 RESULTS AND DISCUSSION

Figure 2-1**(a)** shows a schematic of our assembly approach for attaching Au nanorods onto pre-assembled de novo designed protein nanofibers (see SI for full details). Briefly, we drop cast pre-assembled protein fibers[47] onto positively-charged silane-coated ITO substrates. As a monolayer, t silane, (3-aminopropyl)triethoxysilane, has a pKa of 7.6.[48,49] Thus, when the protein is deposited in a pH 8 tris buffer, the silane surface remains partially protonated, yielding a net positive charge, to which the negatively-charge proteins can attach due to a combination of electrostatic and van der Waals attraction[49,50] . Next, as shown, we exposed these positively-charged surfaces to solutions of preassembled protein nanofibers. We designed the protein nanofibers to have a high-density of carboxylic acid groups on their surfaces,[47] decorating them with a net negative surface charge in buffer at pH 8. Due to the electrostatic attraction, the negatively-charged protein fibers attach at high densities to the charged silane. (see the SEM image in Figure 2-1**a)** The protein nanofibers attach at much lower density (if at all) to substrates without silane treatment (See Figure A-1)

Next, following attachment of the protein nanofibers, we exposed the substrates to solutions of Au nanorods synthesized by the method of Ye et al.[51]. Terminated with cetyl trimethyl

ammonium (CTA) cations, these Au nanorods have a net positive charge,[52] and thus are candidates for electrostatically-driven self-assembly onto the protein nanofibers. Due to the large density of the Au nanorods, we performed the nanorod attachment step with inverted substrates (Figure 2-1) in order to allow the assembly process be dominated by short-range interactions while eliminating the potential for gravitational sedimentation.

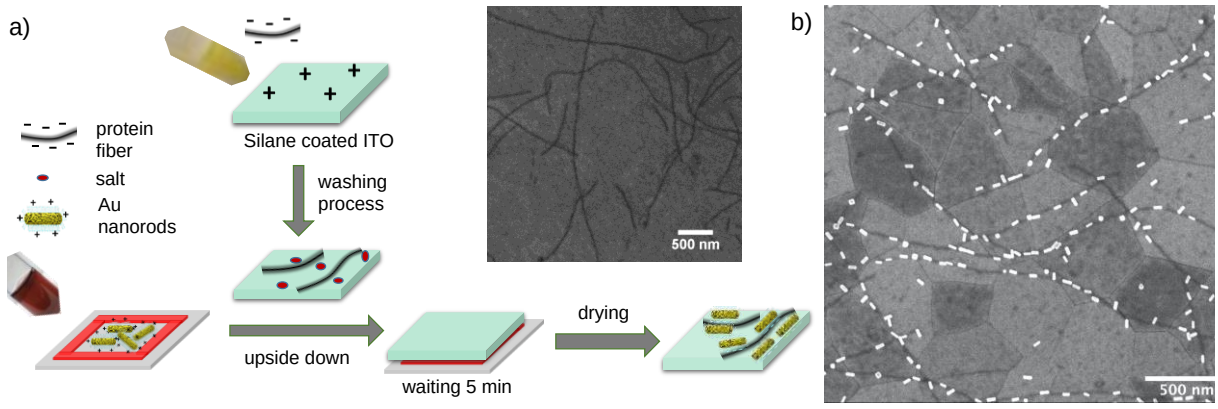

Figure 2-1 Representation of the assembly process. a) Schematic representation of the assembly process and a SEM image of protein fibers on a silane-treated ITO substrate. The protein fibers were drop cast onto the silane-coated ITO substrate and then the substrate was put upside-down on the top of the chamber filled with Au nanorods in solution (to avoid attachment by sedimentation). The substrate was gently washed and dried for imaging. b) A SEM image of Au-decorated de novo designed protein nanofibers.

We then gently washed the substrates with distilled water to remove the excess Au nanorod solution, and imaged the resulting structures using scanning electron microscopy (SEM). Figure 2-1(b) shows the resulting Au nanorod decorated protein nanofibers. In these images the nanofibers appear clearly as black lines stretching for several microns, while the Au nanorods appear as bright white cylindrically-shaped objects. We used the ITO substrate to perform SEM imaging on a conductive surface without the need to sputter over the structures. However, the grain structure of the ITO is also faintly visible in the background. While the ITO grain boundaries

also appear as black lines, they are much lighter and thinner than the protein nanofibers, and are easily distinguished from the protein.

We observed that protein nanofibers attached on functionalized ITO at a reasonably high density (2.5 ± 1.3 μm protein nanofibers/μm^2 area) (See Figure A-2 A) for solutions with different concentrations of salt between 25 mM and 1M, and the average density of protein fibers at each solution are quite similar. On the contrary, the Au nanorods were attached to primarily to the nanofibers or to the substrate with different densities in 25 ± 13 particles/ μm^2 area, which are depending on the sample and attachment conditions (See Figure A-2 B).

In order to explore how different variables affect the fidelity of the assembly of Au nanorods along the protein nanofibers, we performed a number of experiments as a function of nanorod aspect ratio, and ionic strength of the solution. In order to efficiently generate statistically-robust data sets and to automate the workflow for future experiments, we utilized a python-based automatic image analysis tool.[53,54] In this context, the goal for such an image analysis tool is to automatically recognize the protein fibers and label Au nanorods as attached to protein fibers or substrate under different experimental conditions. Here, we consider binding to the protein fibers as specific binding, and to the substrate as nonspecific binding. In addition, the image analysis tool should provide additional information concerning the angle of each Au nanorod on protein fibers, the size distribution of nanorods, and the length of protein fibers. All this information together should help to analyze each individual particle and hopefully will contribute to our understanding of local surface-surface interactions at the nanoscale. Figure 2-2 shows the workflow of the automated image analysis tool. Firstly, the contrast of the images was adjusted and then filtered regional maxima to find bright gold particles in the image. Next, we utilized a fully convolutional neural network (FCNN) to separate the protein fibers from the rest of the image.[55,56] Specifically,

the FCNN perform a semantic segmentation of the input images by categorizing every pixel in the image as belonging to a fiber or to a background.[57] The FCNN structure is based on the custom-build *dilnet* architecture,[58] which uses only a single max-pooling (and the corresponding up-sampling) operation to preserve the maximum amount of information and utilizes dilated convolutions to reduce the total number of weights to train. The latter reduces a computational cost and allows training the FCNN for the large-size images. To train the FCNN, we created image-mask pairs by hand labelling a small subset of experimental images and performing the standard data augmentation procedure, which included random cropping, rotation, and horizontal/vertical flipping. The FCNN weights were optimized using the Adam extension[59] of the stochastic gradient descent algorithm using with the binary cross-entropy loss objective. Once the loss reaches a plateau (with the value of ~0.1), the predicted morphology of protein fibers is adopted for further analysis. Finally, the new image was automatically regenerated from both outputs.

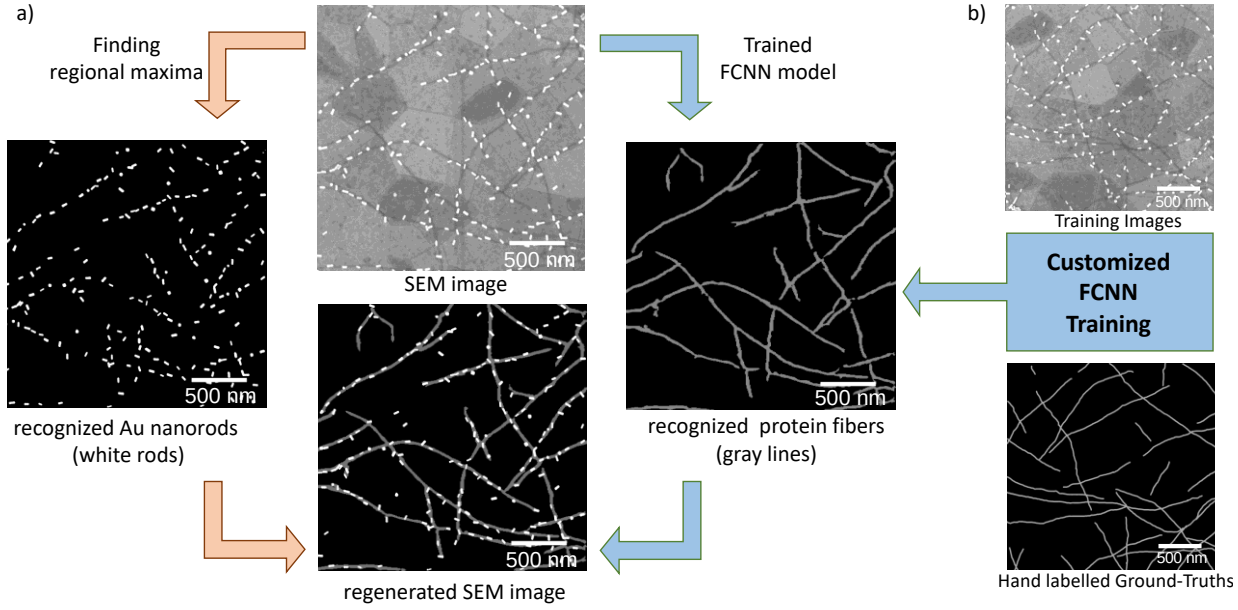

Figure 2-2 Representation of automated workflow. a) Schematic representation of our automated SEM image analysis tool. b) Training process of the Fully Convolutional Neural Network (FCNN) model for recognizing the protein fibers.

After assigning the position of the nanofibers using the FCNN, and the positions of the bright white Au nanorods using thresholding, we then analyzed the resulting labeled images. We used custom automated image analysis tools to compute properties, such as Au nanorod particle size and aspect ratio, and to quantify specific Au nanorods attachment on protein fibers, and nonspecific Au nanorods attachment to the substrate. We also used these tools to compute the angle that the axis of each specifically attached Au nanorod made with respect to the tangent to the local protein nanofiber axis.

After developing the automatic image analysis tool, we first investigated the effect of solution ionic strength on the attachment of the Au nanorods to the protein fibers. We used 25 mM, 75 mM, 150, and 1 M NaCl solutions. We note that these NaCl concentrations were those applied to the protein buffer conditions during assembly, and that the salt concentrations were lower at the final stages of assembly because of the washing steps – but we anticipate that at by washing step the assembly was mostly completed (as verified by the significantly differences in particle assembly results based on ionic strength). Figure 2-3(a) shows the resulting morphologies of the final assembled nanofiber/nanorod solutions. Those images were generated using the image analysis tool, with the protein nanofibers were labelled with a gray line, and the Au nanorods were labelled with a white rod shape. Figure A-2 shows the corresponding raw SEM images.

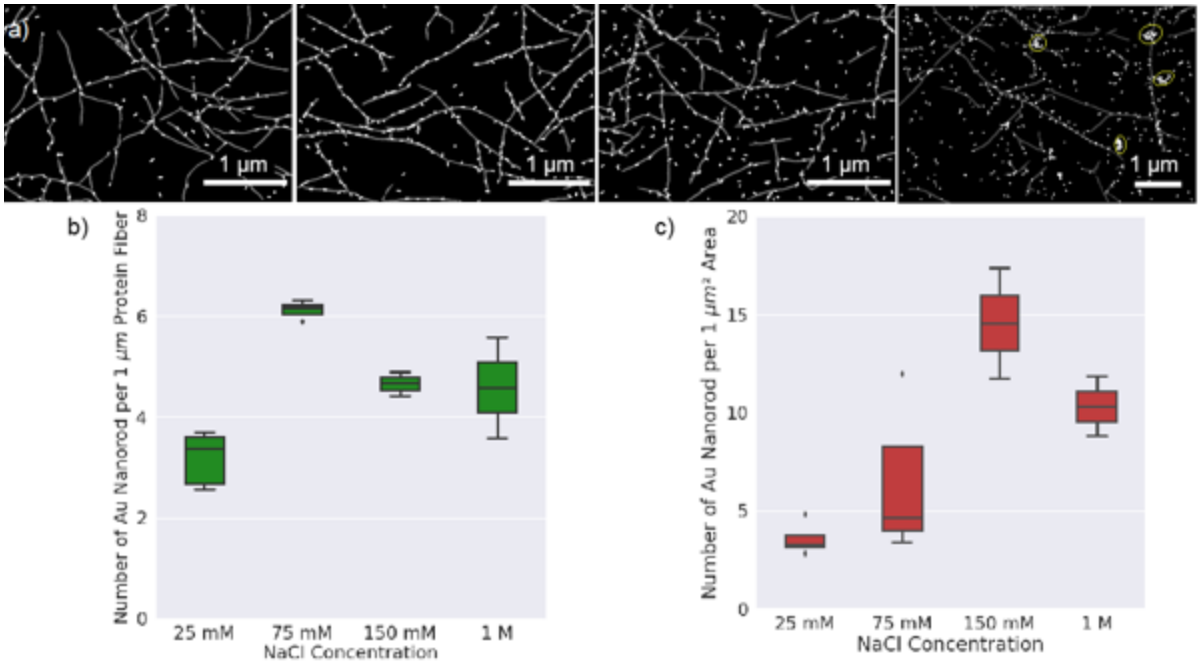

Figure 2-3 Analysis of the SEM images at different NaCl concentration. a) Processed and labelled SEM images at different NaCl concentration from 25 mM to 1M (from left to right). The gray lines show protein fibers and the white particles show the Au nanorods. The aggerated Au NRs were labelled with yellow lines. b) Boxplots show the results from the images analyzed with our automated image analysis tool. The x-axis shows the NaCl concentration whereas the y-axis (the green bins) shows the number of specific Au nanorod attachment per 1 μm protein fibers c) Boxplot of results the above images from the automated image analysis tools. The y-axis (the red bins) shows the number of nonspecific Au nanorod attachments per the 1 μm^2 substrate.

Figure 2-3 (b) shows the number of Au nanorods specifically bound to protein fibers at varying salt concentrations. We quantified the results from the regenerated images as done by the automated image tool. To get statistically-robust results, we analyzed multiple regions on the substrates ensuring that the total number of analyzed particles is more than 500 particles for each sample. The x-axis shows the NaCl concentration while the y-axis (the green bins) shows the number of Au nanorods specifically attached per 1 μm protein fibers. The average length and width for the Au nanorods from this solution are 45 ± 9 nm and 19 ± 5 nm, respectively. Thus, the

maximum loading density of 45 nm-long Au nanorods is 22 nanorods (22 = 1 μm/45nm) per micron of nanofiber if the Au nanorods were to stack perfectly head-to-tail along the protein fibers. We constrained the density to one particle per section of the protein fiber, because the protein fiber and Au nanorods have similar diameters (~15 nm), however in other work with larger fibers it is possible to assemble multiple nanorods on fibers of sufficient width.[36] The highest specific Au attachment to protein fibers was observed when the protein fiber solution is prepared in 75 mM NaCl concentration. As the salt concentration increases, the number of specifically attached Au nanorods first increases at moderate concentration (75 mM NaCl) to maximize the specific electrostatic attraction and then decreases at higher concentration (150 mM and 1 M NaCl) within the limits of experimental uncertainty, we propose this trend can be explained by a balancing of competing factors. When the ionic strength is too low the nanorods can repel each other, and also see repulsion from the background substrate, reducing the attachment density. At intermediate salt concentrations, the rods are screened from all but the shortest-range interactions and can achieve a higher density of attachment to the fibers, while above 75mM, all electrostatic forces are screened too greatly, resulting in poor particle-substrate repulsion and poor particle-fiber attraction, and leading to non-specific van der Waals attraction as the driving force of assembly. Here, we see that the maximum value reaches 6 for the 75mM NaCl solution, indicating that the nanorods are occupying just under 1/3 of the available protein surface.

Figure 2-3c shows the quantitative results for non-specific attachment of Au nanorods (NRs) to the silane-treated ITO substrate. In this case, the y-axis shows the number of attached Au particles per area (1 μm^2) and the area for each image is 9 μm^2. The number of nonspecifically attached Au nanorods (shown in red bins) increases as the salt concentration increases. However, the number of attached Au nanorods decreases at the highest salt concentration (1 M NaCl), most likely

because at such high salt concentrations, the Au nanorods quickly aggregate and settle down to the bottom of the reaction chamber due to the gravitational force. We did not observe any aggerated Au nanorods on the substrate except when the protein fiber solution is prepared in 1 M NaCl. The aggerated Au nanorods were labelled with yellow lines in Figure 2-3a).

We also prepared different Au nanorod solutions with different aspect ratios ranging from 2 to 5, (see Materials and Methods section, Table 2-1) to examine the effects the aspect ratio of Au on the attachment process. While we hypothesized that the aspect ratio of the nanorod might play a role, we observe similar behavior in all cases over the size range that was readily accessible experimentally. (See Figure A-4 and Figure A-5).

Finally, after analyzing the attachment density of the Au nanorods, we further explored how the attachment angle of the nanorods with respect to the protein fibers axis and the inter-rod distance changed with the solution salt concentration. The automated image analysis tool creates a sub image for each individual gold particles and then finds the best fit line for the orientation of the protein fibers in the sub image using Hough Line Tranform.[60] It then calculates the angle at the intersection between Au nanorod and protein fiber for each particle. Figure 2-4 shows the resulting normalized angular distributions of the attached Au nanorod to the protein fibers. We clearly see that at 25 mM, 75 mM salt concentration, the Au nanorods are preferentially aligned with the nanofibers (with nearly 60% of the nanorods aligned with less than a 20° angle with respect to the particle axis). However, when we increase the salt concentration, both mean and median angles of the distribution increase and the spread in attachment angles of the Au nanorods also increases, indicating that attachment becomes more random and less specific at higher salt concentration. By the time we reach 1 M salt concentration, the probability of the attachment angle of the Au nanorod to the protein fiber is almost equal and there is no favorable attachment angle.

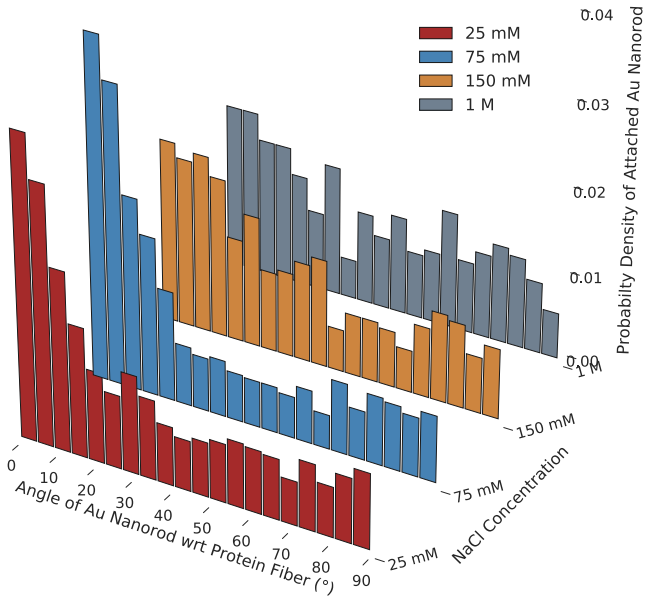

Figure 2-4 Plot of angular distribution of Au nanorods attached to the protein fiber. The x-axis shows the angle of Au nanorods with respect to the protein fiber, the y-axis shows the different salt concentrations, and the z-axis shows the probability density of attached Au NRs. The 25 mM and 75 mM salt concentrations both show peaks at narrow angles < 20 degrees, while at higher salt concentration the attachment becomes more random.

We calculate pair distribution functions (PDFs) for the Au nanorods as a function of ionic strength in Figure A-7. Notably, the PDF at lower salt concentrations has a peak at smaller distances, which we interpret as reflecting the preferential assembly along the nanofibers at low ionic strength leading to more closely spaced rods. At higher salt concentration that random distribution of rods across the entire surface leads to a shift to larger distances, while at very high (1M) salt the extremely high density of random attachment leads to a shift back to smaller values.

Both results from Figure 2-4 and Figure A-7 are consistent with our observations of a transition from specific electrostatic attachment to the protein fibers at low salt (25 mM and 75 mM) to non-

specific binding to the substrate at higher salt (150 mM and 1 M). At low salt concentrations, the Au nanorods tend to maximize the favorable electrostatic attraction between the length of the rod and the length of the fiber by aligning the same direction, whereas at higher salt concentrations, they are randomly aligned and really have little preference for the protein fiber over the substrate. It is possible that at lower salt concentrations the particles are able to move more freely to find the orientation of lowest potential energy whereas at high ionic strength, with the Coulomb repulsion terms that prevent nanorod attachment to the glass substrate screened, nanorods tend to stick where they first attach, with more random orientations.

2.4 CONCLUSIONS

In summary, we show that *de novo* designed proteins with high surface charge densities can be used successfully to assemble, and align, Au nanorods. We further show that the salt concentration can be tuned to achieve ordered and disordered assemblies. The optimum salt concentration of 75 mM for these nanofibers likely achieves as balance between particle-particle and particle-substrate interactions, without resulting in significant particle aggregation and non-specific binding. Importantly, we developed, and have made publicly available,[61] image analysis tools utilizing a fully convolutional neural network to analyze experimental data and gain geometrical information from each nanorod and nanofiber. This work is an important step towards reliable self-assembly of functional inorganic building blocks along these designer protein templates. Future studies may explore alternative de novo designed proteins combining both electrostatic and covalent interactions, as well as seek to explore the emergent optoelectronic properties of these systems. Develop optical screening methods that allow particle assembly to be analyzed in real time and *in situ* during assembly, without requiring *ex situ* electron microscopy analysis would also be valuable.

2.5 MATERIALS AND METHODS

2.5.1 *Materials*

$HAuCl_4.3H_2O$, ≥99.9% trace metals basis, and L-Ascorbic acid, BioXtra, ≥99.0% cyristalline, were purchased from Sigma Aldrich. Hexadecyltrimethylammonium bromide (CTAB) , >98.0%, and 5-Bromosalicylic acid(5-BromoSA), >98.0%, were purchased from TCl America. $AgNO_3$, 99.9%-Ag, and $NaBH_4$, 98%, were purchased from Stream Chemicals.

2.5.2 *Synthesis of Au Nanorods*

The synthesis of Au nanorods were followed by using literature procedure[62]. The seed solution for Au nanorods was prepared as reported previously. A 5 mL amount of 0.5 mM $HAuCl_4$ was mixed with 5 mL of 0.2 M CTAB solution. A 0.6 mL portion of fresh 0.01 M $NaBH_4$ was diluted to 1 mL with water and was then injected into the Au(III)-CTAB solution under vigorous stirring (1200 rpm). The color of the solution was changed from yellow to brownish-yellow, and the stirring was stopped after 2 min. The seed solution was aged at room temperature for 30 min before use. To prepare the 25-ml of growth solution, 0.9 g of CTAB together with 5-bromosalicylic acid, 0.11 g were dissolved in 25 mL of warm water (50-70 °C) in a 125 mL Erlenmeyer flask. The solution was allowed to cool to 30 °C, when a 4 mM $AgNO_3$ solution as detailed in Table S1, was added. The mixture was kept undisturbed at 30 °C for 15 min, after which 25 mL of 1 mM HAuCl4 solution and, if necessary, a small amount of HCl (37 wt % in water, 12.1 M) was added. After 15 min of slow stirring (400 rpm),0.2 ml 0.064 M ascorbic acid (Table 2-1) was added, and the solution was vigorously stirred for 30 s until it became colorless. The growth solution had a CTAB concentration of about 0.05 M and was used right after preparation. Finally, 0.8 mL of seed solution was injected into the growth solution. The resultant mixture was stirred for 30 s and left

undisturbed at 30 °C for 12 h for Au nanorod growth. The reaction products were isolated by centrifugation at 8500 rpm for 25 min followed by removal of the supernatant. The precipitates were re-dispersed in 10 mL of water.

