

White Paper: “Computational Microscopy” (IPAM Long Program, Fall 2022)

Kristopher K. Barr, Benjamin Berkels, Peter Binev, Ruiming Cao, Willem Diepeveen, Marc Aurèle Gilles, Ziyang Hu, Ivo Ihrke, Daniel Jacobs, Béatrice Lessard-Hamel, Siting Liu, Tianhan Liu, Nicholas Marshall, Tingwei Meng, Jianwei (John) Miao, Deanna Needell, Elisa Negrini, Raissa Oblitas, Josiah Park, Minh Pham, Monika Ritsch-Marte, Bogdan Toader, Eric Verbeke, Michael B. Wakin, Paul S. Weiss

Table of Contents:

1. Executive summary
2. Simulation for electron and optical microscopy
3. Inverse problems in cryo-EM and phase retrieval
4. AI & Learning theory
5. Data-driven information extraction from microscopy data
6. Multimodal data processing and acquisition
7. Space-time models
8. Geometry in data processing for microscopy

1. Executive summary

This whitepaper summarizes the activities and outcomes of the long program on “Computational Microscopy” at the Institute of Pure and Applied Mathematics (IPAM) from September 12 to December 16, 2022.

For more than three centuries, lens-based microscopy, such as optical, phase-contrast, fluorescence, confocal, and electron microscopy, has played an important role in the evolution of modern science and technology. In 1999, a novel form of microscopy, known as coherent diffractive imaging (CDI), was developed and transformed our traditional view of microscopy, as the diffraction pattern of a noncrystalline object was first measured and then directly phased to obtain a high-resolution image. The well-known phase problem—the usually unavoidable loss of phase information in the diffraction intensity—was solved by a combination of coherent illumination and computational algorithms. Over the years, various CDI methods including plane-wave CDI, ptychography (i.e. scanning CDI), Bragg CDI and Fourier ptychography have been broadly implemented using synchrotron radiation, X-ray free electron lasers, high harmonic generation, and optical and electron microscopy. Furthermore, the 2017 Nobel Prize in chemistry was awarded to Richard Henderson, Joachim Frank, and Jacques Dubochet for developing cryo-electron microscopy (cryo-EM) for the high-resolution structure determination of biomolecules in solution. All these groundbreaking developments require the use of advanced computational algorithms and mathematical tools. This IPAM long program brought together senior and junior applied mathematicians, physicists, chemists, materials scientists, engineers, and biologists to discuss and debate on the current status and future perspectives of modern microscopy using computation, mathematics, and modeling. The program hosted four workshops focusing on different aspects of computational microscopy:

- Workshop I: “Diffractive Imaging with Phase Retrieval” focused on advanced computational methods to solve the phase problem using iterative algorithms and deep learning.
- Workshop II: “Mathematical Advances for Multi-Dimensional Microscopy” focused on the incorporation of state-of-the-art mathematical and computational methods into multi-dimensional electron microscopy.
- Workshop III: “Cryo-Electron Microscopy and Beyond” focused on the current challenges and future perspectives of the cryo-EM field.
- Workshop IV: “Multi-Modal Imaging with Deep Learning and Modeling” focused on the integration of data acquisition, mathematical modeling, and deep learning in multimodal microscopy.

In addition to these four workshops, we formed seven working groups, including 1) [Simulation for electron and optical microscopy](#), 2) [Inverse problems in cryo-EM and phase retrieval](#), 3) [AI & learning theory](#), 4) [Data-driven information extraction from microscopy data](#), 5) [Multimodal data processing and acquisition](#), 6) [Space-time models](#), and 7) [Geometry in data processing for microscopy](#). The working groups met regularly during the program and tackled a number of outstanding problems in the field. Below we provide the open challenges that we identified in computational microscopy, the progress that we made at IPAM, and the research directions that we will continue to investigate in the future.

2. Simulation for electron and optical microscopy

2.1 Introduction

Experimental data is acquired by real instruments that impose their technical parameters, constraints, and non-idealities on the measurements of the samples being studied. Scientists are usually just interested in the aspects of the sample itself. In order to disambiguate the effects, simulation is an important tool. In order to exploit its full potential, a) accurate device and environmental models, b) their parameters, as well as c) a definition of the classes of samples that are of interest are required in any particular problem under study.