Table 2-1 Initial amounts of each precursor for the synthesis of Au nanorods.

Sample name	Seed solution (ml)	4 mM of Ag solution (ml)	12.1 M of Conc. HCl (μl)	Au Seed solution (ml)	CTAB (g)	5-BromoSA (g)	0.064 M of L-Ascorbic acid (ml)
AR=2	0.06	1.2	0	25	0.9	0.11	0.2
AR=2.5	0.04	1.2	0	25	0.9	0.11	0.2
AR=4	0.04	1.2	420	25	0.9	0.11	0.2

2.5.3 *Synthesis of De novo designed protein fibers*

The *de novo* designed protein fibers were synthesized were followed by using literature procedure[63].

2.5.4 *Assembly of Au nanorods and protein fibers*

Silane coated ITO was washed with ethanol then dry with N_2 blower. Protein Fibers (PF) (18 μM) were diluted 5 times with Tris buffer solution (25 mM Tris and 75 mM NaCl, pH 8). 20 mL, 4 μM of Protein Fiber (PF) was drop casted on silane coated ITO and washed with 500 ml of same solution for dilution and then with 500 ml of water gently and dried by hand shaking. A glass side with chamber was filled with Au NRs (check Table 2-1). 70 mL of Au NRs solution were diluted with 1000 ml of water and waited 30 min. To remove excess CTAB, centrifugated 5 min at 14000 rpm, removed supernatant part and resuspended until 70 mL of total volume.

Silane ITO with PF were put on the top of the chamber (upside-down, see Figure 1) and waited 5 min and then gently washed with water and dried with N_2 blower for further characterization.

2.5.5 *Scanning Electron Microscopy*

The SEM images were obtained using TFS Apreo-S with Lovac Scanning Electron Microscope operating at 2kV and 13 pA.

2.5.6 *UV-Vis Spectroscopy*

The UV-Vis data were obtained using an Agilent 8453 UV−vis spectroscopic system. The spectra were collected from 400 nm to 1000nm, step size 2 nm.

2.5.7 *Zeta Potential*

The zeta potential measurement was performed using a Malvern Zetasizer with a 633 nm laser. All measurements were carried out at 25 °C and the results reported are averages of three readings.

2.5.8 *Data Analysis*

All images were analyzed using the automated image analysis tool. This Jupyter notebook is publicly available at https://github.com/yamanmy/Automated_image_tool_for_Au_PF_image.

2.6 ACKNOWLEDGEMENTS

This material is based upon work supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS--The Center for the Science of Synthesis Across Scales under Award Number DE-SC0019288. SEM imaging was conducted at the University of Washington Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure (NNCI) site which is supported in part by the National Science Foundation, the University of Washington, the Molecular Engineering and Sciences Institute, and the Clean Energy Institute. M.Y. acknowledges support from the University of Washington Clean Energy Institute and the National Science Foundation Research

Traineeship under Award NSF DGE-1633216. D.S.G. acknowledges support from the University of Washington, Department of Chemistry Kwiram Endowment.

Chapter 3. LEARNING AND PREDICTING PHOTONIC RESPONSES OF PLASMONIC NANOPARTICLE ASSEMBLIES VIA DUAL VARIATIONAL AUTOENCODERS

Adapted with permission from Muammer Y. Yaman, Sergei V. Kalinin, Kathryn N. Guye, David Ginger, Maxim Ziatdinov. "Learning and predicting photonic responses of plasmonic nanoparticle assemblies via dual variational autoencoders" Small 2023, 19, 220589, DOI: 10.1002/smll.202205893 Copyright 2023, John Wiley and Sons.

3.1 OVERVIEW

We demonstrate the application of machine learning for rapid and accurate extraction of plasmonic particles cluster geometries from hyperspectral image data via a dual variational autoencoder (dual-VAE). In this approach, the information is shared between the latent spaces of two VAEs acting on the particle shape data and spectral data, respectively, but enforcing a common encoding on the shape-spectra pairs. We show that this approach can establish the relationship between the geometric characteristics of nanoparticles and their far-field photonic responses, demonstrating that we can use hyperspectral darkfield microscopy to accurately predict the geometry (number of particles, arrangement) of a multiparticle assemblies below the diffraction limit in an automated fashion with high fidelity (for monomers (0.96), dimers (0.86), and trimers (0.58). This approach of building structure-property relationships via shared encoding is universal and should have applications to a broader range of materials science and physics problems in imaging of both molecular and nanomaterial systems.

3.2 INTRODUCTION

Assemblies of plasmonic nanoparticles lead to emergent photonic behaviors with applications in fields as diverse as energy harvesting[64–69], medical diagnosis[70–72], sensing[73–77], and catalysis.[78,79] Directed assembly of plasmonic nanoparticles has proven effective in synthesizing hierarchical materials with such desired functionalities.[80–83] Generally, minute changes in the assembly state of the nanoparticles may result in significant changes of the far-field optical response accessible to macroscopic characterization.[84–86] Hence, it is of interest to study whether the optical changes detected via far-field measurements can be used to probe and understand the assembly process, and whether we can predict the optical response of a material system with a known geometry without a complete first principles calculation.

While electron microscopy remains the standard method of probing the assembly process, and of characterizing the resulting structure,[87] the assembly-dependent optical properties of plasmonic nanoparticles also allow the assembly process to be probed directly, in solution,[88] and in near-real time[89] without the challenges or perturbations that may arise from electron beam irradiation.[90,91] Indeed, over the years interparticle plasmonic coupling has been used to probe the evolution of biological structure, particularly with plasmon rulers.[92,93] Going beyond simple dimers, 3D plasmonic rulers have also been proposed in order to probe 3D structure and even demonstrated with lithographically fabricated nanostructures.[94]

More recently, machine learning (ML) methods have emerged as powerful tools to predict and encode complex structure-function relationships, such as those that arise through the coupling of multiple plasmonic particles.[95–98] Conversely, it should also be possible to use ML methods to predict/reconstruct the sub-wavelength structural information encoded in a complex plasmonic scattering spectrum.[95,99–101] At the same time, applications of ML methods to experimental datasets

are limited by the relative scarcity of experimentally accessible data, which is often comparable or smaller than the intrinsic data dimensionality. This sparsity is a ubiquitous challenge facing scientific researchers, limiting the applicability of the classic "big data" supervised learning methods, and thus necessitating the development of novel approaches for building structure-property relationships with a small (compared to intrinsic data dimensionality) number of examples.

Here we develop and demonstrate an approach for connecting the structure-property relationships between far-field optical responses and local nanoparticle cluster geometries in systems of plasmonic gold nanoparticles. We base this approach on dual variational autoencoders for the imaging and spectral data with a shared latent encoding. This approach allows the prediction of the optical properties of a cluster from a known geometry, and also the prediction of the geometry of a cluster from its measured optical response. We apply these methods both to reconstruct and classify the predicted geometries of plasmonic nanoparticle clusters based on their scattering spectra as well as to predict the scattering spectra of resulting clusters.

3.3 RESULTS AND DISCUSSION

As a model system, we examine the plasmonic response of self-assembled, citrate-capped gold nanoparticle clusters, comprising primarily spherical particles with an average diameter of 100 nm. We assemble these gold nanoparticle clusters into different geometries using a salt-induced aggregation approach,[102] as described in detail in the Methods section. These assembled gold clusters were precipitated onto an indium-doped tin oxide (ITO) substrate. The samples were analyzed under both scanning electron microscopy (SEM) and hyperspectral darkfield microscopy. Due to their different geometries, we expect the nanoparticle clusters to generate unique scattering

spectra that depends on the number of particles and interparticle spacings, with some consideration to particle shape variation.[103–105]

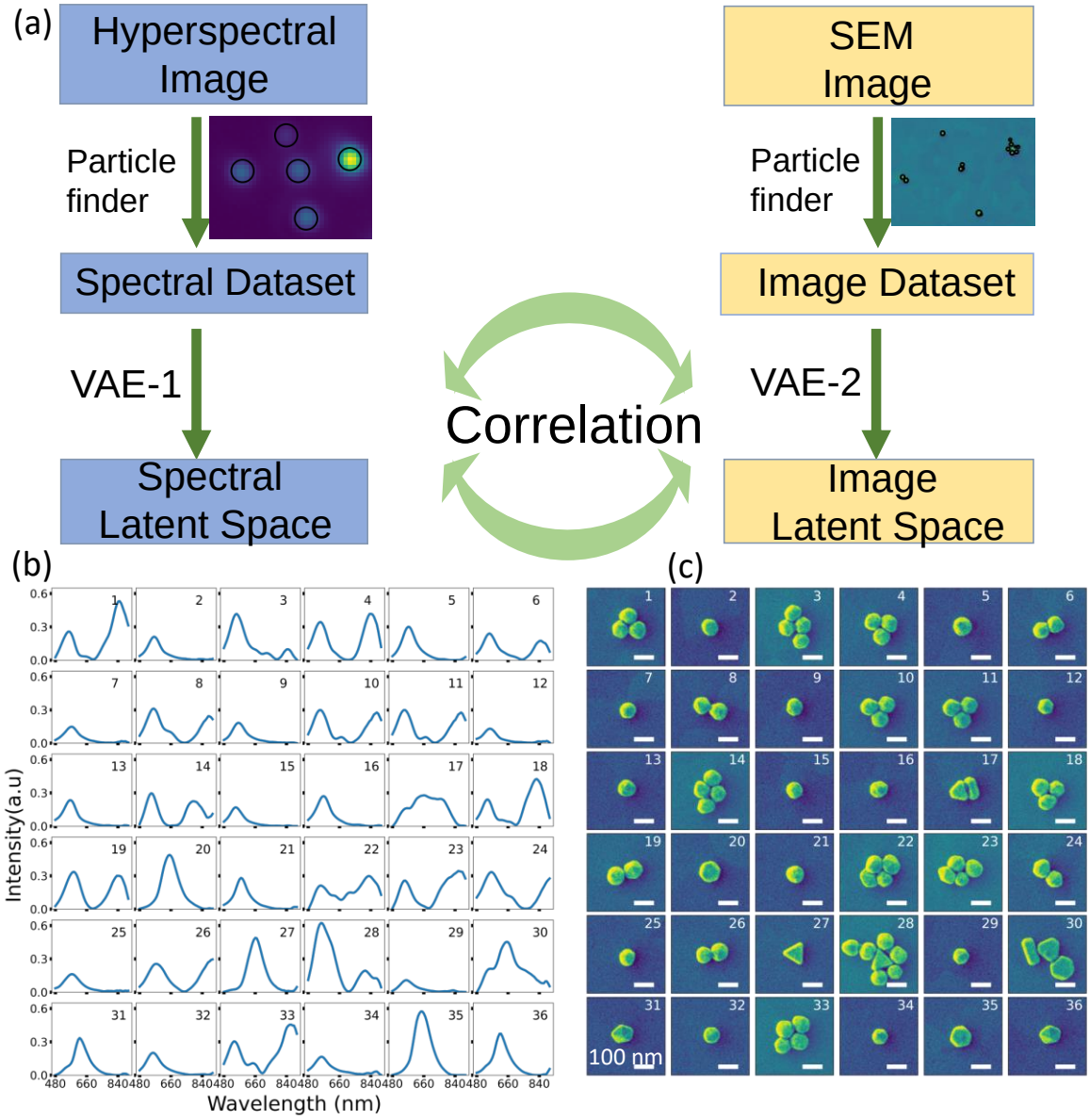

Figure 3-1 Representation of our workflow and datasets. (a) Schematic representation of the dataset identification and correlation workflow. (b) Representative spectral data set and (c) their corresponding SEM images. Sub image numbers are shown right top of each image.

Figure 3-1a depicts our basic approach. In order to relate the geometries of the clusters with their associated plasmonic responses, we collect both high-resolution SEM images, as well as spatially correlated hyperspectral images. By correlating the structures from SEM with spectra from hyperspectral images using the variational autoencoder, we aim to encode the structure-property relationships in the neural networks weights, thereby enabling a prediction of multimer structures based solely on scattering spectral data, and vice versa. To obtain the spectrum of each nanoparticle cluster, we start from identifying the positions of these clusters from hyperspectral images by applying a standard thresholding method, as described in SI section, and then extract the corresponding spectra data at each known position.

Figure 3-1b and Figure 3-1c show an example of scattering spectra and SEM images of the nanoparticle clusters respectively. The full dataset, methods, and algorithms are available for download via https://github.com/ziatdinovmax/dualVAE. We note that the stochasticity of the nanoparticle aggregation process leads to clusters with varying shape, size and orientation, and therefore with various optical responses (varying darkfield scattering spectra). Figure 3-1b, sub image 2, shows a scattering spectrum characteristic of single gold nanosphere which exhibits a single, Lorentzian plasmon scattering peak at 568 nm, as expected for a 100 nm particle on a glass substrate in air.[106] Indeed, the corresponding SEM image (Figure 3-1b, sub image 2) confirms the spectrum came from a single sphere. In contrast, Figure 3-1b, sub-image 6, shows a scattering spectrum exhibiting a transverse plasmon resonance peak at 560 nm and a longitudinal plasmon resonance peak at 850 nm. This dual peak is consistent with the expected scattering spectrum of a dimer of 100 nm gold nanospheres (see Figure B-1).[107] The SEM in Figure 3-1b/c sub-image 2 likewise confirms this cluster geometry assignment. For other dimers observed, the exact position

of the longitudinal peak depends on the interparticle spacing of the dimer as expected (Figure 3-1b/c sub-images 8 and 26).

For trimers and larger clusters, the spectra become increasingly complex (c.f. sub images 1, 4, and 30), again this increase in complexity is expected, since the size, shape, number and orientation of the gold nanoparticles all affect the resulting optical response.[108]

In our first analysis, we tried to independently learn the latent representations of the spectral and structural dataset, and subsequently identify the correlation between them. For the spectral data, we explored the applicability of the VAE with different invariances as we have applied previously for scanning probe microscopy[109] and electron energy loss spectroscopy datasets.[110] Figure B-2a/c shows the performance of the VAE in this case. However, we see the latent variables (z_1, z_2) for the spectral data cannot be differentiated from each other.

Due to the stochastic deposition of clusters, the image dataset necessarily comprises images with significant orientation disorder and possible offsets. In these cases, the standard ML approaches based on direct image analyses using linear decomposition methods, such as principal component analysis and standard deep convolutional neural networks, are known to fail.[111] To address this issue, we have previously implemented a family of the rotation-, shift-, and scale-invariant variational autoencoders for unsupervised-,[109,110,112–119] semi supervised-[120] and joint-learning[121] of disentangled latent representations. These VAE-based approaches allowing for the intrinsic physical invariances in the datasets, or imaging process, have successfully extracted order parameters in disordered systems including chemical transformations on the atomic level [111] and organization in protein nanoparticles,[112] as well as disentangled the domain switching mechanisms in ferroelectric materials,[113] suggesting this approach is fairly universal and could be fruitfully applied to the current problem.

For the structural data, we applied a variational autoencoders (VAE) with translational and rotational invariances to disentangle the latent representations. Figure B-2 b/d shows the performance of VAE with this structural dataset. One can see the importance of rotationally and translationally invariant VAE from the latent space distribution: they are differentiated based on the number of the particles in cluster ($z_{1,structure}$ first variable) and cluster shape ($z_{2,structure}$ second variable), respectively. However, the VAE model cannot separate the spectral dataset well, in Figure B-2 a/c. The first variable ($z_{1,spectra}$) of the VAE model has no change along horizontal axis, whereas the second variable ($z_{2,spectra}$) shows only a slight change in the peak at 840 nm. Furthermore, when we tried to correlate the latent spaces – using the spectral data to predict the structural data, or vice versa, we were unsuccessful and we found (see Figure B-2 e/f) that their latent representation showed no trend, meaning that these two datasets have no correlation.[122]

Because this straightforward approach of trying to correlate the *separate* latent spaces of the spectral and image spaces produces poor results, we next turn to develop the dual-VAE approach for the establishing the structure-property relationships. In a dual-VAE, there is a "communication channel" between latent spaces of two VAE models. Since the data is available in the form of image-spectrum pairs, this correlation allows for several opportunities to control the encoding process. As one such approach, we enforce similarity between structure and spectra latent spaces.

The implementation of dual-VAE is outlined in Algorithm 1. The dual-VAE model consists of two encoders and two decoders. At each training step, the first encoder (*encoder-1*) takes the image data to produce a latent vector, which is then passed through one of the decoders (*decoder-1*) to get a reconstruction of original data. The second encoder (*encoder-2*) takes spectral data to obtain its latent representation which is transformed via a learnable linear transformation before being reconstructed by the second decoder (*decoder-2*). The total loss function is given by:

$$\mathcal{L} = \mathcal{L}_{\mathrm{RE}} + \beta\mathcal{L}_{\mathrm{KLD}} + \gamma L^1$$

where $\mathcal{L}_{\mathrm{RE}}$ is a sum of reconstruction errors for decoded image and spectral data, $\mathcal{L}_{\mathrm{KLD}}$ is a sum of Kullback-Leibler divergence terms[123] between standard normal distribution and encoded distributions of image and spectral data, and L^1 is an L1 score between the latent representation of the *encoder-1* and transformed latent representation of the *encoder-2*. The coefficients β and γ are constant scale factors. The presence of a learnable linear transformation between the two latent representations and the loss term reflecting the 'closeness' of the two representations at each training step allows the alignment of the image and spectral latent spaces for establishing structure-property relationships in the system. We used a standard Adam optimizer for adjusting weights of all the encoders and decoders simultaneously in the end-to-end fashion.[124] The approach outlined in Algorithm 1 works both with standard VAEs and invariant VAEs.

Algorithm 1: Training of dual-VAE (single step)

Inputs: Two experimental datasets, X_1 and X_2.

Pass x_1 through *encoder-1* to get μ_1 and σ_1 parameters of variational distribution
Sample latent vector, $z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$
Compute KL divergence, D_1, between encoded and prior distributions
Pass z_1 through *decoder-1* to obtain x_1'
Compute reconstruction loss, RE_1, between x_1' and x_1

Pass x_2 through *encoder-2* to get μ_2 and σ_2 parameters of variational distribution
Apply a learnable linear transformation, A, such that $\mu_2' = A\mu_2$.
Sample latent vector, $z_2' \sim \mathcal{N}(\mu_2', \sigma_2^2)$
Compute L^1 score between two latent vectors, z_1 and z_2'
Compute KL divergence, D_2, between encoded and prior distributions
Pass z_2' through *decoder-2* to obtain x_2'
Compute reconstruction loss, RE_2, between x_2' and x_2

Compute total loss, $\mathcal{L} = (RE_1 + RE_2) + \beta(D_1 + D_2) + \gamma L^1$
Backpropagate loss and adjust weights in both VAE models

At the prediction stage, the alignment of latent spaces allows predicting spectra from images and vice versa. In this case, the images/spectra are encoded via encoder-1/encoder-2 and decoded into spectra/image via decoder-2/decoder-1 (Figure 3-2).

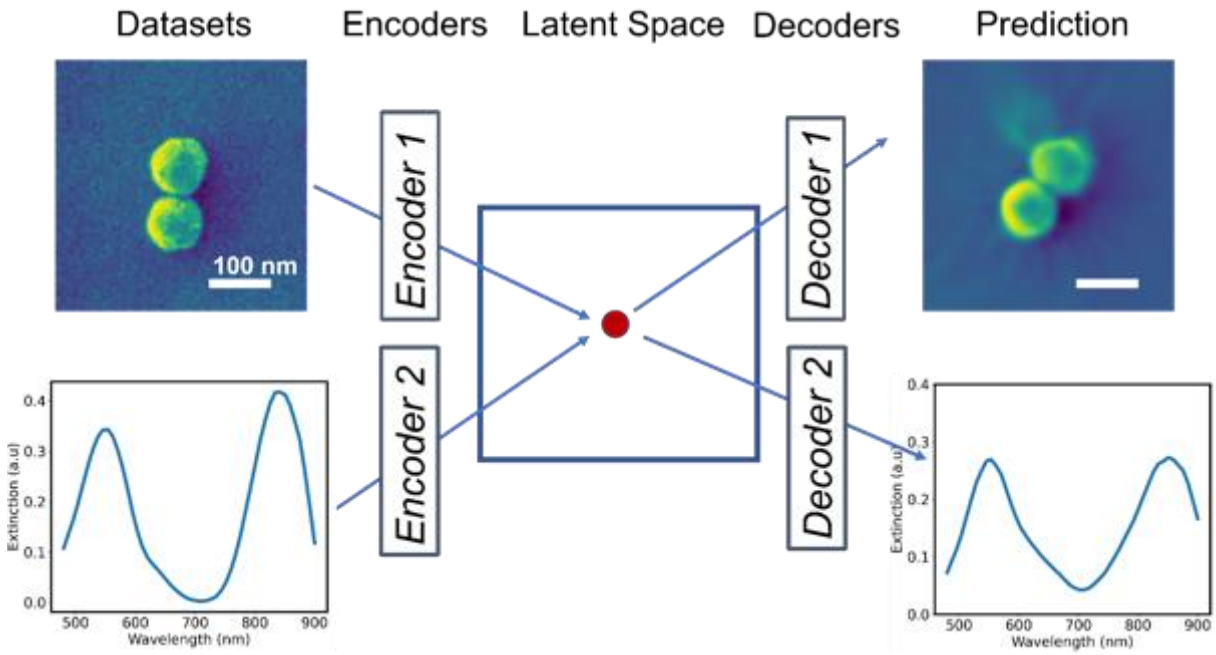

Figure 3-2 Schematic representation of the dual variational autoencoder use for prediction. The left side shows the inputs and right-side shows predictions via the shared latent space. The scale bars are 100 nm. Note that when using rotationally invariant VAE for spatial data, all particles of the same shape (e.g., dimers) are aligned in the same direction in the decoded data, independently of the input orientation.

As discussed above, we created an experimental dataset library using the structure (from SEM) and spectral properties (from hyperspectral imagery) information of the gold nanoparticle clusters. This library is constructed of 898 clusters (526 monomers, 197 dimers, 115 trimers, 60 larger clusters ($n \geq 4$ where n refers to the number of particles in a cluster)). We used randomly chosen 247 clusters (138 monomers. 58 dimers, 26 trimers, and 25 larger clusters) to train the dual-VAE model. Figure 3-3 shows the results generated from the dual-VAE. Here, the structural and spectral data of the clusters are displayed as circle- and star-shape points respectively. Each data point is

post-colored with red, green, blue and purple, representing the structure information of monomers, dimers, trimers, and larger clusters respectively. In contrast to the simple VAE approach, (Figure B-2), we can see that clusters consisting of different numbers of nanoparticles are well separated in both the structural and spectral latent spaces in the shared latent space, Figure 3-3a.

Figure 3-3b shows a plot of the first variable of the spectral latent space as a function of the first variable of the structural latent spaces. Importantly, this plot of $z_{1,spectra}$ vs. $z_{1,structure}$ shows a linear correlation, meaning that these variables are related. Since $z_{1,structure}$ encodes the number of particles in a cluster, the correlation of $z_{1,spectra}$ with $z_{1,structure}$ in Figure 3-3b implies that the dual-VAE with a shared latent space has likely learned that the spectra depend on the number of particles in the cluster, as our physical understanding of the system tells us is the case (indeed total scattering intensity should scale with the number of particles in the cluster). [125]

We test this assumption by decoding the latent space into spectral (Figure 3-3c) and structural (Figure 3-3d) manifolds. We note that: (1) in Figure 3-3e, the spectral intensity increases along the horizontal direction (from right to left) which, as expected, matches well with the increase in the number of nanoparticles comprising the clusters in Figure 3-3f. (2) in Figure 3-3e, the second peak around 800 nm become more obvious along the vertical direction (from bottom to top), which also corresponds well with the cluster shape change shown in Figure 3-3f. Notably then, the dual-VAE is able to distinguish not only the particle number in a cluster, but also different cluster shape in both spectral and structural datasets.

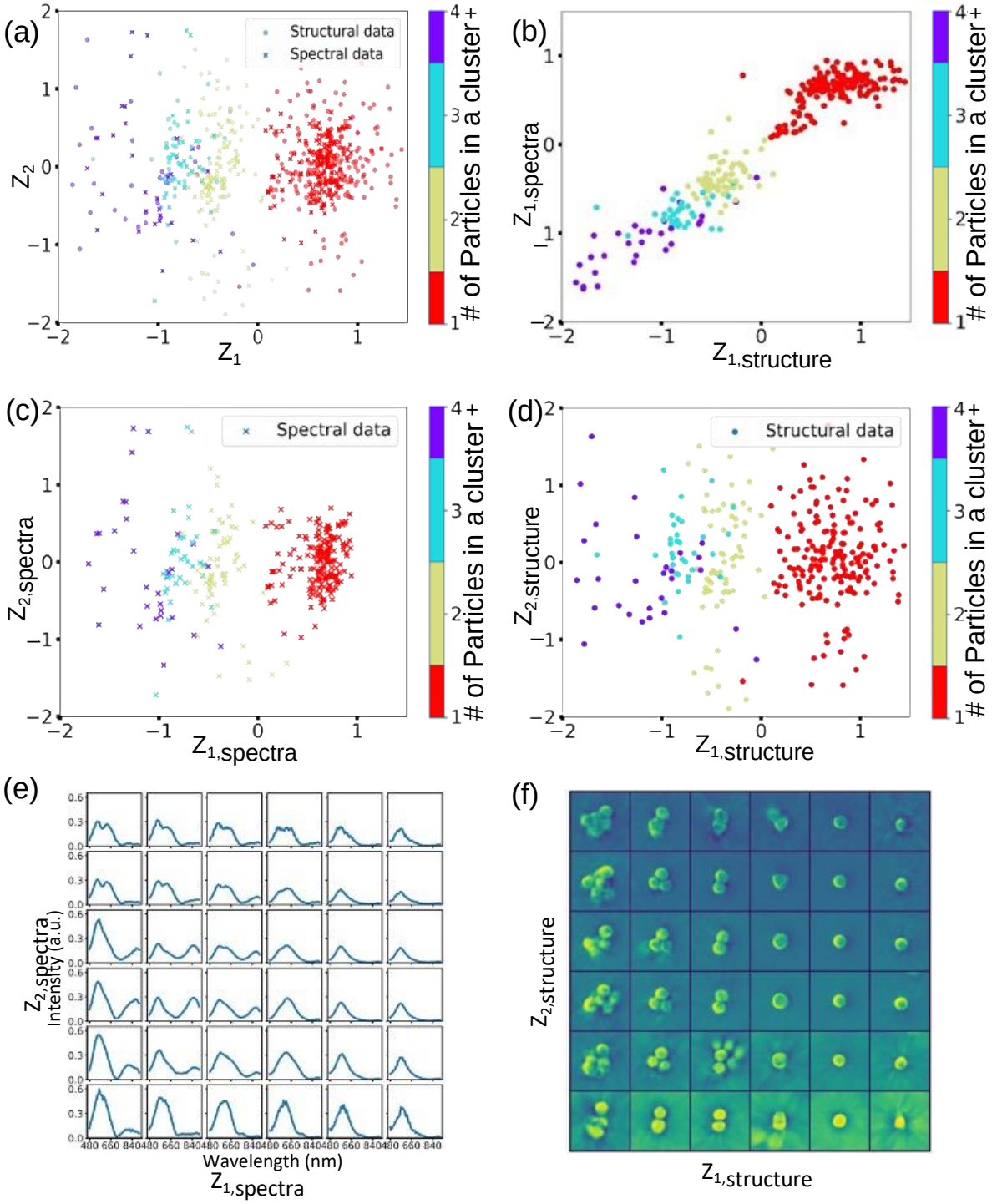

Figure 3-3 Performance of the dual-VAE model. (a) Combined latent space, with star- and circle-shape points representing the spectral and structural data, respectively. (b) First latent variables of structure and spectra datasets. Individual latent space of (c) spectral and (d) structural datasets. The learned manifold representation of (e) spectral latent space and (f) structural latent space.

Figure 3-3 suggests that the dual-VAE can predict the cluster geometry based on the hyperspectral data alone. To test the performance of the dual-VAE model at this task, we compare the model's predicted cluster geometry based on the spectra with the known ground truth as determined by the SEM of the actual cluster. Figure 3-4 shows a comparison for a small subset of 8 clusters. Figure 3-4a shows a color darkfield image of the selected region, while Figure 3-4b shows the correlated SEM image for each cluster in Figure 3-4a. Figure 3-4c shows the predicted geometry of the cluster from the dual-VAE. Given that the dimensionality of the structure data is much larger than the spectral data, and the number of training examples is smaller than the dimensionality of the data set, this task is ill-posed task from the ML perspective. Nevertheless, the reconstructions in Figure 3-4c are surprisingly accurate. Comparing the ground truth in Figure 3-4b and the reconstruction in Figure 3-4c, we note several points: (i) For monomers (particle 4 and 6), the prediction/decoding is nearly perfect. (ii) For dimers, the reconstruction/prediction of cluster 8 is nearly as perfect, albeit the reconstructed dimer is generated for a different orientation, reflecting the intrinsic invariances in the system. However, for cluster 7, the reconstructed image shows a trimer structure, rather than dimer, regardless of the very low intensity of the "added" particle. We attribute this behavior to the non-uniform contrast within the particles on the SEM image. (iii) The dual-VAE decoding still seems to work well for many trimers although the exact shape of each individual particle has not been preserved completely. However, the overall shape of the cluster is well maintained. (iv) The accuracy of the prediction further decreases for more complex tetramers and pentamers (cluster 2 and 3). This decrease may be ascribed to the complexity of particle geometry, the relatively small number of datasets for larger clusters ($n \geq 4$) used to train the model, and the fact that some details do not affect the photonic responses of particles significantly especially with bigger particles (i.e. the difference between a pentamer and

a tetramer is less significant than a monomer to a dimer or dimer to trimer). We consider the statistical performance of this approach more rigorously after we examine a subset of the inverse problem (predicting spectra based on the image).

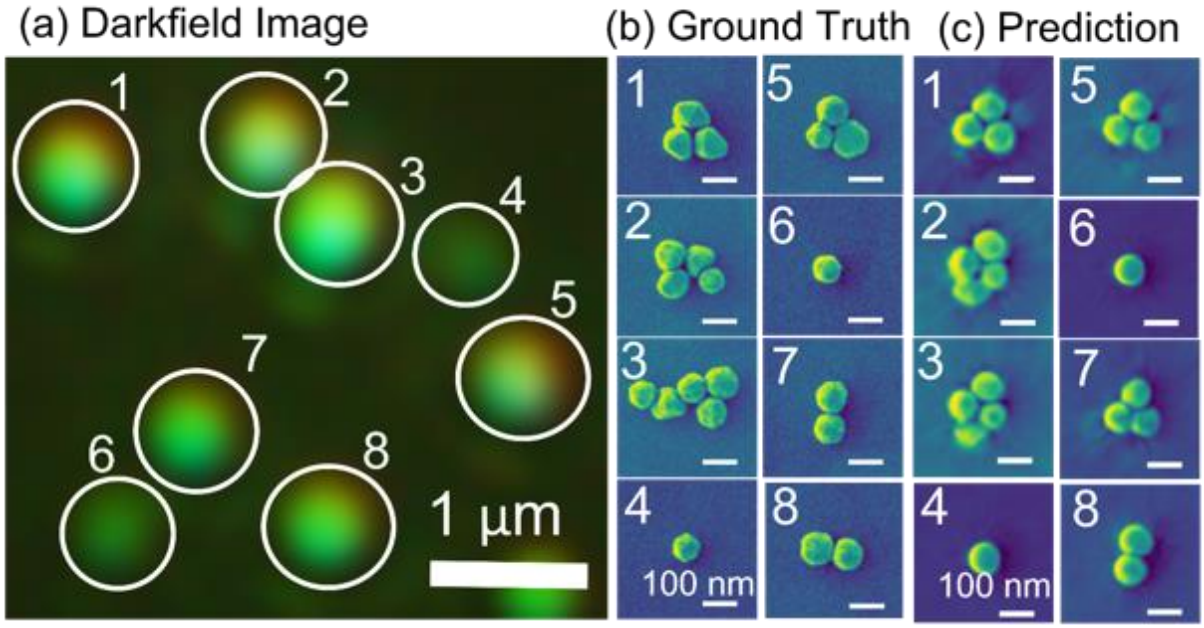

Figure 3-4 Examples of the dual-VAE model's performance on an area of the test datasets. (a) Darkfield scattering image. The (b) ground truth and (c) dual-VAE model prediction of each particle geometry based on the hyperspectral data. The scale bar is 1 mm in (a) and 100 nm in (b-c).

Although our main goal at present is reconstruction of the structural data based on the spectral data in order to determine the structure sub-diffraction-limit structures, we also used this dataset to evaluate the performance of dual-VAE model in predicting the spectral information from the structural data to determine the optical properties of the gold clusters based on their SEM images. As shown in Figure B-3, we compared the spectra predicted by the dual-VAE with the ground truth from the hyperspectral image. (i) For clusters consisting of spherical particles, cluster 3, 6, 8 for example, the dual-VAE model prediction works well. (ii) The dual-VAE model can achieve better prediction on the transverse resonance peak at around 560 nm than on the longitudinal

plasmon peaks, with the likely reason being that the longitudinal plasmon peak depends strongly on the distance between two particles, whereas the transverse resonance is relatively static, as we have already mentioned above. (iii) We point out that, in our case, the model still faces some challenges, as it is not easy to predict the spectra information (which, in essence, is determined by the cluster structures in three dimensions) out of SEM images that only display structure information in two dimensions. For example, cluster 1 and 2 consist of pentagonal and triangular shaped particles, while the model predicts them as spherical. The reason is that these particles overlap with other particles at the edges, thus demonstrating themselves as spherical in two-dimensional SEM images. Precise prediction of the spectra without knowing the exact 3D shape is a difficult hard task even for a human, therefore, we think that the dual-VAE model predictions are promising.

While Figure 3-4 examines a typical region of the image to help visualize the performance of the dual-VAE, we also examined the fidelity of the reconstructions with more statistical rigor. To examine the reliability of our dual-VAE model, we quantitatively analyzed the error between the predicted structure (Figure 3-5) / spectra (Figure B-4) against the ground truth of the clusters as determined by examining the corresponding SEM. We started from calculating the accuracy of the predicted structural data by comparing the predicted number of particles constituting each cluster with the ground truth in each cluster category. We have 651 clusters (388 monomers, 139 dimers, 89 trimers, and 35 tetramers) for testing and summarized the results in the confusion matrix shown in Figure 3-5. For example, out of 388 ground truth monomers, the model correctly identified 374 (374/388=0.96) objects as monomers, while the remaining 14 (14/388=0.04) were incorrectly classified as dimers. Figure 3-5a shows that model achieves better prediction accuracy for monomers (0.96), and dimers (0.86), than it does for trimers (0.58), and tetramers (0.11). In the

case it fails, the dual-VAE model sometimes predicts a dimer, when it should be trimer, or a trimer, when it should be tetramer.

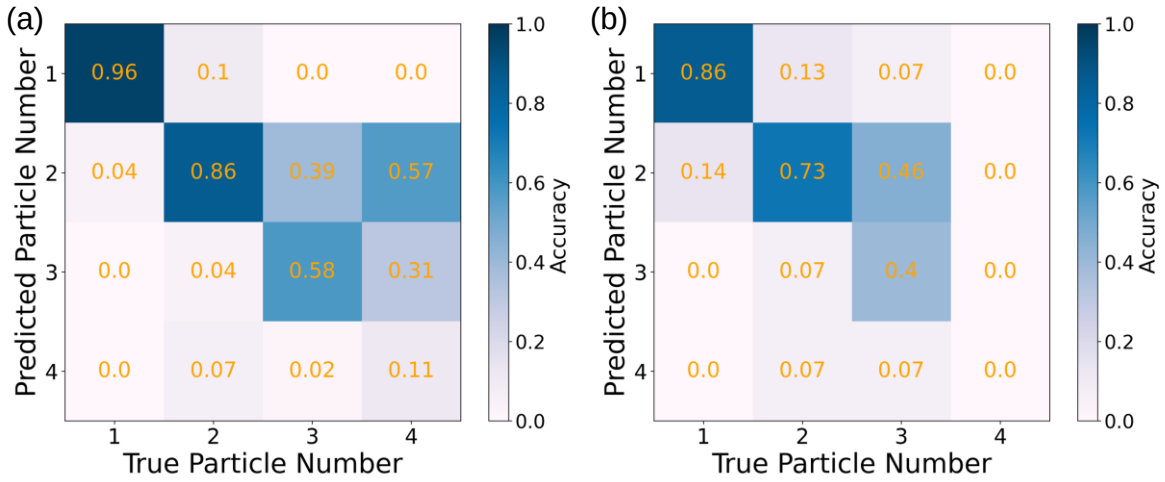

Figure 3-5 Accuracy of the dual-VAE structural prediction. Our model confusion matrix for structure prediction is based on particle numbers in a cluster in (a) our datasets, and (b) sampled literature datasets. Predicted particle numbers are our model prediction.

We also tested our dual-VAE model against literature data (15 data for each class and 3 different sources [126–130]). Here, for literature data, the model prediction accuracy is 0.86 for monomers, 0.73 for dimers and 0.4 for trimers (See Figure 3-5b). Going beyond simply classifying particles as monomers, dimers, trimers, etc., we also use the structural similarity index (SSI)[131] to evaluate the quality of model reconstruction. The Appendix B includes both the confusion matrix and the SSI results on images and shows that this dual-V AE model can predict the structural information with high precision.

3.4 Conclusions

Over, the dual-VAE represents a promising machine learning approach to build correlative relationships between the nanoparticle cluster geometries and their optical properties, with the particular approach of identifying sub-diffraction-limit particle geometries based solely on

diffraction-limited far-field hyperspectral image data. Enforcing similarity between the latent representations of the structural and spectral variational autoencoders during the training appears to be a key factor in successful reconstructions based on the shared latent space. This work assumes that the encoded structure-property relationships can be approximated by a linear function, but the approach can be easily extended to more complicated dependencies.

We note that this approach of building structure-property relationships via common encoding is universal and can be applied to a broader range of materials science and physics problems in imaging and building structure-functionality relationships. The latter includes imaging/spectroscopic data emerging in the context of techniques such as (scanning) transmission electron microscopy – electron energy loss spectroscopy, the gamut of scanning probe microscopies and spectroscopies including atomic force microscopy, piezoresponse microscopy, and scanning tunneling microscopy, and chemical and optical imaging. Beyond imaging studies, this approach can be applied in molecular systems e.g., for the joint encoding of structures parametrized via graph, SELFIES, or SMILES representations and corresponding property vectors. In the case of optical systems, we anticipate that in the future it should be possible to reconstruct both particle orientations and interparticle distances within clusters, both by expanding the data collected, as well as expanding the training sets, likely with a concomitant expansion of the size of the latent space.

3.5 MATERIAL AND METHODS

3.5.1 *Materials*

Citrate-coated gold nanoparticle solution with diameters of 100 nm, was purchased from BBI solutions.

3.5.2 *Sample Preparation*

Suspension of gold solution were precipitated out in 25 mM NaCl solution onto indium-doped tin oxide (ITO) glass. The process was watched under darkfield scattering microscopy to ensure the particles are far from each other and there is no spectral overlapping. The ITO glass with the gold particles on top was carefully analyzed under both scanning electron microscope (SEM) and hyperspectral darkfield scattering microscope.

3.5.3 *Scanning Electron Microscopy*

The SEM images were obtained using TFS Apreo-S with Lovac Scanning Electron Microscope operating at 2kV and 13 pA.

3.5.4 *Hyperspectral Darkfield Scattering Microscopy*

Hyperspectral darkfield scattering image were collected on a Photon Etc. IMA system, using a Nikon Ni-U upright microscope and a 40x objective lens (Nikon, NA 0.6). Full-field scattered light was passed through the tunable volume Bragg grating filter with a step size of 2 nm and integration time of 2 seconds/step. The scattered light at each wavelength step was collected by a CCD camera (Thorlabs, 1501M-US-TE), building the hyperspectral data-cube by accumulating each wavelength slice from 400 nm to 1000nm. Scattering spectra were normalized by dividing by a white light reference spectrum and then background corrected. We used the Photon Etc. PHySpec software for the calculation. The rest of analysis was done on Google Colab Jupyter notebook.

3.5.5 *Data Analysis*

The detailed methodologies of dual VAE analysis on structure and spectra data set are established in Jupyter notebooks and are available from https://github.com/ziatdinovmax/dualVAE.

3.6 ACKNOWLEDGEMENTS

This work was supported (M.Y.Y., K.N.G., D.G., S.V.K.) by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS–The Center for the Science of Synthesis Across Scales–under Award NumberDE-SC0019288, located at University of Washington and performed (M.Z.) at Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS), a U.S. Department of Energy, Office of Science User Facility. M.Y.Y. and D.S.G. acknowledge support from CNMS user facility, project number CNMS2021-B-00847. SEM imaging was conducted at the University of Washington Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure (NNCI) site which is supported in part by the National Science Foundation, the University of Washington, the Molecular Engineering and Sciences Institute, and the Clean Energy Institute.

.

Chapter 4. PREDICTING PHOTONIC RESPONSES AND ORIENTATION OF PLASMONIC NANORODS VIA 3D - DUAL VARIATIONAL AUTOENCODERS

4.1 INTRODUCTION

Assemblies of plasmonic gold nanorods display unique photonic properties suitable for various applications, including energy harvesting, medical diagnostics, sensing[132], and catalysis. Utilizing directed assembly methods for these nanorods enables the creation of complex hierarchical materials with the desired functionalities. Notably, even subtle variations in the assembly of these nanorods can lead to marked alterations in their far-field optical responses that can be detected macroscopically[133,134]. It's intriguing to explore if these optical variations, observed through far-field measurements, can offer insights into the assembly dynamics and whether it's feasible to forecast the optical behavior of a system with a predefined geometry without delving into comprehensive primary calculations.

While electron microscopy remains the go-to technique for inspecting the assembly process and deciphering the subsequent structures, the inherent optical traits of these plasmonic gold nanorods offer a direct means to examine the assembly in solution and in near-real time. This bypasses any potential issues or disruptions resulting from electron beam exposure. Historically, the coupling between individual plasmonic particles has been instrumental in investigating biological structures, especially with the aid of plasmon rulers. Advanced 3D plasmonic rulers have even been devised to inspect 3D formations, as showcased in lithographically engineered nanostructures.[8,135]

Recently, the advent of machine learning (ML) has provided potent tools to decode and anticipate intricate structure-function dynamics, especially those emerging from the interaction of multiple plasmonic gold nanorods. In parallel, ML can potentially be harnessed to predict or decipher the intricate structural details encapsulated in the scattering spectra of these plasmonic entities. However, the practical application of ML in analyzing experimental datasets is constrained by the limited availability of data, which is typically on par or even less than the inherent data dimensionality. This limitation poses challenges for researchers and underscores the need for innovative strategies to elucidate structure-property correlations using limited data samples.

In this context, we introduce and validate a technique to bridge the structure-property correlations between far-field optical signals and local nanorod cluster configurations in plasmonic gold nanorod systems. Our method leverages dual variational autoencoders for image and spectral data combined with a mutual latent encoding. This allows for predicting the optical attributes of a cluster based on its known structure and vice versa. We employ these techniques to both reconstruct and categorize the anticipated structures of nanorod clusters from their scattering spectra and to forecast the scattering spectra of the resulting configurations.

4.2 RESULTS AND DISCUSSION

In our chosen model system, we study the plasmonic behavior of a self-assembled, citrate-capped individual gold nanorod, which is primarily rod-shaped with an aspect ratio ranging from 1 to 2.5 (see Figure C-1). We position this single gold nanorod using a salt-induced placement technique, elaborated further in the Methods section. We drop-casted gold nanorod was then deposited onto an indium-doped tin oxide (ITO) platform. We subjected this sample to both

scanning electron microscopy (SEM) and hyperspectral darkfield microscopy for analysis. Given its shape and positioning, we anticipate the gold nanorod to produce a specific scattering spectrum, influenced by its rod-shaped morphology and surrounding environment.

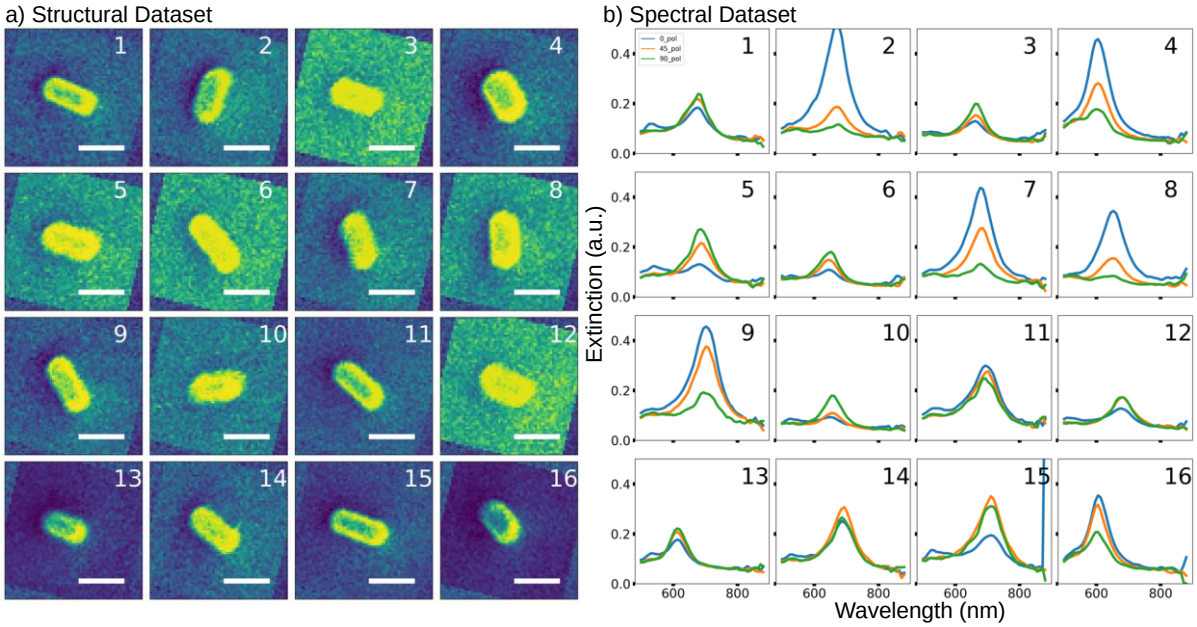

Figure 4-1 Selected datasets for correlated structural image and spectral datasets. (a) Representative SEM images of Au nanorods with different aspect ratio and orientation. Sub image numbers are shown top right of each image. (b) Corresponding polarization spectral dataset with different polarized light. Blue, orange, and green lines show a spectrum under 0-, 45- and 90- degree polarized light.