A physically rigorous simulation on the intended application scales is, unfortunately, often intractable. Therefore, appropriate choices need to be made in particular applications, the most important of which we discuss below in sections [Device and Sample Models](#), [Modeling Propagation](#), and [Simulation of Transmission Electron Microscopy](#).

The main application areas of simulation are device design and parameter selection, the generation of ground truth data, [Simulation for Inverse Problems](#), and more recently the generation of training data for machine learning, [Simulation for Learning](#).

2.2 Device and sample models

While the models and methods for device simulation in optics and electron optics are to some extent similar, the parameter spaces encountered are quite different. In optics, a large numerical aperture (NA) is necessary to achieve high resolution whereas in electron microscopy higher energies and relatively low NAs are used. This is reflected in the simulation challenges on the device side. High-NA optical simulations have strong discretization requirements and need to

handle non-paraxial settings efficiently. In both optical imaging and electron optics device simulation, accurate knowledge of the aberrations is necessary.

Equally important to the simulation process are representative sample models. In the case of atomic resolution transmission electron microscopy (TEM), these consist of 3D atomic potentials for different types of atoms in crystalline granular structures or for amorphous materials. A variety of methods exist for calculating a 3D potential energy function with different trade-offs. The dynamic imaging environment, which can include vitreous ice, carbon contamination, and atomic motion, adds significant complexity in the number of atomic potentials and degrees of freedom that need to be handled. In Workshop III, Khoo suggested a modeling method using tensor networks to simulate transition paths between metastable states, which is particularly useful for the study in molecular dynamics. For optical microscopy, samples often come from biology. Important simulation cases are unstained cells which are usually treated as thick phase objects, possibly with some absorption. A current trend in this area is towards the imaging of 3D cell clusters, cancer spheroids, and organoids. Obtaining realistic sample models for these optically diffuse samples is difficult. Other important settings are the imaging of fluorescently labeled specimens, especially in strongly scattering settings such as brain tissue.

While initial work has been done on sample models, the full complexity of the real acquisition settings has not been reached. Important parameters such as realistic phase gradients, potential profiles in complex settings, and their susceptibility to temporal change are not well understood, requiring the interaction of experimental observation and simulation for improvements.

2.3 Modeling propagation

2.3.1 Physical effects and model accuracy

The choice of mathematical model underlying a simulation influences the replicability of observed physical phenomena as well as the simulation time and memory complexity and thus its suitability for online or offline use. Imaging techniques operating on different scales (from μm to atomic-scale) face different challenges and cross-fertilization between them may be beneficial but is largely unexplored. Moreover, sample complexity may vary significantly.

In optical applications, light is propagated through a target sample. The samples are volumetric with spatially varying scattering potentials, so multiple scattering must be considered. This aspect was covered as a theme of Workshop I. A rigorous solution can be obtained from the Lippmann-Schwinger equations and the first applications in inverse scattering have appeared (Pham *et al.*, *Trans. Comp. Img.*, 2020, G. Osnabrugge *et al.*, *J. Comput. Phys.*, 2016). More standard treatments employ the Born or Rytov approximations. The latter only model single

forward scattering and are hardly suitable for samples of non-negligible thickness or pronounced refractive index contrast. Modifications of the Born approximation that are more accurate appear for forward and inverse modeling use. Another important but computationally expensive feature of optical microscopic imaging is partial coherence.

In electron microscopy, multiple scattering effects ultimately limit the sample thickness to be investigated. Currently, the multislice simulation method is the method of choice for practical applications. This method decomposes a volumetric object into a set of layers, which are sequentially treated as 2D scattering planes. It can therefore model forward multiple scattering but relies on an appropriate layer sampling. At current electron energies of about 300 keV, backward scattering followed by subsequent forward scatter is negligible. Lower electron beam energies and thicker samples may change this.

2.3.2 Noise modeling

Accurate physical modeling necessitates the characterization of the noise type and level. Technical noise, e.g. from the detection system, needs to be distinguished from stochastic noise resulting, e.g. from the interaction of the sample with the probing beam such as jitter of the excitation beam, motion of the particle, and similar effects. Typical detector noise models consist of a larger number of sources that are usually assumed to be statistically independent and therefore separately treated. More complex noise sources due to sample interaction and/or device changes are often handled using Monte-Carlo approaches.