Figure 4-1 presents our partial datasets and Figure C-2 and Figure C-3 show widefield view of correlated datasets. To connect the configurations of individual nanorods to their specific plasmonic responses, we acquire both detailed SEM images and spatially related hyperspectral imaging with polarized light (see the experimental setup in Figure C-4). By associating SEM structures with the hyperspectral data using the variational autoencoder, our objective is to encapsulate the structure-property dynamics within the neural network's weightings. This enables the anticipation of single nanorod configurations based on merely scattering spectral information,

and the other way around. To ascertain the spectrum of each individual nanorod, we pinpoint their locations from hyperspectral visuals through a conventional thresholding technique, elaborated further in the Appendix C section, and subsequently retrieve the pertinent spectral information from recognized locales.

Figure 4-1a and Figure 4-1b exhibit an instance of scattering spectrums and SEM visuals of the gold nanorods, respectively. Comprehensive data sets, methodologies, and algorithms can be accessed at https://github.com/yamanmy/dual3DVAE . It's noteworthy that the unpredictability inherent in the nanorod positioning process yields nanorods of diverse shapes, dimensions, and alignments, thus resulting in a range of optical feedbacks (diverse darkfield scattering spectrums). We further assessed our correlated structural and spectral datasets in Figure C-5. We calculated max peak intensity at 45° polarized light using intensity using 0° and 90° polarized light and shows linear correlation with experimental observed intensity. Also, we show the linear correlation between angles in structural datasets and spectral datasets in Figure C-5b.

When it comes to larger assemblies, the spectral representations grow progressively intricate. Such escalation in intricacy is foreseeable as the dimension, form, quantity, and orientation of the gold nanorods profoundly influence the optical output.[136] However, we mainly work individual nanorods to their specific plasmonic responses to simplify our approach.

In our initial analysis, we examined the suitability of the dual-VAE with various invariances for the spectral data, similar to our prior applications for scanning probe microscopy and electron energy loss spectroscopy datasets. The dual VAE's efficacy is illustrated in Figure C-6. However, it's evident that the two latent variables (z1, z2) for the spectral data and structural are indistinguishable.

Given the random deposition of gold nanorods, the image dataset naturally includes images with pronounced orientation disarray and potential offsets. Traditional ML techniques that rely on direct image analyses, such as principal component analysis and regular deep convolutional neural networks, have been proven ineffective in such situations. To mitigate this, we previously introduced a set of rotation-, shift-, and scale-invariant variational autoencoders for unsupervised, semi-supervised, and joint-learning of separated latent representations[110,115,118–120,137,138]. These VAE-centric methods, which account for the inherent physical invariances in the datasets or imaging process, have adeptly identified order parameters in chaotic systems. Examples include atomic-level chemical changes and organization in protein nanoparticles indicating the method's versatility and its potential applicability to our current challenge.

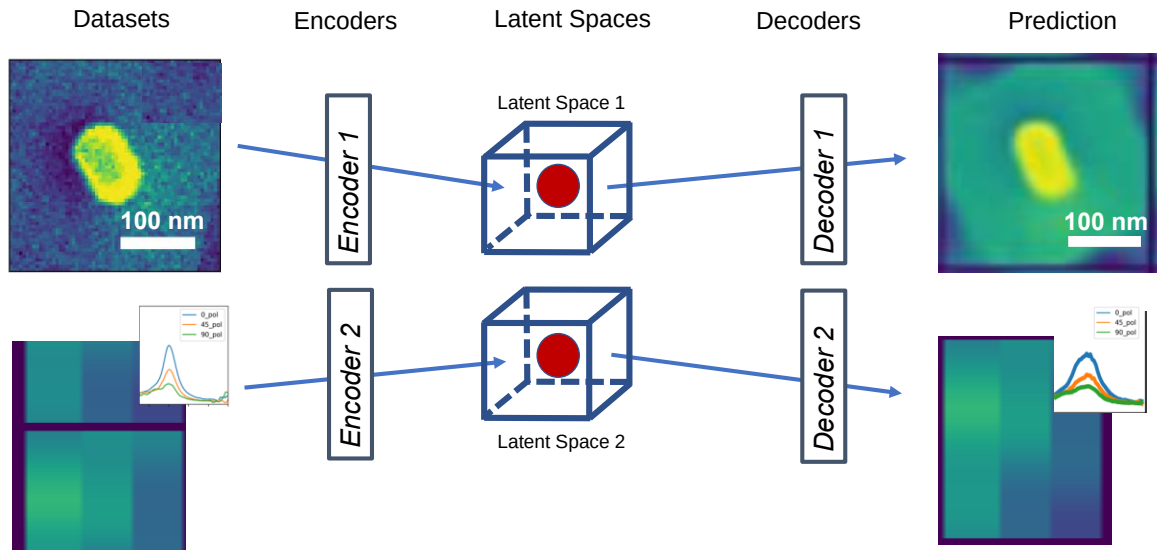

Figure 4-2 Schematic representation of the dual 3D variational autoencoder (VAE) use for prediction. The left side shows the inputs and right-side shows predictions via the 3D shared latent spaces. The scale bars are 100 nm. Note that polarized spectra are combined into 2D image as an input.

Algorithm 1 details the setup of the dual-3D-VAE model, which includes two encoders and two decoders. During each training phase, encoder-1 processes the image data to generate 3-dimensional latent vectors. These vectors are then fed into decoder-1 for a reconstruction of the original data. Encoder-2, on the other hand, handles 2-D spectra-image data (combined 3 different polarized spectrum) to derive its latent representation. This representation undergoes a modifiable linear transformation before decoder-2 reconstructs it. The total loss function is given by:

$$\mathcal{L} = \mathcal{L}_{\mathrm{RE}} + \beta\mathcal{L}_{\mathrm{KLD}} + \gamma\mathcal{L}^{1}$$

where $\mathcal{L}_{\mathrm{RE}}$ is a sum of reconstruction errors for decoded image and spectral data, $\mathcal{L}_{\mathrm{KLD}}$ is a sum of Kullback-Leibler divergence terms[123] between standard normal distribution and encoded distributions of image and spectral data, and $\mathcal{L}^{1}$ is an L1 score between the latent representation of the *encoder-1* and transformed latent representation of the *encoder-2*. The coefficients β and γ are constant scale factors (2 and 10, respectively).

Algorithm 1: Training of dual-VAE (single step)

Inputs: Two experimental datasets, X_1 and X_2.

Pass x_1 through *encoder-1* to get μ_1 and σ_1 parameters of variational distribution
Sample latent vector, $z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$
Compute KL divergence, D_1, between encoded and prior distributions
Pass z_1 through *decoder-1* to obtain x_1'
Compute reconstruction loss, RE_1, between x_1' and x_1

Pass x_2 through *encoder-2* to get μ_2 and σ_2 parameters of variational distribution
Apply a learnable linear transformation, A, such that $\mu_2' = A\mu_2$.
Sample latent vector, $z_2' \sim \mathcal{N}(\mu_2', \sigma_2^2)$
Compute L^1 score between two latent vectors, z_1 and z_2'
Compute KL divergence, D_2, between encoded and prior distributions
Pass z_2' through *decoder-2* to obtain x_2'
Compute reconstruction loss, RE_2, between x_2' and x_2

Compute total loss, $\mathcal{L} = (RE_1 + RE_2) + \beta(D_1 + D_2) + \gamma L^1$
Backpropagate loss and adjust weights in both VAE models

At the prediction stage, the alignment of latent spaces allows predicting spectra-images from images and vice versa. In this case, the images/spectra-images are encoded via encoder-1/encoder-2 and decoded into spectra-images/image via decoder-2/decoder-1 (Figure 4-2).

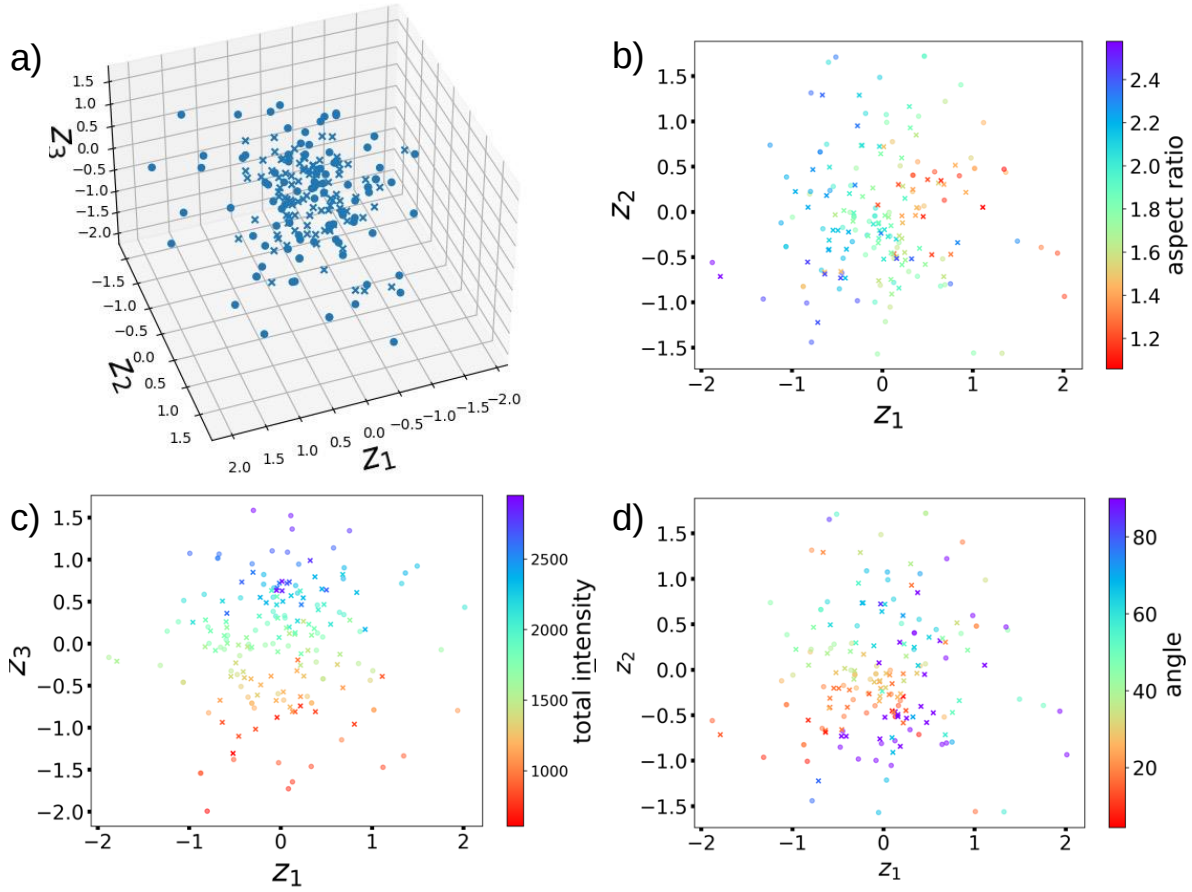

Figure 4-3 Performance of the dual 3D VAE model. (a) Combined 3D latent space, with cross- and circle-shape points representing the spectral and structural data, respectively. 2D representation of first latent variable (z_1) with respect to second latent variable (z_2) are in b) and d). 2D representation of first latent variable with respect to third latent variable (z_3) are in c). The data points represent one input and are post-colored based on aspect ratio in b), total intensity in c) and angle in d).

As discussed above, we created an experimental dataset library using the structure (from SEM) and spectral properties (from hyperspectral image) information of the gold nanoparticle clusters.

This library is constructed of 372 datasets. We used randomly chosen 279 datapoints (with varying aspect ratio) to train the dual-VAE model. Figure 4-3 shows the results generated from the dual-3D-VAE. Figure 4-3a shows combined 3D latent space, with cross- and circle-shape points representing the spectral and structural data, respectively. 2D representation of first latent variable (z_1) with respect to second latent variable (z_2) are in Figure 4-3 b) and d). 2D representation of first latent variable with respect to third latent variable (z_3) are in Figure 4-3c). The data points represent one input and are post-colored based on aspect ratio in Figure 4-3 b), total intensity in c) and angle in d). Here, the structural and spectral data of the clusters are displayed as circle- and star-shape points respectively. Each data point is post-colored with red, green, blue and purple, representing the structure information of monomers, dimers, trimers, and larger clusters respectively. In contrast to the dual 2D VAE approach, (Figure C-6), we can see that clusters consisting of different numbers of nanoparticles are well separated in both the structural and spectral latent spaces in the shared latent space, Figure 4-3a.

Figure 4-4 shows a plot of the first variable of the spectral latent space as a function of the first variable of the structural latent spaces. Importantly, this plot of $z_{1,spectra}$ vs. $z_{1,structure}$ shows a linear correlation, meaning that these variables are related. Since $z_{1,structure}$ encodes the aspect ratio of gold nanorods, the correlation of $z_{1,spectra}$ with $z_{1,structure}$ Figure 4-4b implies that the dual-3D- VAE with a shared latent space has likely learned that the spectra depend on the aspect ratio of gold nanorods, as our physical understanding of the system tells us is the case (aspect ratio has impact on the red shift of the plasmonic particles). [125] In Figure 4-4, $z_{2,spectra}$ vs. $z_{2,structure}$ shows linear correlation with angle information and $z_{3,spectra}$ vs. $z_{3,structure}$ shows linear correlation with total intensity of images/spectra images.

We validate our hypothesis by translating the 3D latent space into structural (Figure 4-5) and spectral (Figure 4-6) domains using the dual 3D VAE. We expand the detail of the structural manifold representation along with the third (z_3) latent variable direction. Observations include: (1) In Figure 4-5, there's an increase in aspect ratio of Au nanorods from left to right, which aligns with the expected red shifted in the spectrum of nanoparticles within the clusters depicted in Figure 4-6. (2) In Figure 4-5, the angle of Au nanorods are changing along with z_2 axis (from bottom to top). This correlates with the modification in the spectral shape difference illustrated in Figure 4-6. It's noteworthy that the dual 3D VAE can identify not just the aspect ratio of nanorods particles but also the angle/orientation of nanorods across both spectral and structural datasets.

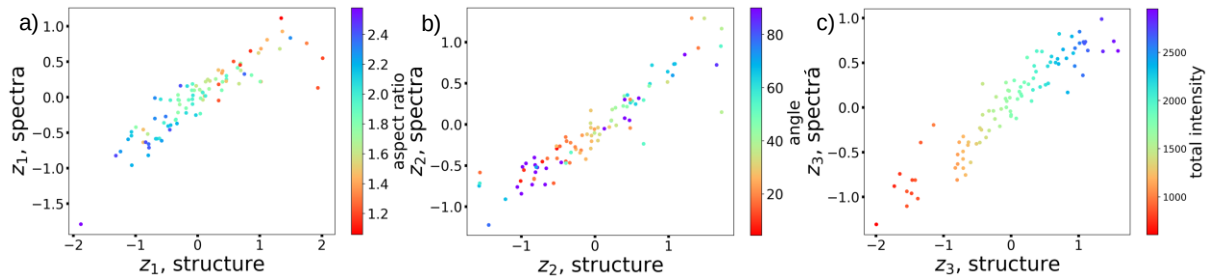

Figure 4-4 Analysis of each latent variables. Relationship between first latent variable (z_1) **(**a), second latent variable (z_2) **(**b), and third latent variable (z_3) **(**c) of structural (shown in x-axis) and spectral (shown in y-axis) datasets. The data points represent one input and are post-colored based on aspect ratio in a), angle in b) and total intensity in c)

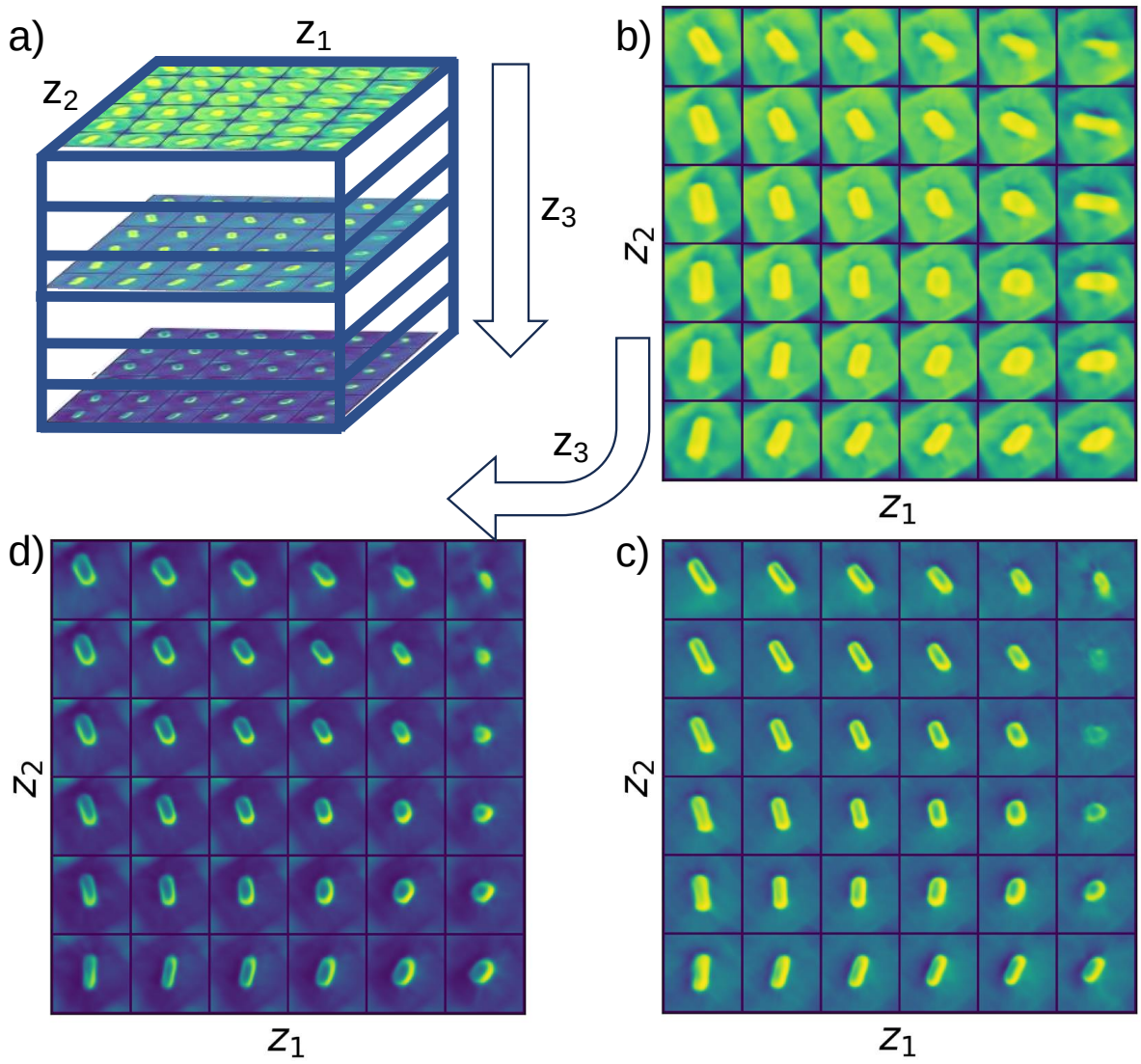

Figure 4-5 Structural manifold representation of the dual 3D VAE model. (a) Combined 3D (z_1, z_2, z_3) structural manifold representation. b), c) and d) show the detail of the structural manifold representation along with the third (z_3) latent variable direction, respectively.

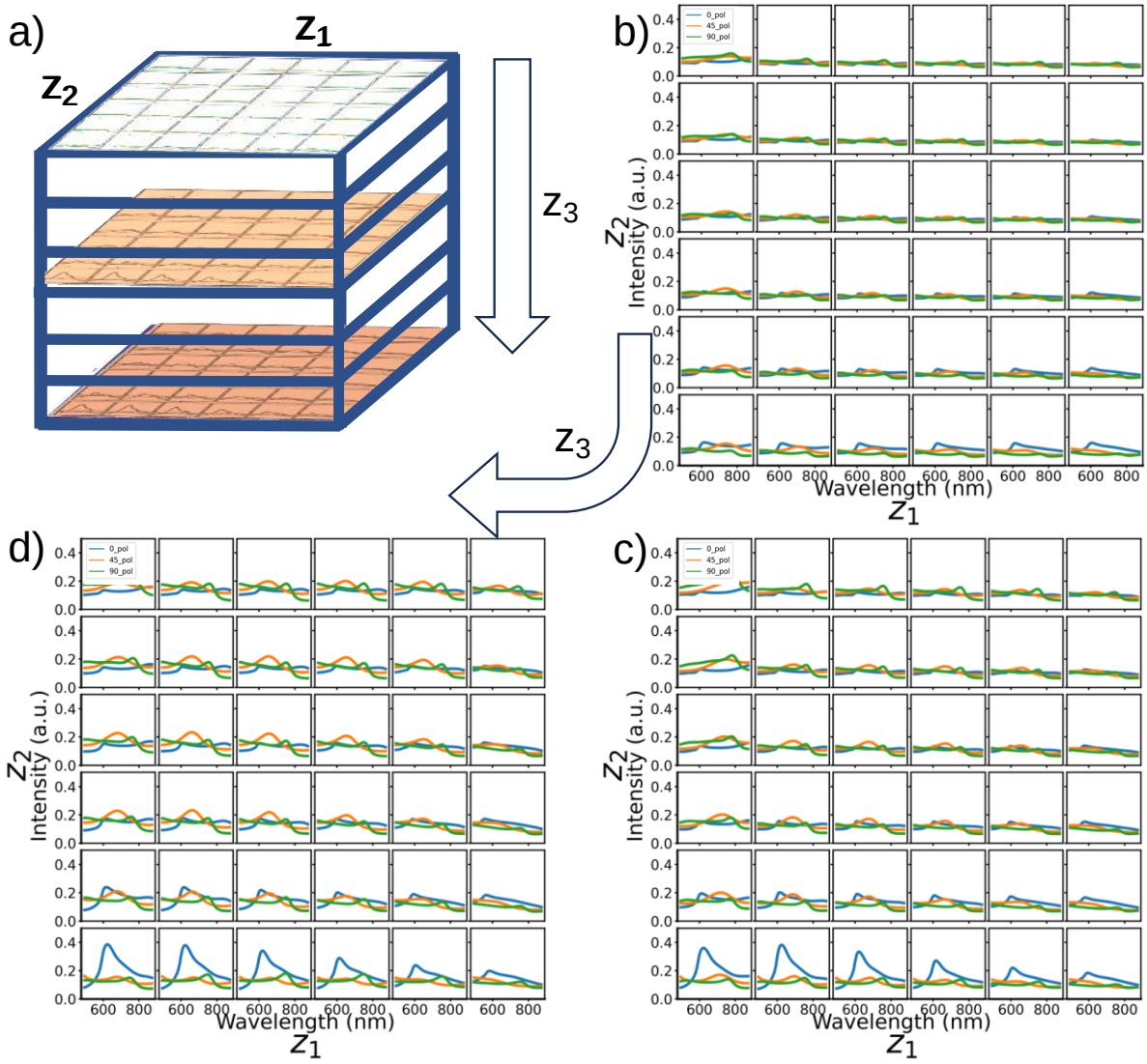

Figure 4-6 Spectral manifold representation of the dual 3D VAE model. (a) Combined 3D (z_1, z_2, z_3) structural manifold representation. b), c) and d) show the detail of the structural manifold representation along with the third (z_3) latent variable direction, respectively. Blue, orange, and green lines show **a** spectrum under 0-, 45- and 90 - degree polarized light.

The Appendix C includes both the confusion matrix and the SSI results on images and shows that this dual-3D- VAE model can predict the structural information with high precision.