In addition to the sample changes discussed in [Section 2.2](#), sample damage from energy deposited by the beam can interfere with image reconstruction and analysis. This issue was discussed at length in Workshop III for TEM and 4D-scanning transmission electron microscope (4D-STEM) modes. On the hardware-level different scanning strategies can be employed to enable a relaxation period in different spatial positions before rescanning neighboring points as presented by Reed in Workshop II. In optical microscopy, light intensity has to stay below a certain level to avoid sample damage. Laser scanning microscopy, e.g. involves considerable intensities. Nonlinear optical microscopy techniques require high laser power, which is potentially harmful to delicate samples. In fluorescence microscopy the fluorescent labels undergo bleaching which grows with the optical power on the sample.

2.4 Simulation of transmission electron microscopy (TEM)

Simulation of possible structures is an integral part in analyzing data from TEM and STEM. A precise mathematical model involves the Dirac equation whose numerical solution requires enormous computational work that cannot be performed on current computational platforms. The

usual approach is to solve an approximation of the Dirac equation known as the Schrödinger equation for fast electrons. The solution of this equation provides an exit wave from which the desired results for the (S)TEM simulations are derived. The general problem is to derive this wave from given potential functions or atomic structures and the initial wave expressing the electron wave before entering the specimen. The dominant numerical approach is called the “multislice method” (E. Kirkland, *Advanced Computing in Electron Microscopy*, 3rd ed., 2020), a variant of the forward Euler method. This increases computational efficiency while maintaining accuracy in most cases. However, the computational cost is often too high, especially for STEM, where each position of the scanning probe yields an initial wave for which a different solution must be calculated.

One possible improvement of the computational efficiency of the multislice algorithm for STEM can be achieved by the PRISM method that expresses each scanning probe exit wave as a linear combination of a set of basic waves (harmonics of limited frequencies). Since the number of basic waves is much lower than the number of probes, PRISM has better computational efficiency than the standard multislice method, albeit with lower accuracy. One possible improvement of PRISM advanced during the IPAM Long Program on Computational Microscopy is composing a set of basic waves consisting of well-localized trigonometric functions instead of the global harmonics. The use of local initial waves allows a noticeable reduction of the amount of computations to obtain the corresponding exit waves without decreasing the accuracy significantly. This allows an implementation with a relatively large amount of local basic waves that is more computationally efficient than PRISM and maintains good accuracy.

In the simulation of STEM, the potential term in the Schrödinger equation depends on the configuration of atoms in the specimen. However, due to atomic temperature vibrations, the potential term also changes as a function of time during data acquisition. To decrease the computational complexity, this dependence is approximated by randomly perturbing each of the atoms, calculating hundreds of realizations, and averaging the results. For these repetitive simulation procedures, existing approaches typically carried out each realization independently. An alternative method is introducing an iterative process to solve the equation, with an initial guess possibly precalculated using an existing approach. In such a way, there is a possibility of improvement in both computational cost and efficiency. On the other hand, the multislice method computes the Schrödinger equation by tracing the two-dimensional electron wave as it advances through the domain. A numerical method that solves the whole physical domain brings potential improvement in computational accuracy, especially toward better diffraction results. One idea here is to cast the Schrödinger equation to a min-max problem by introducing a Lagrange multiplier. An implicit discretization scheme (e.g., backward Euler method) can be used to discretize the equation into a min-max problem solvable with iterative methods such as the

primal-dual hybrid gradient descent method. Alternative methods can produce a single solution to serve as an initial guess to solve the Schrödinger equation again with perturbed potential. This framework can potentially improve accuracy and efficiency while maintaining a reasonable computational cost in the context of STEM.

2.5 Simulation for inverse problems

In the inverse problem setting, simulation acquires the role of the forward operator. Many optimization techniques are formulated in terms of operators defined by the forward model, e.g. the direct application of the forward imaging operator, its Hermitian conjugate, or its gradient. Since the forward operator is called in the main loop of the optimization algorithms, we refer to this as an online simulation setting. Computational efficiency is a key consideration, since inverse problem solvers perform iterative updates with many iterations.

In generic optical and electron optical microscopy, the intensity of the wave field incident on the sensor is measured, resulting in a non-linear image formation model. For tractability in an inverse problem setting, differentiability of the forward model is a major feature. The majority of the significant physical effects should be incorporated into the forward operator while maintaining computational feasibility. In electron microscopy with samples of standard thickness, multislice has seen good agreement with experiments. In optical settings several multislice methods exist, such as the “beam propagation method” or the “wave propagation method”. Multiple scattering effects can be sufficiently pronounced to render multislice methods problematic for accurate inversion (Lee et al. *Optica* 9(2), 2022). Therefore, currently considerable research effort is directed towards finding efficient and accurate propagators for this setting.