4.3 CONCLUSIONS

Over, the dual-3D-VAE presents an innovative machine learning method for drawing correlations between nanorod geometries and their optical characteristics, notably pinpointing sub-diffraction-limit nanorod designs using solely diffraction-limited far-field hyperspectral imagery. A crucial component for effective reconstructions from the shared latent space is the imposition of congruity between the latent representations of both the structural and spectral variational autoencoders during training. This methodology presumes that the encoded relationships between structure and properties can be represented linearly. However, it can be adapted to cater to more intricate relationships.

This strategy of deriving structure-property correlations through mutual encoding is versatile and can cater to a vast array of challenges in materials science and physics, especially in delineating structure-function relationships. This spans across imaging and spectroscopic data, originating from techniques like (scanning) transmission electron microscopy – electron energy loss spectroscopy, a plethora of scanning probe methods and spectroscopies encompassing atomic force microscopy, piezoresponse microscopy, and scanning tunneling microscopy, as well as chemical and optical imaging. Beyond imaging, this methodology proves invaluable for molecular setups, for instance, in jointly encoding structures described via graphs, SELFIES, or SMILES representations and their corresponding properties. For optical systems, future endeavors may facilitate reconstructions of both nanorod orientations and distances between nanorods in clusters. This could be achieved by amplifying the collected data, enriching training datasets, and possibly expanding the latent space.

4.4 MATERIAL AND METHODS

4.4.1 *Materials*

Citrate-coated gold nanoparticle solution was purchased from Nanopartz. Product number of "Sample Au_700" is A12-50-700-CTAB-DIH-1-1. Product number of "Sample Au_750" is A12-70-750-CTAB-DIH-1-1.

4.4.2 *Sample Preparation*

Suspension of gold solution were precipitated out in 25 mM NaCl solution onto indium-doped tin oxide (ITO) glass. The process was watched under darkfield scattering microscopy to ensure the particles are far from each other and there is no spectral overlapping. The ITO glass with the gold particles on top was carefully analyzed under both scanning electron microscope (SEM) and hyperspectral darkfield scattering microscope.

4.4.3 *Scanning Electron Microscopy*

The SEM images were obtained using TFS Apreo-S with Lovac Scanning Electron Microscope operating at 2kV and 13 pA.

4.4.4 *Hyperspectral Darkfield Scattering Microscopy*

Hyperspectral darkfield scattering image were collected on a Photon Etc. IMA system, using a Nikon Ni-U upright microscope and a 40x objective lens (Nikon, NA 0.6). Full-field scattered light was passed through the tunable volume Bragg grating filter with a step size of 2 nm and integration time of 2 seconds/step. The scattered light at each wavelength step was collected by a CCD camera (Thorlabs, 1501M-US-TE), building the hyperspectral data-cube by accumulating

each wavelength slice from 400 nm to 1000nm. Scattering spectra were normalized by dividing by a white light reference spectrum and then background corrected. We used the Photon Etc. PHySpec software for the calculation. The rest of analysis was done on Google Colab Jupyter notebook.

4.4.5 *UV-Vis Spectroscopy*

The UV-Vis data were obtained using an Agilent 8453 UV−vis spectroscopic system. The spectra were collected from 400 nm to 1000nm, step size 2 nm.

4.4.6 *Data Analysis*

The detailed methodologies of dual VAE analysis on structure and spectra data set are established in Jupyter notebooks, available from https://github.com/yamanmy/dual3DVAE .

4.5 ACKNOWLEDGEMENTS

This work was supported (M.Y.Y., D.G., S.V.K.) by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS–The Center for the Science of Synthesis Across Scales–under Award NumberDE-SC0019288, located at University of Washington and performed (M.Z.) at Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS), a U.S. Department of Energy, Office of Science User Facility. M.Y.Y. and D.S.G. acknowledge support from CNMS user facility, project number CNMS2021-B-00847. SEM imaging was conducted at the University of Washington Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure (NNCI) site which is supported in part by the National Science Foundation, the

University of Washington, the Molecular Engineering and Sciences Institute, and the Clean Energy Institute.

Chapter 5. CONCLUSIONS AND FUTURE DIRECTION

In conclusion, this dissertation undertakes a comprehensive exploration of plasmonic assembly principles, spanning from individual particle levels to expansive area synthesis via protein fiber templates. It further seeks to establish the structure-property relationship pivotal for predicting the optical characteristics of plasmonic nanoparticles.

In Chapter 2, the focus is predominantly on comprehending the interplay between protein fibers and gold nanoparticles, and the ensuing crystalline transformations during the electrochemical doping process. We examine the role of external variables, such as ionic strength and particle aspect ratio, on the assembly mechanism. Utilizing advanced image analysis tools, the distinctive behavior of Au nanoparticles post varied assembly conditions is unraveled.

Chapter 3 delves deeper into the structure-property dynamics specific to plasmonic nanoparticle clusters, notably nanospheres. By directly gauging the plasmonic response of gold nanoparticle clusters, we glean rich structural insights through high-resolution imagery. This chapter culminates in the proposition of a non-invasive methodology for the assembly of plasmonic constituents under the watchful eye of hyperspectral microscopy.

Chapter 4 turns its attention to the orientation dynamics of anisotropic plasmonic nanoparticles. Through the lens of polarized hyperspectral scattering methods, we dissect the plasmonic reactions of gold entities across diverse aspect ratios. This study offers a holistic view of the structure-property correlations intrinsic to both nanospheres and nanorods, furthering our ambition of providing in-situ analyses of macromolecular-templated hierarchical assembly of plasmonic elements.

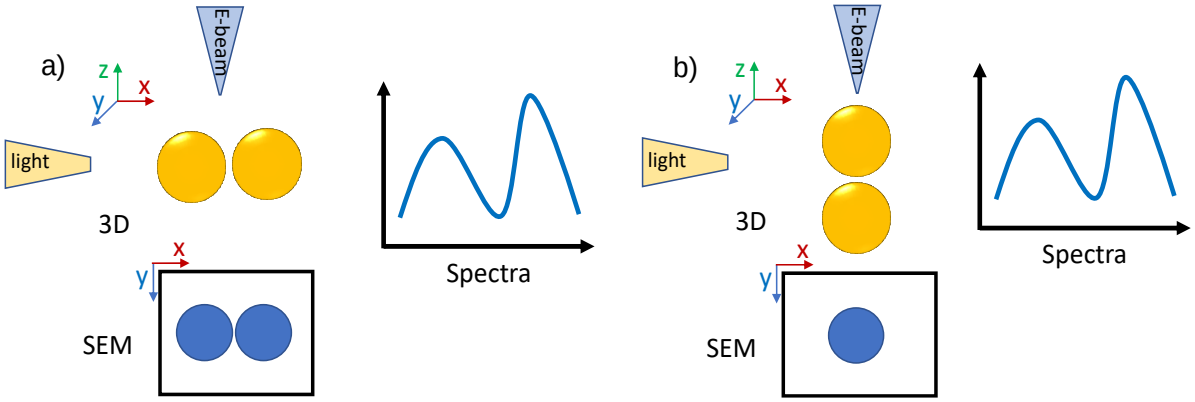

Figure 5-1 Specific case where the dual-VAE model fails. (a) and (b) have same structure and plasmonic response but have different z-axis alignment. Particles are overlapped in b) and could not observed in SEM.

In sum, this dissertation illuminates the sophisticated intricacies of regulating and forecasting the assembly of plasmonic gold nanoparticles. By leveraging cutting-edge machine learning tools and characterization methodologies and meticulously dissecting experimental procedures across scales, it furnishes a substantial contribution to our understanding of plasmonic nanoparticle assembly and the prediction of their optical attributes. For example, Figure 5-1 shows a specific case where our dual-VAE model fails. To achieve better representation of the nanoparticles, we can take advantage of 3D FDTD simulation. Figure 5-2 shows schematic representation of simulation space and the simulation results of 80 x 50 nm Au nanorods under different polarized light. This simulation datasets are generated by python scripts in automated fashion. The python scripts is available at https://github.com/yamanmy/Python_Scripts_for_FDTD_automation . When the dual 3D-VAE can be modified to accept 3D input data, it would be predicted better for the optical properties of plasmonic nanoparticle (see Figure 5-2).

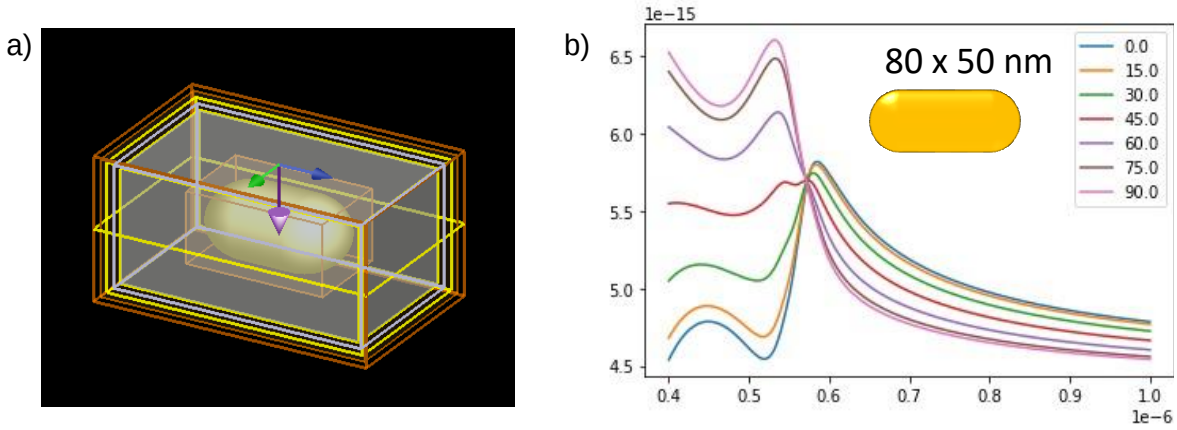

Figure 5-2 3D FDTD simulation. a) Schematic representation of simulation space. b) The simulation results of 80 x 50 nm Au nanorods under different polarized light.

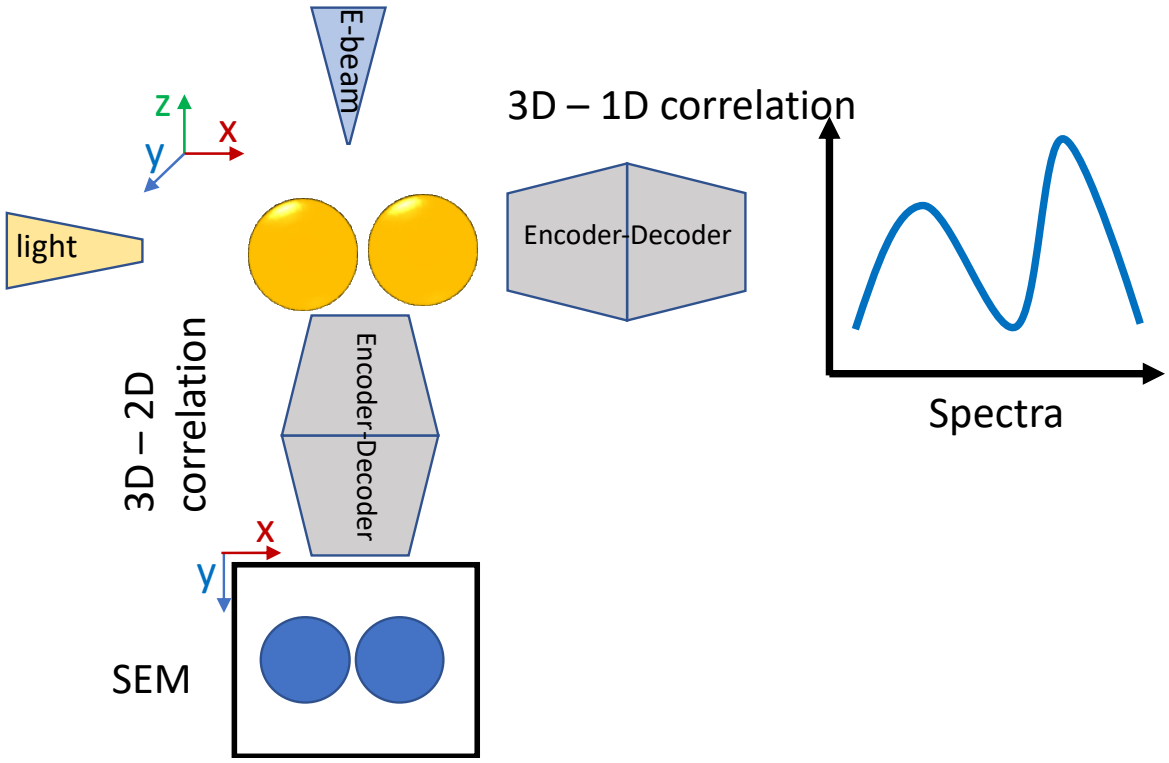

Figure 5-3 Schematic representation of dual 3D VAE model for 3D inputs.

Figure 5-4 shows dimension reduction of hyperspectral data cubes decreasing from 300 to 2 dimensional representation using the VAE model. The hyperspectral data cubes is reduced to 3 dimension from 300 dimension without losing the information. There is sweet area where the acquisition is fast and fruitful spectral collection. This optimization will allows to do fast in-situ characterization of plasmonic particle assembly.

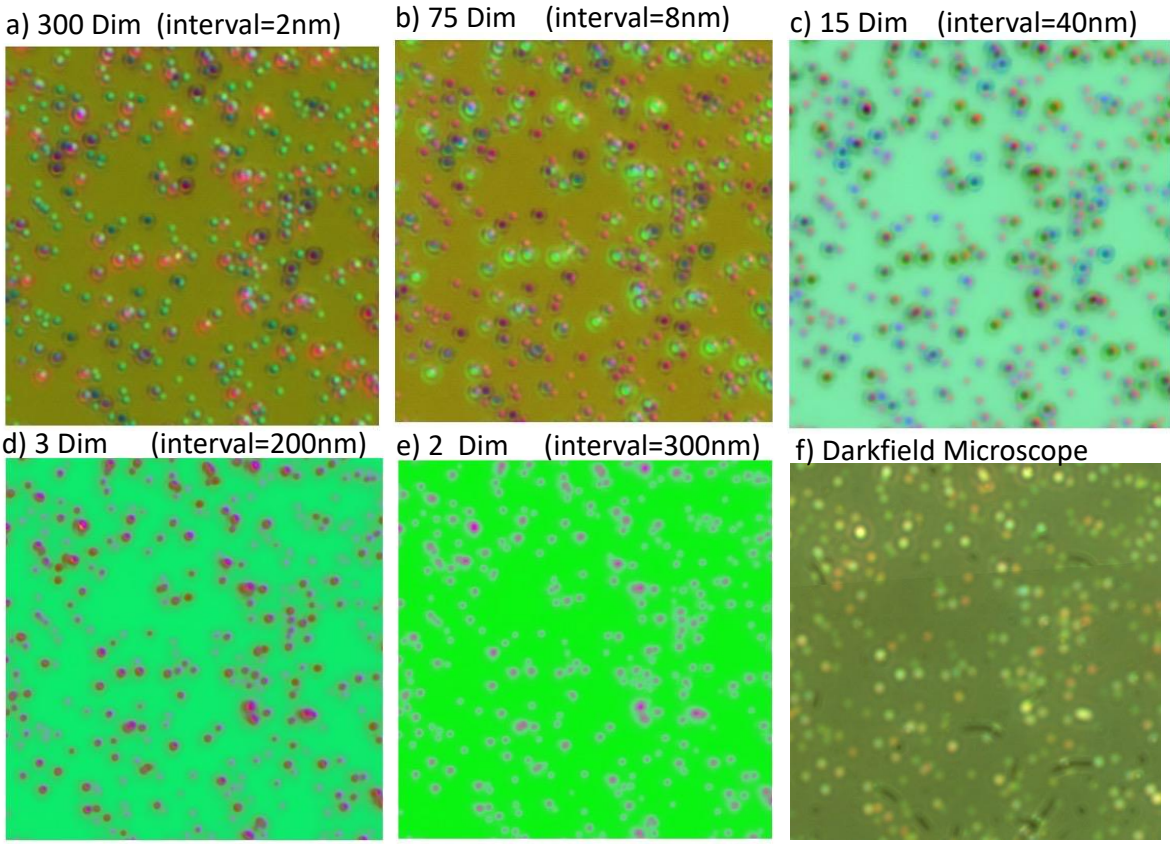

Figure 5-4 Dimension reduction of hyperspectral data cube. The dimension of hyperspectral data cubes decreases from 300 to 2-dimension representation (from a to e). There is sweet area where the acquisition is fast and fruitful spectral collection.

At the culmination of this chapter, we find ourselves poised at an intersection of revelation and enlightenment, fortified by the wisdom and discernment accumulated through these investigations. The path that unfolds before us is laden with opportunities and possibilities. With anticipation, I look forward to observing the further progression and unfolding of this mesmerizing domain.

Appendix A. SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Appendix A accompanies Chapter 2: Alignment of Au Nanorods Along de novo Designed Protein Nanofibers Studied with Automated Image Analysis.

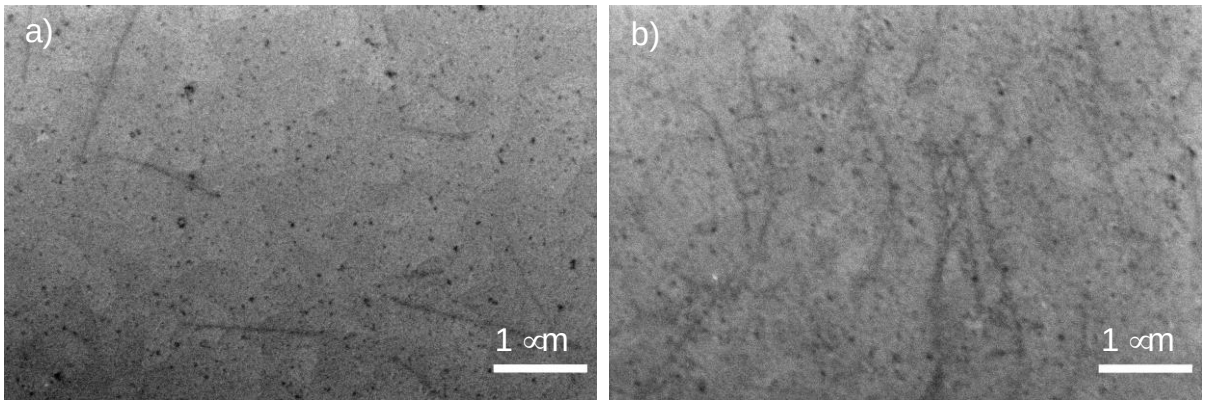

Figure A-1 SEM images of the protein nanofibers attach to substrates. (a) without silane treatment, (b) with silane treatment.

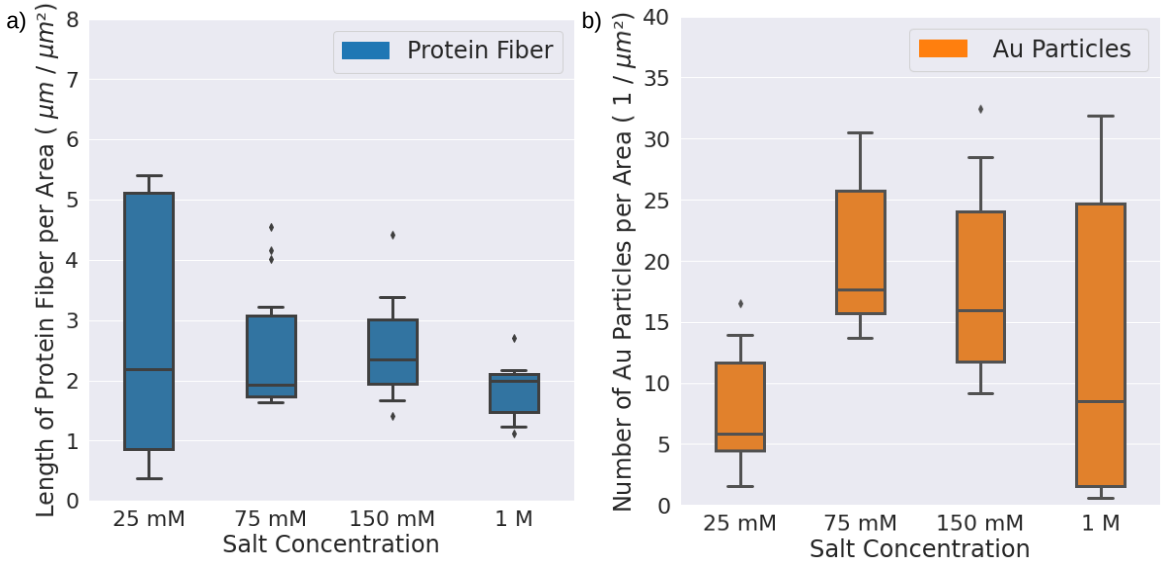

Figure A-2 Quantity results for the density of protein fibers and Au particles on substrate. a) Boxplot of quantity results for the density of protein fibers on substrate at different salt concentration. b) Boxplot of quantity results for the density of Au particles on substrate at different salt concentration.

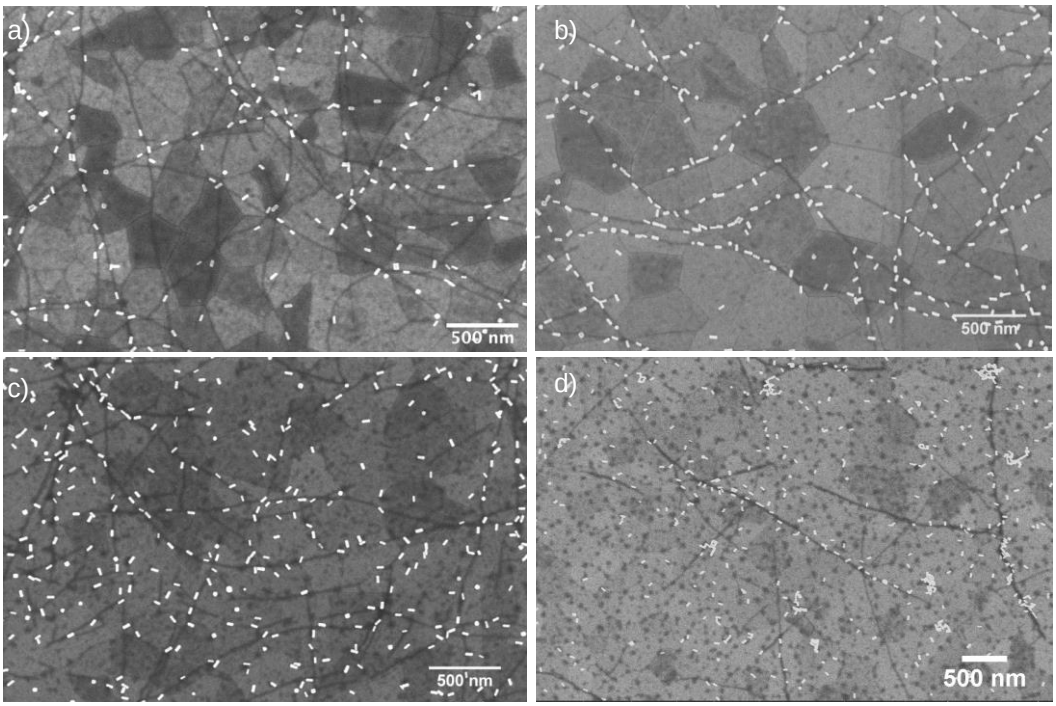

Figure A-3 Unmodified SEM images at different salt concentration. From 25 mM (a), 75mM (b), 150mM (c), to 1M (d).

Table A-1 Quantity analysis of Au nanorods with different aspect ratio.

Sample name	Zeta Potential (mV)	Width of Au Nanorods (nm)	Height of Au Nanorods (nm)	Aspect Ratio of Au Nanorods
AR=2	45 ± 1	18 ± 5	37 ± 8	2.1 ± 0.4
AR=2.5	43 ± 1	19 ± 5	44 ± 9	2.4 ± 0.3
AR=4	43 ± 1	17 ± 7	56 ± 13	3.8 ± 1.2

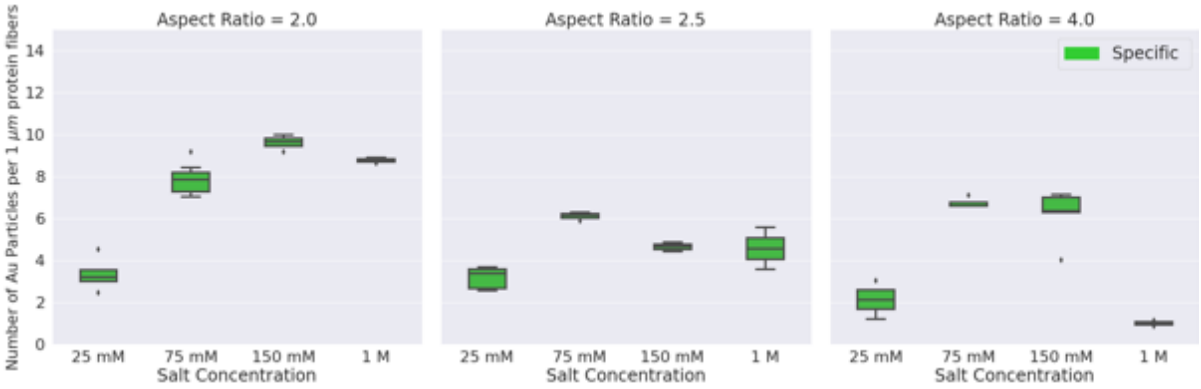

Figure A-4 Boxplot of quantity results for specific Au attachment on protein fibers at different aspect ratio of Au nanorods. The y-axis shows the number of attached Au particles per 1 μm protein fiber. The green bins show specific Au particle attachment to the fibers.

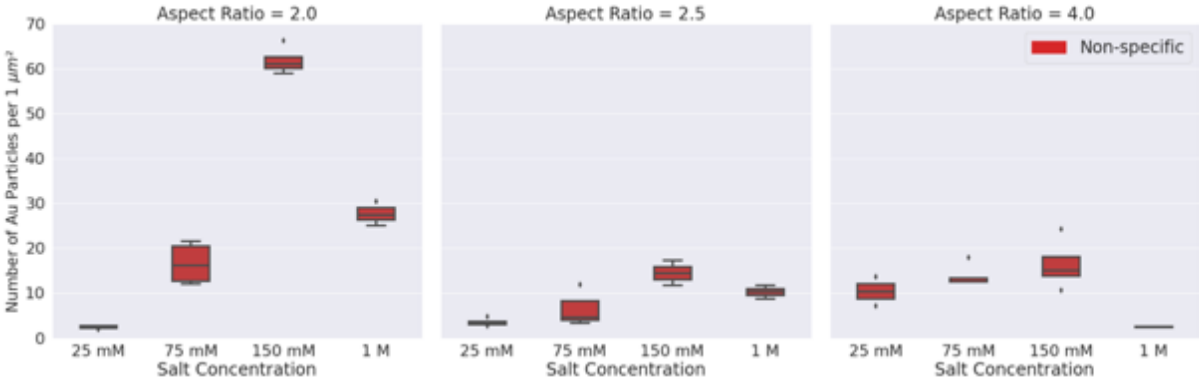

Figure A-5 Boxplot of quantity results for non-specific Au attachment on protein fibers at different aspect ratio of Au nanorods. The y-axis shows the number of attached Au particles per 1 μm protein fiber. The green bins show specific Au particle attachment to the

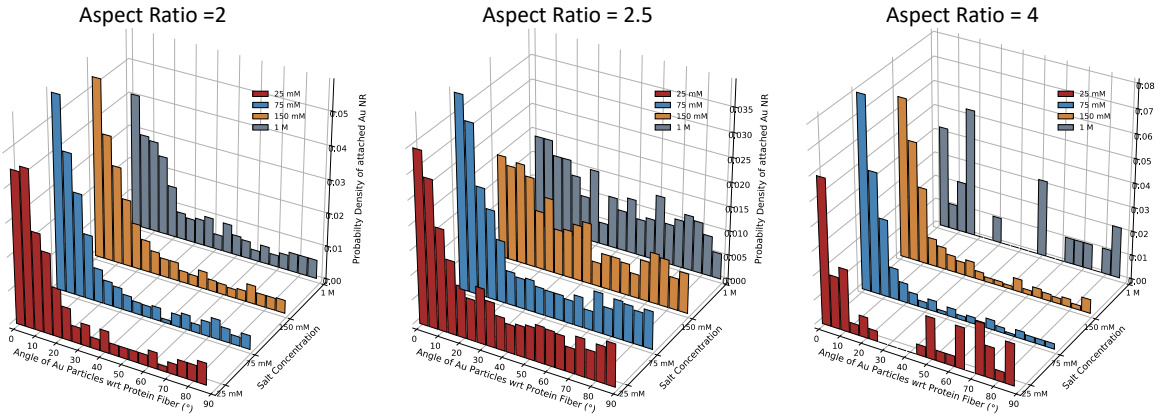

Figure A-6 3D plot of angle distribution of attached Au nanoparticles on the protein fiber at different Au aspect ratios. (Increase from 2 to 4). The x-axis shows the angle of Au nanorods with respect to the protein fiber, the y-axis shows the different salt concentration, and the z-axis shows the probability density of attached Au NRs.

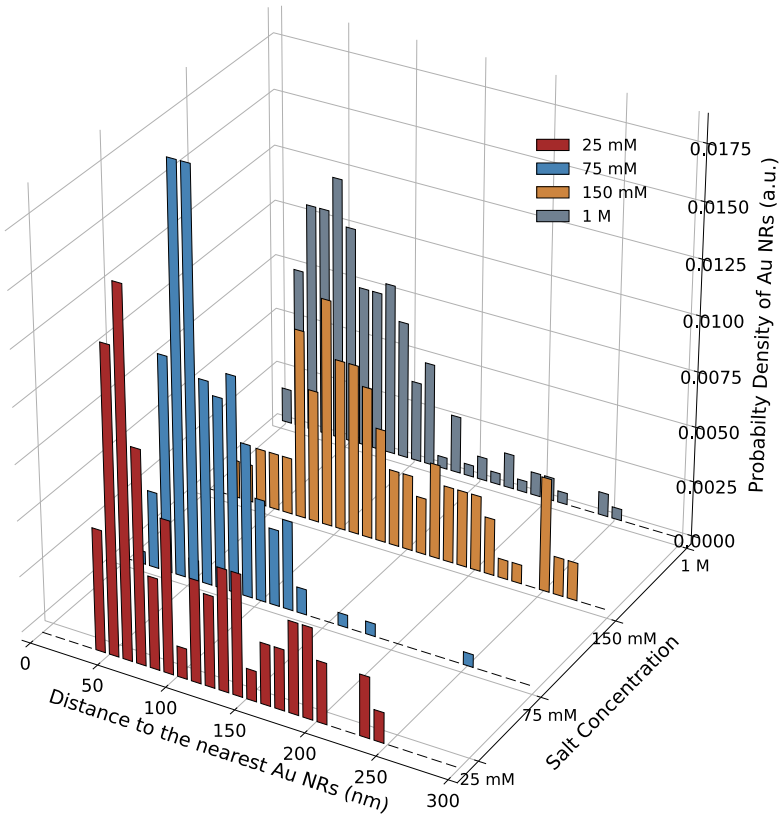

Figure A-7 3D plot of pair distribution of Au nanorods at different salt concentrations. The x-axis shows the distance of Au nanorods to the nearest Au NRs, the y-axis shows the different salt concentration, and the z-axis shows the probability density of the distance of Au

Appendix B. SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Appendix B accompanies Chapter 3: Learning and predicting photonic responses of plasmonic nanoparticle assemblies via dual variational autoencoders.

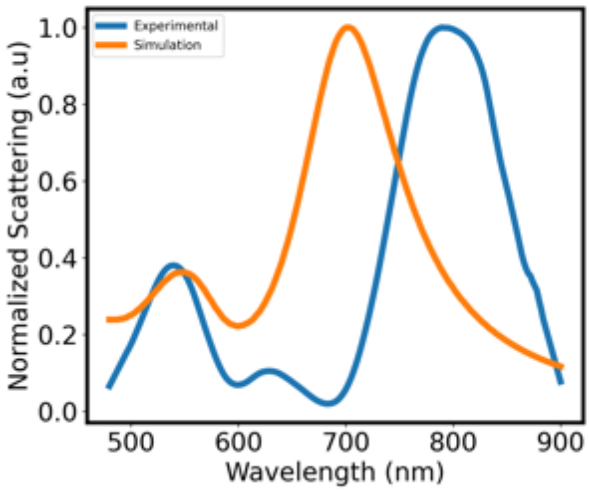

Figure B-1 Finite-Difference Time-Domain Simulation and experimental scattering spectra of a 100 nm dimer particle on ITO.

Finite-difference time-domain (FDTD) numerical simulations were performed using Ansys Lumerical 2021 R1.4 FDTD solver software. The simulation box was constructed of perfectly matched layer boundaries, a mesh size of 2 nm, and an override mesh of 0.1 nm over the cluster structure. 100 nm diameter Au nanospheres were modeled using the optical properties of gold as reported by Johnson and Christy on a thin layer of indium tin oxide (ITO) as modeled by König et al.[139,140] The ITO layer extended through the simulation boundary in each direction. Nanoparticles were modeled with a 0.5 nm dielectric capping layer shell with a refractive index of 1.1 to simulate the thin organic capping layer. A broadband total-field scattered-field source between 400-1000

nm at a step size of 2 nm was chosen as the illumination source. Scattered light was collected via frequency domain power monitors placed on each border on the exterior of the illumination source. Measured nanocluster scattering spectra are typically slightly redshifted from the simulated spectra, likely due to faceting and the unique microstructure at the individual nanoparticle surface [126].

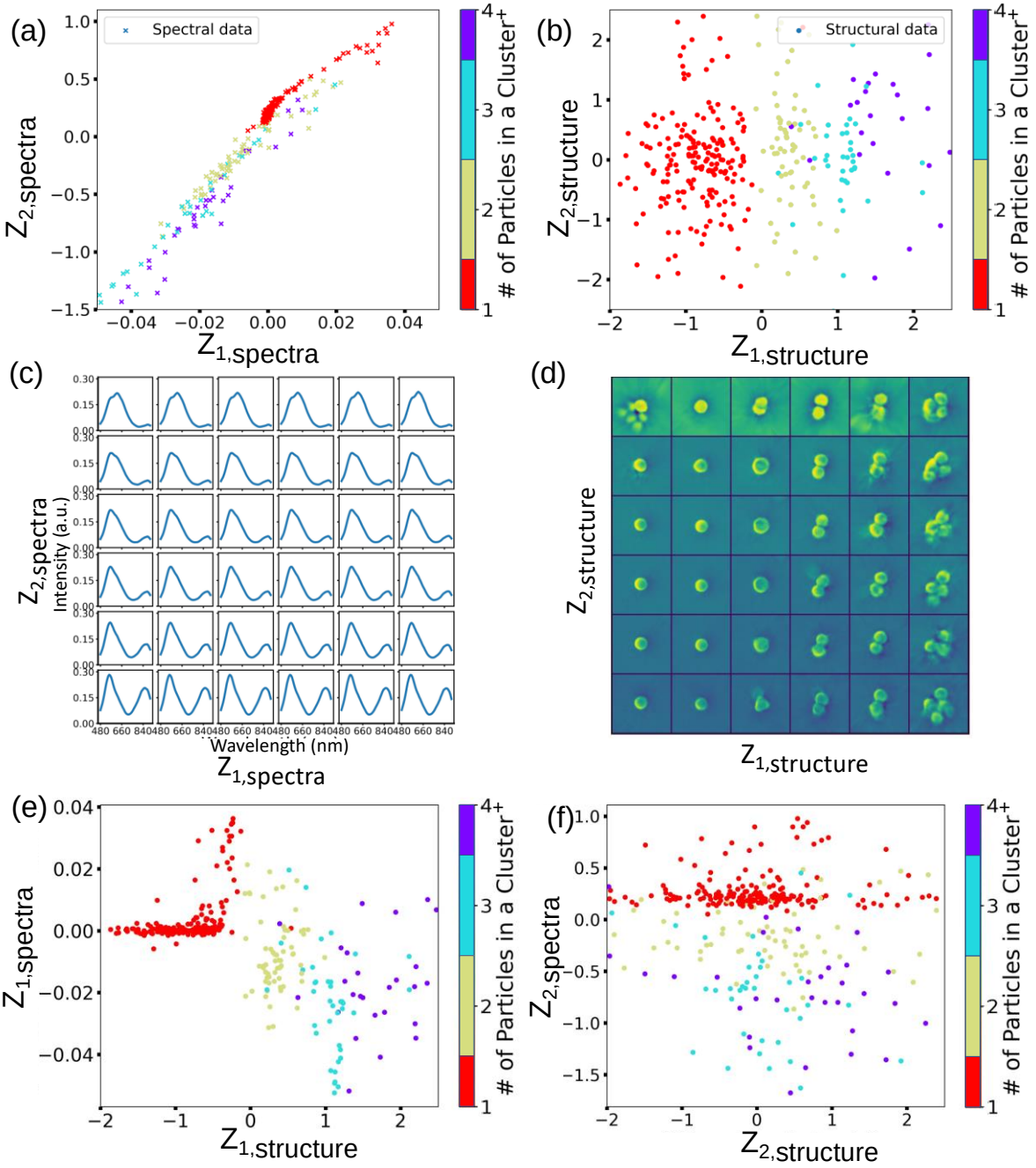

Figure B-2 Analysis of spectra and structure datasets via vanilla variational autoencoder. Individual latent space of (a) spectral and (b) structural datasets. The manifold representation of (c) spectral latent space and (d) structural latent space. e) First and f) second latent variable correlations between structural and spectral datasets.

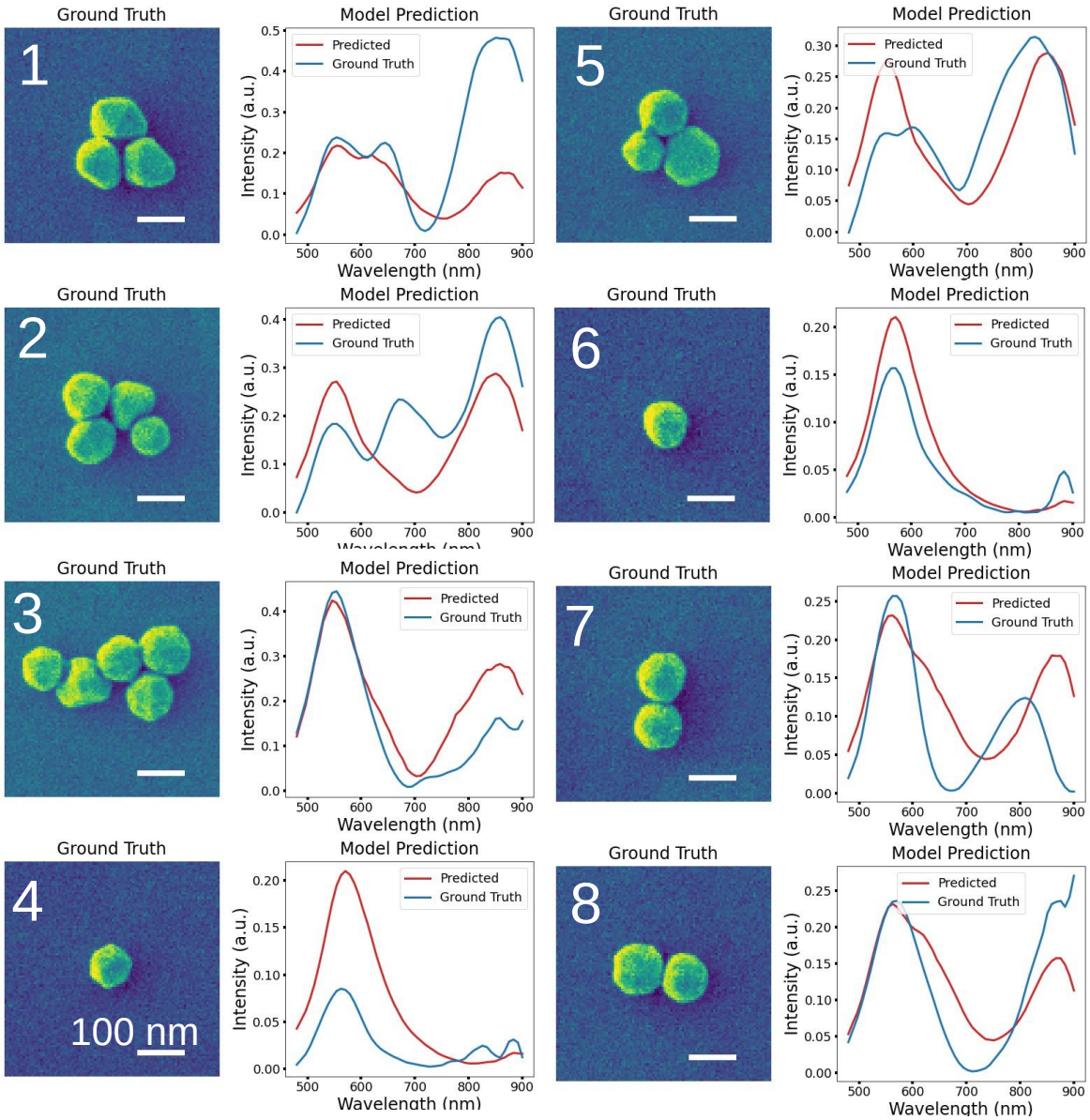

Figure B-3 Examples of the dual im2spec model performance on the SEM images in test dataset.. SEM images are used as input. The red lines are the model prediction whereas blue lines are the ground truth.

We also use structural similarity index (SSI) to evaluate the quality of model reconstruction. SSI is commonly used image quality metrics and depend on image luminance, contrast and structure. When all parameters are taken into consideration, our dual-VAE model has an SSI of 0.22. The SSI reaches up to 0.86 (86% similarity) when we make image binary (to remove the

effects of luminance and contrast) to see only the effect of the structure. Both the confusion matrix and the SSI results showed that our dual-VAE model can predict the structural information with high precision.

We then calculated the mean absolute error (MAE) for spectral prediction in Figure S4. The solid black line shows the mean of MAE across spectral data, and the other thinner lines show the MAE of each cluster based on particle numbers (monomer, dimer, etc.). The model prediction on traverse peak region is better than on longitudinal peak region. Where the transverse resonance peak is more related to cluster size, the longitudinal resonance peak is more related to cluster shape and geometry. When one compares to the predicted spectra with ground truth, it is obvious that the performance of the dual-VAE is high.

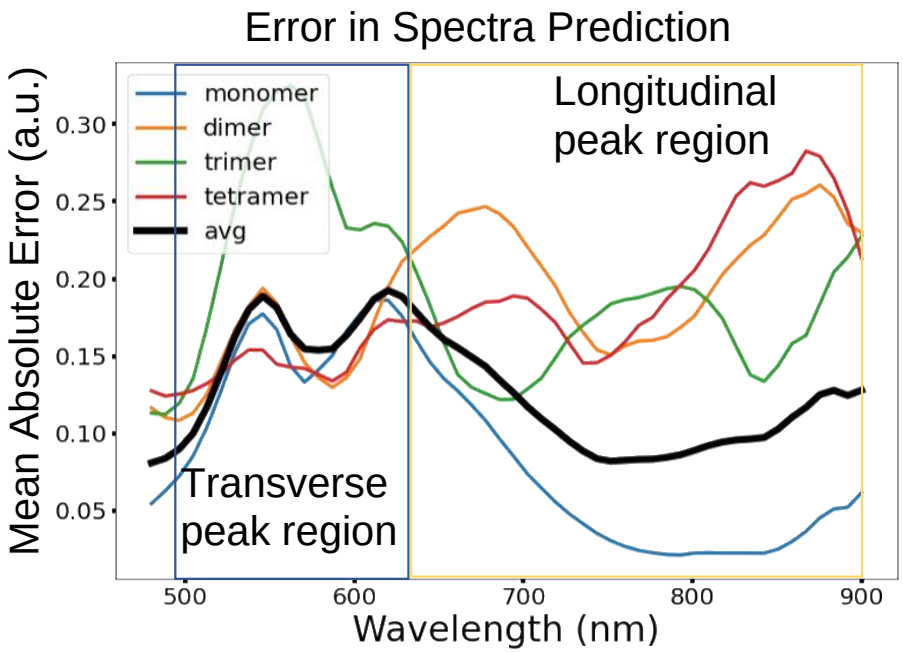

Figure B-4 Error calculation on the dual-VAE prediction.

. The solid black line shows the mean of MAE across spectral data, and the other thinner lines show the MAE of each cluster based on particle numbers (monomer, dimer, trimer and

Appendix C. SUPPLEMENTARY INFORMATION FOR CHAPTER 4

Appendix C accompanies Chapter 4: Predicting Photonic Responses And Orientation Of Plasmonic Nanorods Via 3d - Dual Variational Autoencoders

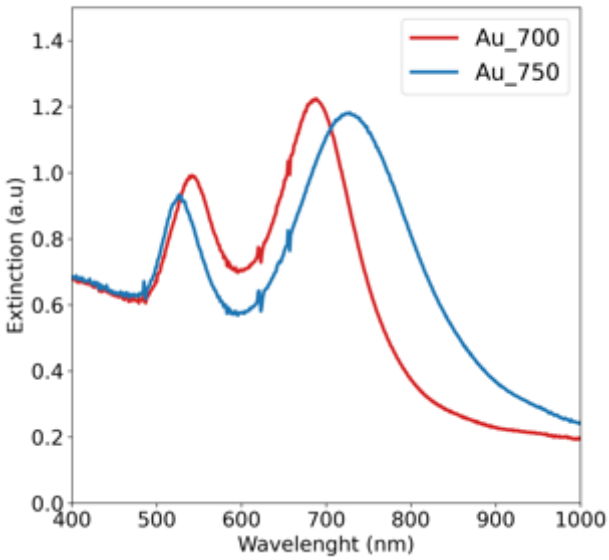

Figure C-1 UV-Vis Spectra of two different Au Nanorod stock solution. The red line shows the spectrum of Au nanorod with 50nm x110 nm dimensions solution. The blue line shows the spectrum of Au nanorod with 70nm x120 nm dimensions solution.

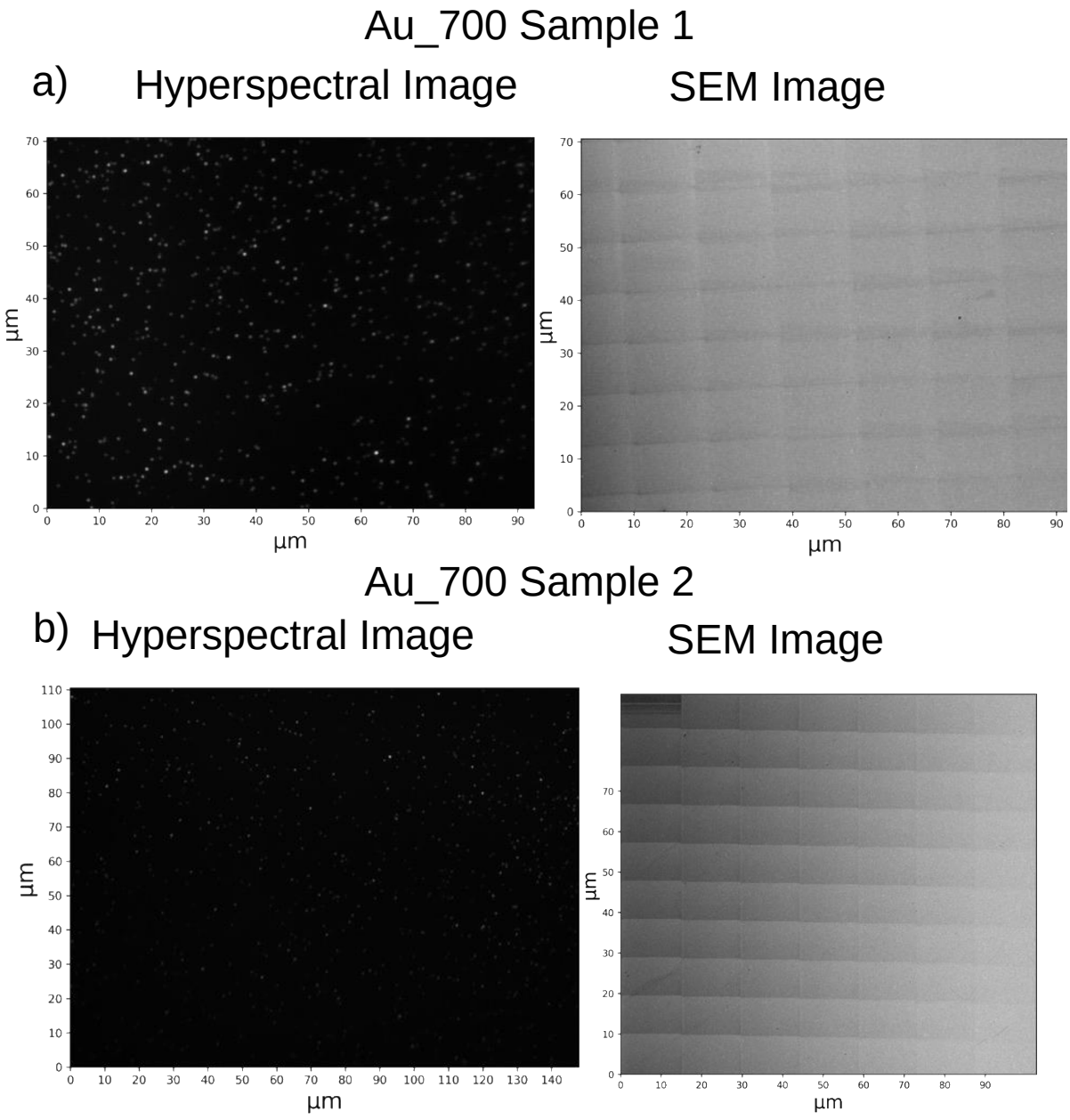

Figure C-2 Widefield view of Au_700 samples with correlated hyperspectral and SEM images.

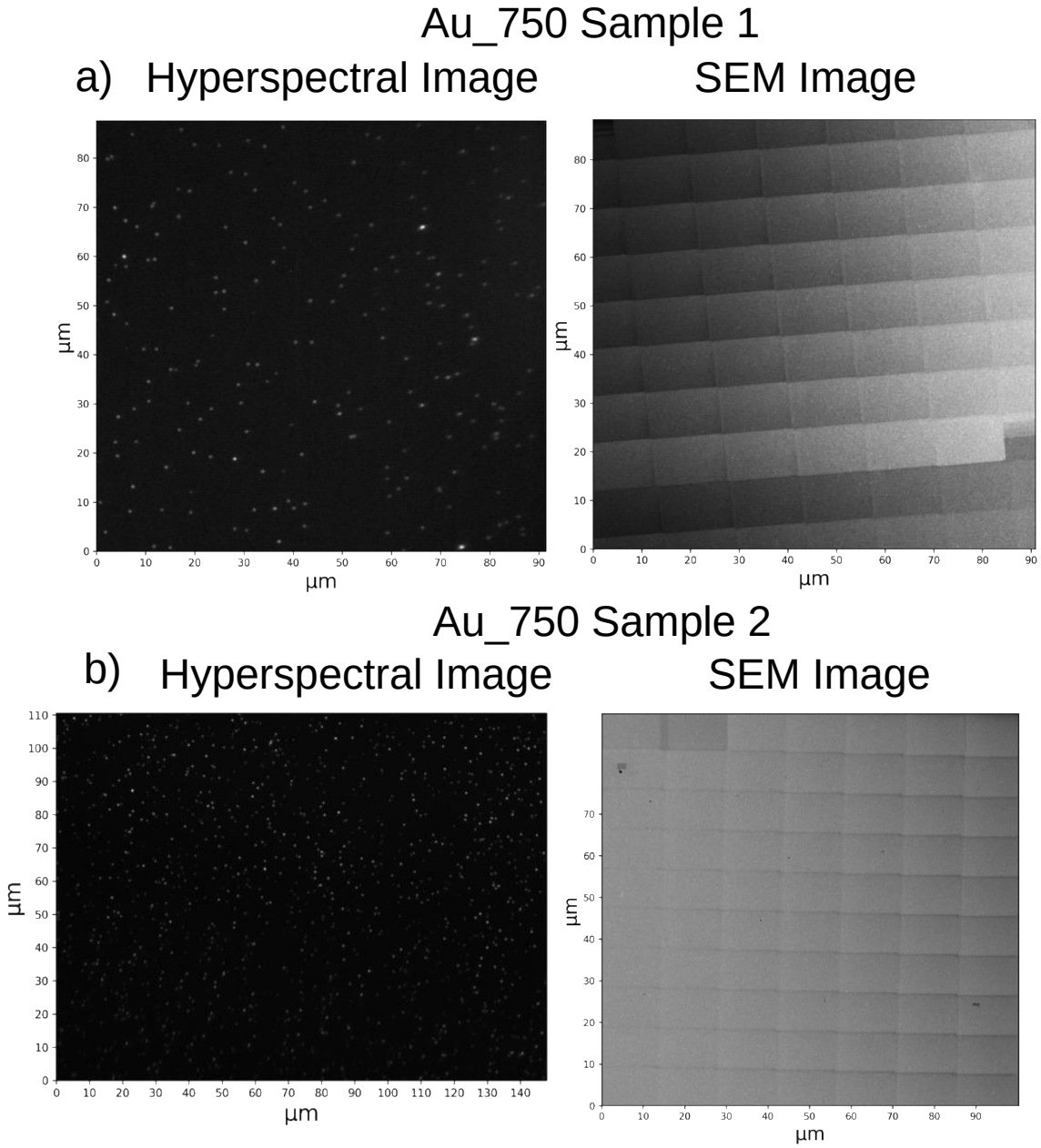

Figure C-3 Widefield view of Au_700 samples with correlated hyperspectral and SEM images.

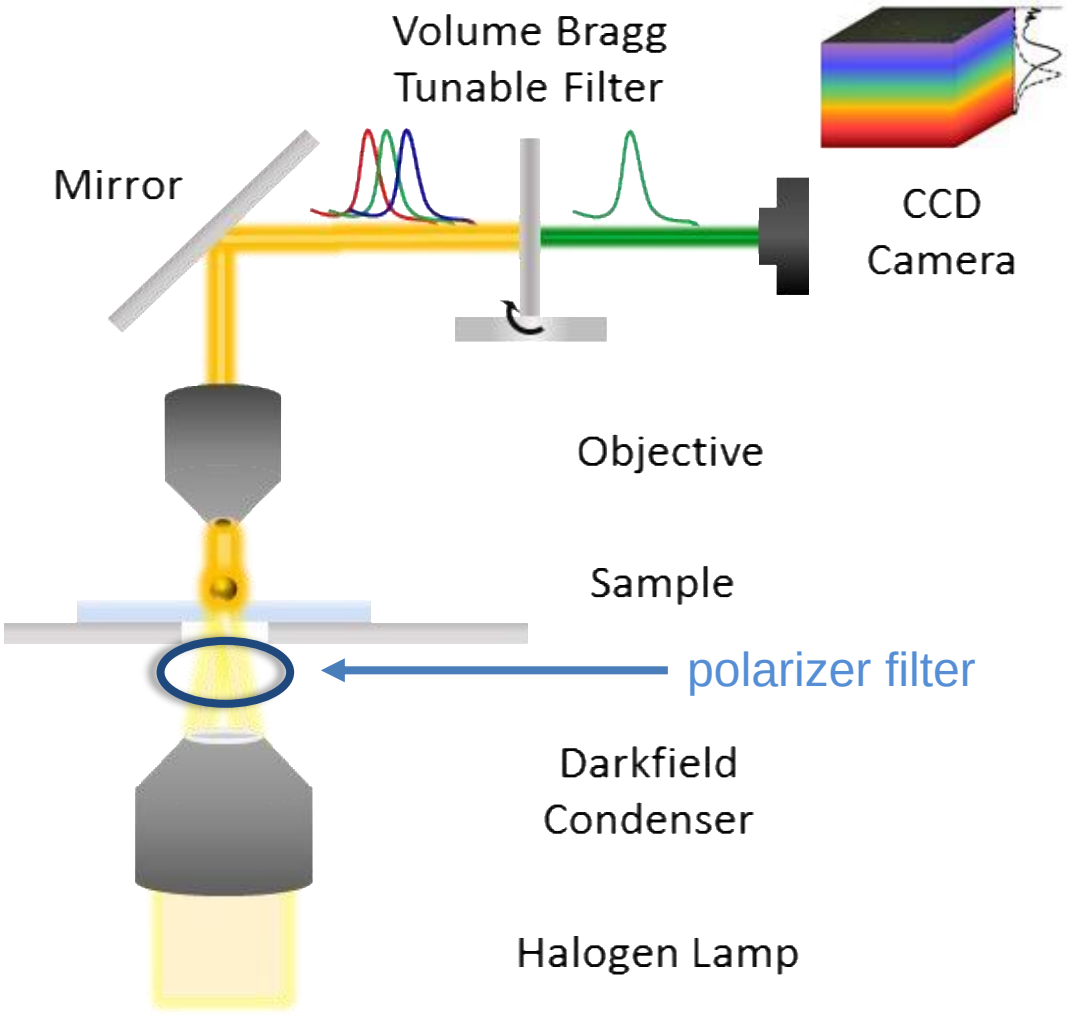

Figure C-4 Experimental setup for polarized hyperspectral scattering microscope. The polarized filter was inserted between sample and darkfield condenser. The angle was controlled by manual 360° rotation stage.

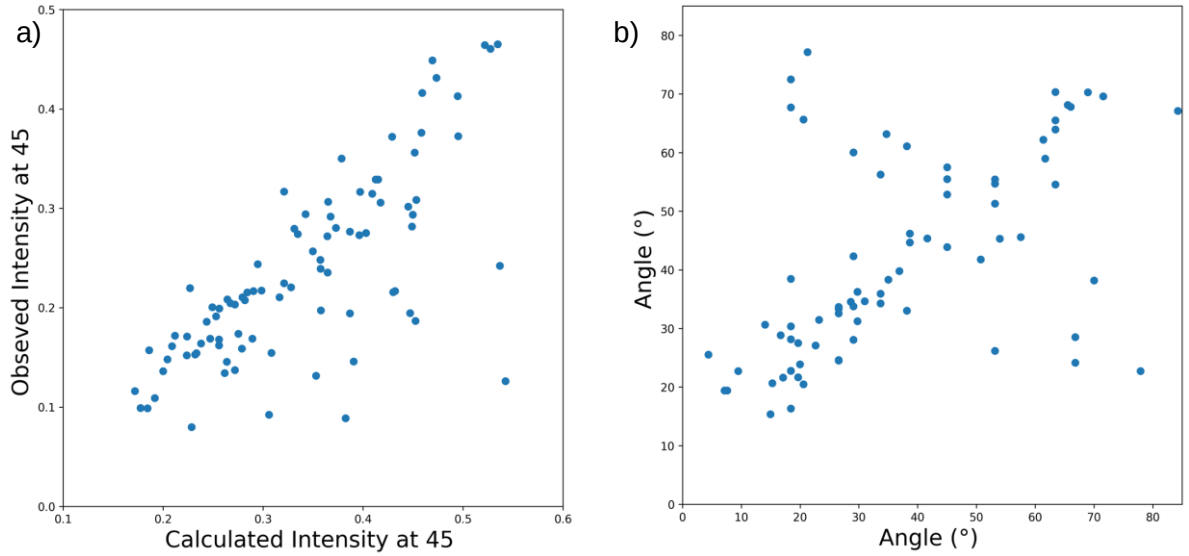

Figure C-5 Analysis of our datasets. a) Intensity assessment at 45° polarized light between calculated intensity using 0° and 90° polarized light and observed intensity. They show linear correlation. b) Angle assessment between structural datasets (x-axis) and spectral datasets (y-axis).

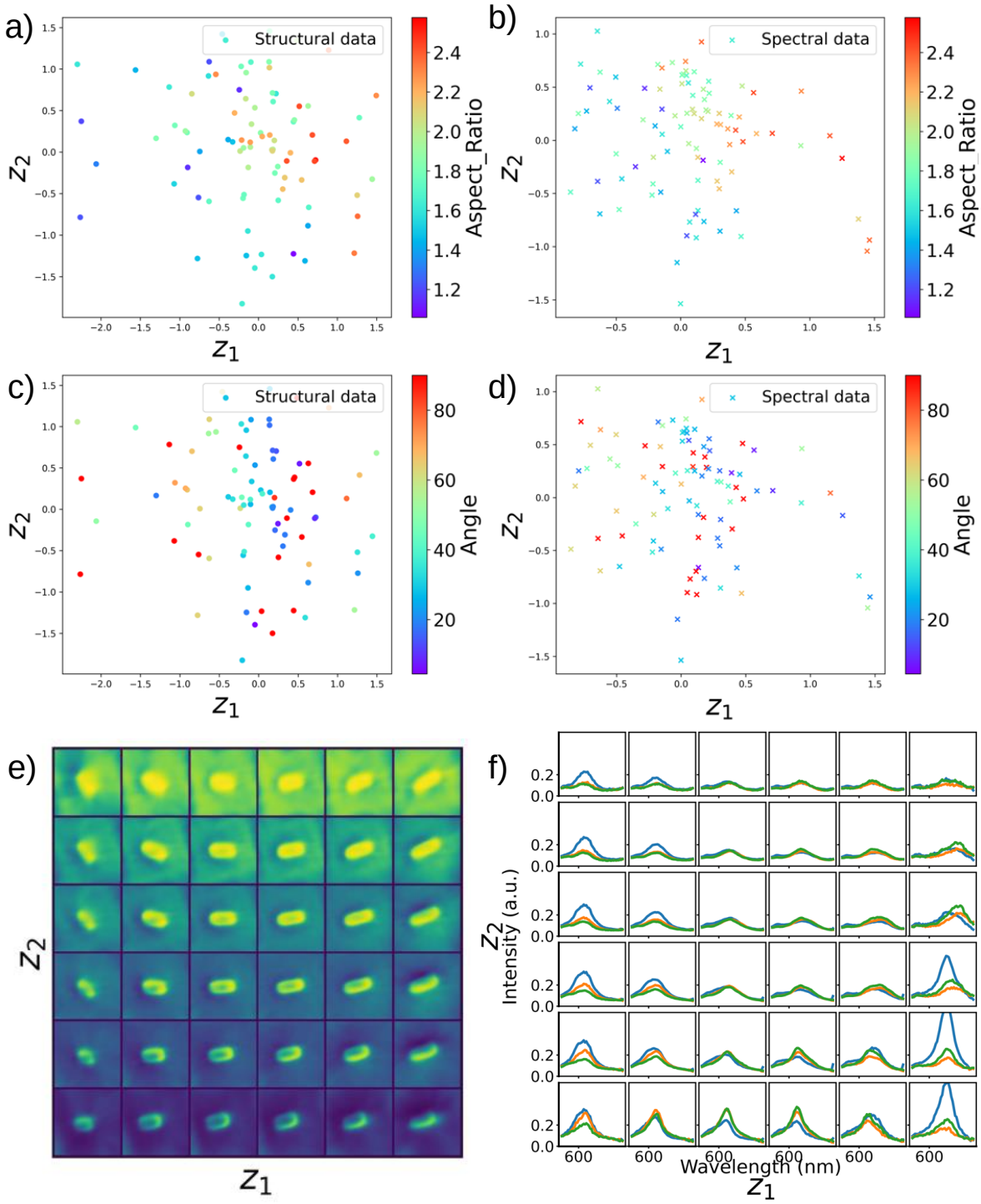

Figure C-6 Performance of the dual-VAE model on Au nanorod datasets. Individual latent space of structural (a, c) and spectral (b, d) datasets. Star- and circle-shape points represent the spectral and structural data, respectively. The datapoints are post-colored based on aspect ratio in a) and b), angle in c) and d). The learned manifold representation of (e) spectral latent space and (f) structural latent space.

VITA

EDUCATION

2023 Ph.D. in Chemistry: Data Science – University of Washington,USA
Advisor: Prof. David S. Ginger

2017 M.S. in Chemistry – Bilkent University, Turkey
Advisor: Prof. Ömer Dağ

2015 B.S. in Chemistry – Bilkent University, Turkey
Advisor: Prof. Ömer Dağ

PUBLICATIONS

1. M. Y. Yaman, S. V. Kalinin, K. N. Guye, D. S. Ginger, M. Ziatdinov. Learning and predicting photonic responses of plasmonic nanoparticle assemblies via dual variational autoencoders. Small, 2023, 19 (25), 2205893.
2. Z. Li, S. Wang, U. Nattermann, A. K. Bera, A. J. Borst, M. Y. Yaman, M. J. Bick, … , D. S. Ginger, D. Baker. Accurate Computational Design of 3D Protein Crystals. Nature Materials, 2023, in press.
3. M. Y. Yaman, S. V. Kalinin, M. Ziatdinov, D. S. Ginger. Predicting plasmonic response and orientation of gold nanorods via 3D dual variational autoencoders. In preparation.
4. T. R. Rana, M. Abbas, E. Schwartz, F. Jiang, M. Y. Yaman, D. S. Ginger, D. MacKenzie. Scalable Passivation Strategies to Improve Efficiency of Slot Die-Coated Perovskite Solar Cells. Submitting.
5. F. Jiang, Y. Shi, … , M. Y. Yaman, ... , D. S. Ginger. Architecture Optimization Dramatically Improves Reverse Bias Stability in Perovskite Solar Cells (Under revision)
6. M. Ziatdinov, M. Y. Yaman, Y. Liu, D. S. Ginger, S. V. Kalinin. Semi-supervised learning of images with strong rotational disorder: assembling nanoparticle libraries. 2023, (Under revision in Patterns), preprint arXiv:2105.11475.
7. M. Y. Yaman, K. N. Guye, M. Ziatdinov, H. Shen, D. Baker, S. V. Kalinin, D.S. Ginger. Alignment of Au nanorods along de novo designed protein nanofibers studied with automated image analysis. Soft Matter, 2021,17, 6109 – 6115.
8. K.N Guye, H. Shen, M. Y. Yaman, G. Y. Liao, D. Baker, D. S. Ginger. Importance of Substrate-Particle Repulsion for Protein-Templated Assembly of metal Nanoparticles. Langmuir, 2021, 37(30), 9111 – 9119.

BIBLIOGRAPHY

1. Anker, J. N. *et al.* Biosensing with plasmonic nanosensors. *Nat Mater* **7**, 442–453 (2008).
2. Azzam, S. I. *et al.* Ten years of spasers and plasmonic nanolasers. *Light Sci Appl* **9**, 90 (2020).
3. Yu, H., Peng, Y., Yang, Y. & Li, Z.-Y. Plasmon-enhanced light–matter interactions and applications. *NPJ Comput Mater* **5**, 45 (2019).
4. Zheng, Y. B., Kiraly, B., Weiss, P. S. & Huang, T. J. Molecular plasmonics for biology and nanomedicine. *Nanomedicine* **7**, 751–770 (2012).
5. Boriskina, S. V, Ghasemi, H. & Chen, G. Plasmonic materials for energy: From physics to applications. *Materials Today* **16**, 375–386 (2013).
6. Wriedt, T. Mie Theory: A Review. in *The Mie Theory: Basics and Applications* (eds. Hergert, W. & Wriedt, T.) 53–71 (Springer Berlin Heidelberg, 2012). doi:10.1007/978-3-642-28738-1_2.
7. Koya, A. N. *et al.* Nanoporous Metals: From Plasmonic Properties to Applications in Enhanced Spectroscopy and Photocatalysis. *ACS Nano* **15**, 6038–6060 (2021).
8. Lee, S. *et al.* Controlled Assembly of Plasmonic Nanoparticles: From Static to Dynamic Nanostructures. *Advanced Materials* **33**, 2007668 (2021).
9. Park, S. Y. *et al.* DNA-programmable nanoparticle crystallization. *Nature* **451**, 553–556 (2008).
10. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
11. Merg, A. D. *et al.* Peptide-Directed Assembly of Single-Helical Gold Nanoparticle Superstructures Exhibiting Intense Chiroptical Activity. *J Am Chem Soc* **138**, 13655–13663 (2016).
12. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
13. Lee, H.-E. *et al.* Amino-acid- and peptide-directed synthesis of chiral plasmonic gold nanoparticles. *Nature* **556**, 360–365 (2018).
14. Li, Z. *et al.* Accurate computational design of three-dimensional protein crystals. *Nat Mater* (2023) doi:10.1038/s41563-023-01683-1.
15. Klocke, F., Raedt, H.-W. & Hoppe, S. 2D-FEM SIMULATION OF THE ORTHOGONAL HIGH SPEED CUTTING PROCESS. *Machining Science and Technology* **5**, 323–340 (2001).
16. Huiping, L., Guoqun, Z., Shanting, N. & Chuanzhen, H. FEM simulation of quenching process and experimental verification of simulation results. *Materials Science and Engineering: A* **452–453**, 705–714 (2007).
17. Alsawafta, M., Wahbeh, M. & Truong, V.-V. Plasmonic Modes and Optical Properties of Gold and Silver Ellipsoidal Nanoparticles by the Discrete Dipole Approximation. *J Nanomater* **2012**, 457968 (2012).

18. Yurkin, M. A. & Huntemann, M. Rigorous and Fast Discrete Dipole Approximation for Particles near a Plane Interface. *The Journal of Physical Chemistry C* **119**, 29088–29094 (2015).
19. Lin, R., Zhai, Y., Xiong, C. & Li, X. Inverse design of plasmonic metasurfaces by convolutional neural network. *Opt Lett* **45**, 1362–1365 (2020).
20. Sajedian, I., Kim, J. & Rho, J. Finding the optical properties of plasmonic structures by image processing using a combination of convolutional neural networks and recurrent neural networks. *Microsyst Nanoeng* **5**, 27 (2019).
21. Dou, J. *et al.* De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
22. Fratzl, P. & Weinkamer, R. Nature's hierarchical materials. *Prog Mater Sci* **52**, 1263–1334 (2007).
23. Nyström, G., Roder, L., Fernández-Ronco, M. P. & Mezzenga, R. Amyloid Templated Organic–Inorganic Hybrid Aerogels. *Adv Funct Mater* **28**, 1703609 (2018).
24. Yan, F. *et al.* Controlled synthesis of highly-branched plasmonic gold nanoparticles through peptoid engineering. *Nat Commun* **9**, 2327 (2018).
25. Ma, J. *et al.* Nanoparticle-Mediated Assembly of Peptoid Nanosheets Functionalized with Solid-Binding Proteins: Designing Heterostructures for Hierarchy. *Nano Lett* **21**, 1636–1642 (2021).
26. Monahan, M. *et al.* Peptoid-directed assembly of CdSe nanoparticles. *Nanoscale* **13**, 1273–1282 (2021).
27. Chen, C.-L., Zhang, P. & L. Rosi, N. A New Peptide-Based Method for the Design and Synthesis of Nanoparticle Superstructures: Construction of Highly Ordered Gold Nanoparticle Double Helices. *J Am Chem Soc* **130**, 13555–13557 (2008).
28. Yu, Y. *et al.* Biomimetic Mineralized Organic–Inorganic Hybrid Macrofiber with Spider Silk-Like Supertoughness. *Adv Funct Mater* **30**, 1908556 (2020).
29. Kijima, M., Oaki, Y., Munekawa, Y. & Imai, H. Synthesis and Morphogenesis of Organic and Inorganic Polymers by Means of Biominerals and Biomimetic Materials. *Chemistry – A European Journal* **19**, 2284–2293 (2013).
30. Bridonneau, N., Noël, V., Zrig, S. & Carn, F. Self-Assembly of Gold Nanoparticles with Oppositely Charged, Long, Linear Chains of Periodic Copolymers. *J Phys Chem B* **124**, 900–908 (2020).
31. Myroshnychenko, V. *et al.* Modelling the Optical Response of Gold Nanoparticles. *Chem. Soc. Rev.* **37**, 1792 (2008).
32. Bai, P. *et al.* Diversifying Nanoparticle Assemblies in Supramolecule Nanocomposites Via Cylindrical Confinement. *Nano Lett* **17**, 6847–6854 (2017).
33. Liu, Z., Huang, H. & He, T. Large-Area 2D Gold Nanorod Arrays Assembled on Block Copolymer Templates. *Small* **9**, 505–510 (2013).
34. Yi, C. *et al.* Self-limiting directional nanoparticle bonding governed by reaction stoichiometry. *Science (1979)* **369**, 1369 LP – 1374 (2020).
35. Liu, K. *et al.* Step-Growth Polymerization of Inorganic Nanoparticles. *Science (1979)* **329**, 197 LP – 200 (2010).
36. Lee, C. H., Tian, L., Abbas, A., Kattumenu, R. & Singamaneni, S. Directed assembly of gold nanorods using aligned electrospun polymer nanofibers for highly efficient SAB - Nonspherical metal nanoparticles are very attractive plasmonic nanostructures owing to

the facile tunability of the plasmonic properties and the pr. *Nanotechnology* **22**, 275311 (2011).
37. D. Merg, A. *et al.* Peptide-Directed Assembly of Single-Helical Gold Nanoparticle Superstructures Exhibiting Intense Chiroptical Activity. *J Am Chem Soc* **138**, 13655–13663 (2016).
38. Neubrech, F., Hentschel, M. & Liu, N. Reconfigurable Plasmonic Chirality: Fundamentals and Applications. *Advanced Materials* **32**, 1905640 (2020).
39. Qian, Z. & S. Ginger, D. Reversibly Reconfigurable Colloidal Plasmonic Nanomaterials. *J Am Chem Soc* **139**, 5266–5276 (2017).
40. Amendola, V., Pilot, R., Frasconi, M., Maragò, O. M. & Iatì, M. A. Surface plasmon resonance in gold nanoparticles: a review. *Journal of Physics: Condensed Matter* **29**, 203002 (2017).
41. Hao, P., Wu, Y. & Li, F. Improved sensitivity of wavelength-modulated surface plasmon resonance biosensor using gold nanorods. *Appl Opt* **50**, 5555 (2011).
42. Kabashin, A. V *et al.* Plasmonic Nanorod Metamaterials for Biosensing. *Nat. Mater.* **8**, 867 (2009).
43. Xue, X., Wang, F. & Liu, X. Emerging Functional Nanomaterials for Therapeutics. *J. Mater. Chem.* **21**, 13107 (2011).
44. Jiang, W. *et al.* Emergence of complexity in hierarchically organized chiral particles. *Science (1979)* eaaz7949 (2020) doi:10.1126/science.aaz7949.
45. Lu, J. *et al.* Enhanced optical asymmetry in supramolecular chiroplasmonic assemblies with long-range order. *Science (1979)* **371**, 1368 LP – 1374 (2021).
46. Pyles, H., Zhang, S., De Yoreo, J. J. & Baker, D. Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **571**, 251–256 (2019).
47. Shen, H. *et al.* De novo design of self-assembling helical protein filaments. *Science (1979)* **362**, 705 (2018).
48. Fang, N., Lee, H., Sun, C. & Zhang, X. Sub-Diffraction-Limited Optical Imaging with a Silver Superlens. *Science (1979)* **308**, 534 (2005).
49. Bhat, R. R. & Genzer, J. Tuning the number density of nanoparticles by multivariant tailoring of attachment points on flat substrates. *Nanotechnology* **18**, 025301 (2007).
50. Valtiner, M., Banquy, X., Kristiansen, K., W. Greene, G. & N. Israelachvili, J. The Electrochemical Surface Forces Apparatus: The Effect of Surface Roughness, Electrostatic Surface Potentials, and Anodic Oxide Growth on Interaction Forces, and Friction between Dissimilar Surfaces in Aqueous Solutions. *Langmuir* **28**, 13080–13093 (2012).
51. Ye, X. *et al.* Improved Size-Tunable Synthesis of Monodisperse Gold Nanorods through the Use of Aromatic Additives. *ACS Nano* **6**, 2804–2817 (2012).
52. Kim, J.-Y. *et al.* Dipole-like electrostatic asymmetry of gold nanorods. *Sci Adv* **4**, e1700682 (2018).
53. Ziatdinov, M. *et al.* Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. *ACS Nano* **11**, 12742–12752 (2017).
54. Ziatdinov, M. *et al.* Quantifying the Dynamics of Protein Self-Organization Using Deep Learning Analysis of Atomic Force Microscopy Data. *Nano Lett* **21**, 158–165 (2020).
55. Kalinin, S. V. *et al.* Exploring particle dynamics during self-organization processes via rotationally invariant latent representations. (2020).

56. Kalinin, S. V *et al.* Disentangling Rotational Dynamics and Ordering Transitions in a System of Self-Organizing Protein Nanorods via Rotationally Invariant Latent Representations. *ACS Nano* **15**, 6471–6480 (2021).
57. Chollet, F. *Deep learning with Python*. (2018).
58. Maxim Ziatdinov. AtomAI, GitHub repository. https://github.com/pycroscopy/atomai.
59. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014).
60. Duda, R. O. & Hart, P. E. Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Commun. ACM* **15**, 11–15 (1972).
61. Yaman, M. GitHub repository. https://github.com/yamanmy/Automated_image_tool_for_Au_PF_image.
62. Ye, X. *et al.* Improved Size-Tunable Synthesis of Monodisperse Gold Nanorods through the Use of Aromatic Additives. *ACS Nano* **6**, 2804–2817 (2012).
63. Shen, H. *et al.* De novo design of self-assembling helical protein filaments. *Science (1979)* **362**, 705 (2018).
64. Karker, N., Dharmalingam, G. & Carpenter, M. A. Thermal energy harvesting plasmonic based chemical sensors. *ACS Nano* **8**, 10953–10962 (2014).
65. Smith, J. G., Faucheaux, J. A. & Jain, P. K. Plasmon resonances for solar energy harvesting: a mechanistic outlook. *Nano Today* **10**, 67–80 (2015).
66. Du, M. & Tang, G. H. Plasmonic nanofluids based on gold nanorods/nanoellipsoids/nanosheets for solar energy harvesting. *Solar Energy* **137**, 393–400 (2016).
67. Hamed, M. S. G., Ike, J. N. & Mola, G. T. Plasmonic nano-particles mediated energy harvesting in thin-film organic solar cells. *J Phys D Appl Phys* **55**, 015102 (2021).
68. P. Kulkarni, A., M. Noone, K., Munechika, K., R. Guyer, S. & S. Ginger, D. Plasmon-Enhanced Charge Carrier Generation in Organic Photovoltaic Films Using Silver Nanoprisms. *Nano Lett* **10**, 1501–1505 (2010).
69. Yao, K. *et al.* A General Route to Enhance Polymer Solar Cell Performance using Plasmonic Nanoprisms. *Adv Energy Mater* **4**, 1400206 (2014).
70. Huang, X., H. El-Sayed, I., Qian, W. & A. El-Sayed, M. Cancer Cell Imaging and Photothermal Therapy in the Near-Infrared Region by Using Gold Nanorods. *J Am Chem Soc* **128**, 2115–2120 (2006).
71. M. Gobin, A. *et al.* Near-Infrared Resonant Nanoshells for Combined Optical Imaging and Photothermal Cancer Therapy. *Nano Lett* **7**, 1929–1934 (2007).
72. Choi, M.-R. *et al.* A Cellular Trojan Horse for Delivery of Therapeutic Nanoparticles into Tumors. *Nano Lett* **7**, 3759–3765 (2007).
73. Maier, S. A. *et al.* Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides. *Nat Mater* **2**, 229–232 (2003).
74. Jiang, Bosnick, K., Maillard, M. & Brus, L. Single Molecule Raman Spectroscopy at the Junctions of Large Ag Nanocrystals. *J Phys Chem B* **107**, 9964–9972 (2003).
75. L. Rosi, N. & A. Mirkin, C. Nanostructures in Biodiagnostics. *Chem Rev* **105**, 1547–1562 (2005).
76. Nikoobakht, B. & A. El-Sayed, M. Surface-Enhanced Raman Scattering Studies on Aggregated Gold Nanorods. *J Phys Chem A* **107**, 3372–3378 (2003).
77. Chen, J. I. L., Durkee, H., Traxler, B. & Ginger, D. S. Optical Detection of Protein in Complex Media with Plasmonic Nanoparticle Dimers. *Small* **7**, 1993–1997 (2011).

78. Maier, S. A. *et al.* Plasmonics—A Route to Nanoscale Optical Devices. *Advanced Materials* **13**, 1501–1505 (2001).
79. Huang, W. Y., Qian, W. & El-Sayed, M. A. Gigahertz Optical Modulation Resulting from Coherent Lattice Oscillations Induced by Femtosecond Laser Pumping of 2D Photonic Crystals of Gold-Capped Polystyrene Microspheres. *Advanced Materials* **20**, 733–737 (2008).
80. N. Guye, K. *et al.* Importance of Substrate–Particle Repulsion for Protein-Templated Assembly of Metal Nanoparticles. *Langmuir* **37**, 9111–9119 (2021).
81. Yaman, M. Y. *et al.* Alignment of Au nanorods along de novo designed protein nanofibers studied with automated image analysis †. *Soft Matter* **17**, 6109 (2021).
82. Kuzyk, A. *et al.* DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483**, 311–314 (2012).
83. Shen, X. *et al.* Rolling Up Gold Nanoparticle-Dressed DNA Origami into Three-Dimensional Plasmonic Chiral Nanostructures. *J Am Chem Soc* **134**, 146–149 (2011).
84. M. Funston, A., Novo, C., J. Davis, T. & Mulvaney, P. Plasmon Coupling of Gold Nanorods at Short Distances and in Different Geometries. *Nano Lett* **9**, 1651–1658 (2009).
85. Gandra, N., Abbas, A., Tian, L. & Singamaneni, S. Plasmonic Planet–Satellite Analogues: Hierarchical Self-Assembly of Gold Nanostructures. *Nano Lett* **12**, 2645–2651 (2012).
86. Kim, J.-Y. *et al.* Assembly of Gold Nanoparticles into Chiral Superstructures Driven by Circularly Polarized Light. *J Am Chem Soc* **141**, 11739–11744 (2019).
87. Yao, L., Ou, Z., Luo, B., Xu, C. & Chen, Q. Machine Learning to Reveal Nanoparticle Dynamics from Liquid-Phase TEM Videos. *ACS Cent Sci* **6**, 1421–1430 (2020).
88. I. L. Chen, J., Chen, Y. & S. Ginger, D. Plasmonic Nanoparticle Dimers for Optical Sensing of DNA in Complex Media. *J Am Chem Soc* **132**, 9600–9601 (2010).
89. Samai, S., Qian, Z., Ling, J., N. Guye, K. & S. Ginger, D. Optical Properties of Reconfigurable Polymer/Silver Nanoprism Hybrids: Tunable Color and Infrared Scattering Contrast. *ACS Applied Materials & Interfaces* **10**, 8976–8984 (2018).
90. Koh, A. L. *et al.* Electron energy-loss spectroscopy (EELS) of surface plasmons in single silver nanoparticles and dimers: influence of beam damage and mapping of dark modes. *ACS Nano* **3**, 3015–3022 (2009).
91. Szczerbiński, J., Gyr, L., Kaeslin, J. & Zenobi, R. Plasmon-driven photocatalysis leads to products known from E-beam and X-ray-induced surface chemistry. *Nano Lett* **18**, 6740–6749 (2018).
92. K. Jain, P., Huang, W. & A. El-Sayed, M. On the Universal Scaling Behavior of the Distance Decay of Plasmon Coupling in Metal Nanoparticle Pairs: A Plasmon Ruler Equation. *Nano Lett* **7**, 2080–2088 (2007).
93. T. Hill, R. *et al.* Plasmon Ruler with Angstrom Length Resolution. *ACS Nano* **6**, 9237–9246 (2012).
94. Na, L., Mario, H., Thomas, W., Paul, A. A. & Harald, G. Three-Dimensional Plasmon Rulers. *Science (1979)* **332**, 1407–1410 (2011).
95. Malkiel, I. *et al.* Plasmonic nanostructure design and characterization via Deep Learning. *Official journal of the CIOMP* 2047–7538 doi:10.1038/s41377-018-0060-7.
96. Wiecha, P. R., Lecestre, A., Mallet, N. & Larrieu, G. Pushing the limits of optical information storage using deep learning. doi:10.1038/s41565-018-0346-1.

97. Zhou, J., Huang, B., Yan, Z. & Bünzli, J.-C. G. Emerging role of machine learning in light-matter interaction. *Official journal of the CIOMP* 2047–7538 doi:10.1038/s41377-019-0192-4.
98. Tan, E. X. *et al.* Incorporating plasmonic featurization with machine learning to achieve accurate and bidirectional prediction of nanoparticle size and size distribution. *Nanoscale Horiz* **7**, 626–633 (2022).
99. Zhang, T. *et al.* Efficient spectrum prediction and inverse design for plasmonic waveguide systems based on artificial neural networks. *Photonics Res* **7**, 368–380 (2019).
100. Kojima, K., Koike-Akino, T., Tang, Y. & Wang, Y. Inverse design for integrated photonics using deep neural network. in *OSA Advanced Photonics Congress 2021* IF3A.6 (Optica Publishing Group, 2021). doi:10.1364/IPRSN.2021.IF3A.6.
101. Avci, O., Yurdakul, C. & Selim Ünlü, M. Nanoparticle classification in wide-field interferometric microscopy by supervised learning from model. *Appl Opt* **56**, 4238–4242 (2017).
102. Sun, M. *et al.* Salt-induced aggregation of gold nanoparticles for photoacoustic imaging and photothermal therapy of cancer †. *Nanoscale* **8**, 4452 (2016).
103. Lance Kelly, K., Coronado, E., Lin Zhao, L. & C. Schatz, G. The Optical Properties of Metal Nanoparticles: The Influence of Size, Shape, and Dielectric Environment. *J Phys Chem B* **107**, 668–677 (2002).
104. Tabor, C., Murali, R., Mahmoud, M. & A. El-Sayed, M. On the Use of Plasmonic Nanoparticle Pairs As a Plasmon Ruler: The Dependence of the Near-Field Dipole Plasmon Coupling on Nanoparticle Size and Shape. *J Phys Chem A* **113**, 1946–1953 (2008).
105. Hee Yoon, J., Selbach, F., Schumacher, L., Jose, J. & Schlücker, S. Surface Plasmon Coupling in Dimers of Gold Nanoparticles: Experiment and Theory for Ideal (Spherical) and Nonideal (Faceted) Building Blocks. *ACS Photonics* **6**, 642–648 (2019).
106. Hee Yoon, J., Selbach, F., Schumacher, L., Jose, J. & Schlücker, S. Surface Plasmon Coupling in Dimers of Gold Nanoparticles: Experiment and Theory for Ideal (Spherical) and Nonideal (Faceted) Building Blocks. *ACS Photonics* **6**, 642–648 (2019).
107. Zhang, Q., Li, G.-C., Lo, T. W. & Lei, D. Y. Polarization-resolved optical response of plasmonic particle-on-film nanocavities. *Journal of Optics* **20**, 024010 (2018).
108. Murphy, C. J. *et al.* Anisotropic Metal Nanoparticles: Synthesis, Assembly, and Optical Applications. *J. Phys. Chem. B* **109**, 13857 (2005).
109. Liu, Y. *et al.* Decoding the shift-invariant data: applications for band-excitation scanning probe microscopy *. *Mach Learn Sci Technol* **2**, 045028 (2021).
110. Roccapriore, K. M., Ziatdinov, M., Cho, S. H., Hachtel, J. A. & Kalinin, S. V. Predictability of Localized Plasmonic Responses in Nanoparticle Assemblies. *Small* **17**, 2100181 (2021).
111. Ziatdinov, M. *et al.* Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. *ACS Nano* **11**, 12742–12752 (2017).
112. Kalinin, S. v. *et al.* Exploring particle dynamics during self-organization processes via rotationally invariant latent representations. (2020).
113. V, K. S., Ondrej, D., Stephen, J. & Maxim, Z. Exploring order parameters and dynamic processes in disordered systems via variational autoencoders. *Sci Adv* **7**, eabd5084 (2021).

114. Oxley, M. P. *et al.* Probing atomic-scale symmetry breaking by rotationally invariant machine learning of multidimensional electron scattering. *NPJ Comput Mater* **7**, 65 (2021).
115. Liu, Y., Proksch, R., Wong, C. Y., Ziatdinov, M. & Kalinin, S. V. Disentangling Ferroelectric Wall Dynamics and Identification of Pinning Mechanisms via Deep Learning. *Advanced Materials* **33**, 2103680 (2021).
116. Kalinin, S. v *et al.* Deep Bayesian local crystallography. *NPJ Comput Mater* **7**, 181 (2021).
117. Ziatdinov, M., Wong, C. Y. & Kalinin, S. v. Finding simplicity: unsupervised discovery of features, patterns, and order parameters via shift-invariant variational autoencoders. *arXiv:2106.12472 [cond-mat]* (2021).
118. Kalinin, S. V, Kelley, K., Vasudevan, R. K. & Ziatdinov, M. Toward Decoding the Relationship between Domain Structure and Functionality in Ferroelectrics via Hidden Latent Variables. *ACS Appl Mater Interfaces* **13**, 1693–1703 (2021).
119. Roccapriore, K. M. *et al.* Predictability of Localized Plasmonic Responses in Nanoparticle Assemblies. (2021) doi:10.1002/smll.202100181.
120. Ziatdinov, M., Yaman, M. Y., Liu, Y., Ginger, D. & Kalinin, S. V. Semi-supervised learning of images with strong rotational disorder: assembling nanoparticle libraries. (2021).
121. Ziatdinov, M. & Kalinin, S. Robust Feature Disentanglement in Imaging Data via Joint Invariant Variational Autoencoders: from Cards to Atoms. (2021).
122. Amit Singhal. Modern information retrieval: a brief overview. *BULLETIN OF THE IEEE COMPUTER SOCIETY TECHNICAL COMMITTEE ON DATA ENGINEERING* **24**, (2001).
123. Hershey, J. R. & Olsen, P. A. Approximating the Kullback Leibler divergence between Gaussian mixture models. in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* vol. 4 IV–317 (IEEE, 2007).
124. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
125. Yan, Y., I. L. Chen, J. & S. Ginger, D. Photoswitchable Oligonucleotide-Modified Gold Nanoparticles: Controlling Hybridization Stringency with Photon Dose. *Nano Lett* **12**, 2530–2536 (2012).
126. Yoon, J. H., Selbach, F., Schumacher, L., Jose, J. & Schlücker, S. Surface Plasmon Coupling in Dimers of Gold Nanoparticles: Experiment and Theory for Ideal (Spherical) and Nonideal (Faceted) Building Blocks. *ACS Photonics* **6**, 642–648 (2019).
127. Samai, S., Choi, T. L. Y., Guye, K. N., Yan, Y. & Ginger, D. S. Plasmonic Nanoparticle Dimers with Reversibly Photoswitchable Interparticle Distances Linked by DNA. *The Journal of Physical Chemistry C* **122**, 13363–13370 (2018).
128. Crut, A., Maioli, P., del Fatti, N. & Vallée, F. Optical absorption and scattering spectroscopies of single nano-objects. *Chem Soc Rev* **43**, 3921–3956 (2014).
129. Schopf, C. *et al.* Self-Assembly of Gold Nanocrystals into Discrete Coupled Plasmonic Structures. doi:10.3390/cryst6090117.
130. Pal, S. K., Chatterjee, H. & Ghosh, S. K. Manipulating the confinement of electromagnetic field in size-specific gold nanoparticles dimers and trimers. *RSC Adv* **9**, 42145–42154 (2019).

131. Dosselmann, R. & Yang, X. D. A comprehensive assessment of the structural similarity index. *Signal Image Video Process* **5**, 81–91 (2011).
132. Chang, W.-S., Ha, J. W., Slaughter, L. S. & Link, S. Plasmonic nanorod absorbers as orientation sensors. *Proceedings of the National Academy of Sciences* **107**, 2781–2786 (2010).
133. Mahmoud, M. A. Controlling the orientations of gold nanorods inside highly packed 2D arrays. *Physical Chemistry Chemical Physics* **16**, 26153–26162 (2014).
134. Zhang, H. *et al.* Direct Assembly of Vertically Oriented, Gold Nanorod Arrays. *Adv Funct Mater* **31**, 2006753 (2021).
135. Sherman, Z. M. *et al.* Plasmonic Response of Complex Nanoparticle Assemblies. *Nano Lett* **23**, 3030–3037 (2023).
136. Nepal, D. *et al.* Control over Position, Orientation, and Spacing of Arrays of Gold Nanorods Using Chemically Nanopatterned Surfaces and Tailored Particle–Particle–Surface Interactions. *ACS Nano* **6**, 5693–5701 (2012).
137. Yaman, M. Y., Kalinin, S. V, Guye, K. N., Ginger, D. S. & Ziatdinov, M. Learning and Predicting Photonic Responses of Plasmonic Nanoparticle Assemblies via Dual Variational Autoencoders. *Small* **19**, 2205893 (2023).
138. Ovchinnikov, O. S., Jesse, S., Bintacchit, P., Trolier-McKinstry, S. & Kalinin, S. V. Disorder Identification in Hysteresis Data: Recognition Analysis of the Random-Bond--Random-Field Ising Model. *Phys Rev Lett* **103**, 157203 (2009).
139. Johnson, P. B. & Christy, R. W. Optical Constants of the Noble Metals. *Phys Rev B* **6**, 4370–4379 (1972).
140. König, T. A. F. *et al.* Electrically Tunable Plasmonic Behavior of Nanocube–Polymer Nanomaterials Induced by a Redox-Active Electrochromic Polymer. *ACS Nano* **8**, 6182–6192 (2014).
141. Alba-Molina, D., Giner-Casares, J. J. & Cano, M. Bioconjugated Plasmonic Nanoparticles for Enhanced Skin Penetration. *Top Curr Chem* **378**, 8 (2019).