A second use of simulation in inverse problems is the generation of realistic ground truth data for the development, verification, and comparison of different types of optimization algorithms. In this case, the highest possible accuracy is desirable, whereas the required computation times are less relevant. We refer to this setting as an offline simulation setting.

2.6 Simulation for learning

Many image reconstruction methods use optimization principles to solve inverse problems for a given forward model. Deep learning has emerged as a popular approach to such inverse problems in microscopy. This approach requires large amounts of training data, which can be obtained either through experiments or simulations. Experimental data encapsulates a truly exact forward physical process, but it can be expensive to obtain and often fails to generalize beyond the

specific conditions present during data acquisition, requiring a new training process for every experiment. Covering a large variety of application settings is often difficult.

As such, simulated training data enables more robust, generalizable neural networks for processing microscopy data given a well-defined forward model. The applicability of deep learning methods to cover real experimental data, however, depends heavily on the accuracy of the forward model. Further progress in this area requires improving the completeness of forward models in electron and optical microscopy.

2.7 Work done at IPAM

One of the projects to emerge from this collaboration focuses on developing a forward model for electron ptychography of cryogenic biological samples. Recent applications of ptychography to biological samples have shown promising results but are not yet competitive with established methods in cryo-EM or cryo-ET. Further improvements to this model include an accurate treatment of sample drift during the scan process, which can prevent accurate reconstruction by invalidating the overlap constraints necessary for ptychographic phase retrieval. We hope to use this forward model to explore the potential for ptychography in cryo-electron tomography and to inform robust reconstruction methods for biological samples.

The second project that was considered is a simulation setting for optical Fourier ptychography in the presence of thick samples. We identified a shortcoming in current Fourier ptychographic modeling: the microscope is usually treated as having no internal obstructions. As such, vignetting effects that are prominently presented in acquired data from real experiments are not taken into account by current models. Challenges for developing the simulation tools requires dealing with a) cascaded diffraction effects from different obstructing parts of the microscope, b) handling the partial coherence of the LED illumination, and c) efficiently and accurately treating high-NA.

The third project that was considered is an iterative method for simulating STEM. As discussed earlier, an iterative approach has the potential to improve the computational efficiency as well as accuracy with proper numerical discretization approaches.

The fourth project that was considered is an improvement of the PRISM approach to STEM simulation. The well localized basis of bandlimited trigonometric functions is used to define a set of functions to replace the basis of standard harmonics used in PRISM. The code is in the final stages of its development and will be made available to the Computational Microscopy community when completed.

2.8 Outlook

In the following, we derive some insights for driving forces in the field and likely future developments, while outlining the associated challenges and potential impact. We have left aside a number of open issues such as the simulation of transient phenomena, Raman and hyperspectral imaging, the relation between cryo-EM and NMR/EPR analysis, and others.

Opportunities & Challenges: Advances in available computing power have long pushed the envelope of technical feasibility in simulations. The dominating way of increasing compute capability today is a hierarchical organization of parallelism with different processing models. Multi-threaded GPU computations via multi-GPU cluster nodes and classical inter-node parallelism are popular ways of enhancing throughput. This presents both an opportunity and a challenge since engineering massively parallel software is a major task.

At the same time, the massive increase in compute power has given rise to dramatic improvements in machine learning, in particular deep learning. Harvesting the power of these techniques for function approximation in the context of simulation promises to be a major innovation potentially enabling fast evaluations of complex imaging models and their inverses. Challenges in this domain are the assurance of accurate output for new samples as well as a possible boot-strapping due to limited availability of real data. A major challenge to this scheme is the curse of dimensionality as we strive for the description of ever higher dimensional data, e.g. 4D + time in future 4D-STEM, 3D + spectral in 3D fluorescence microscopy with multiple labels, etc.

Generally, improved observation techniques enabled by simulation components rest on available knowledge about the investigated system. We envision a positive feedback loop between improved observation capabilities, enhanced understanding of physical sources, and improved modeling and simulation. This feedback loop could enable studies of samples that are currently beyond the scope of state-of-the-art imaging. A challenge to be overcome is a two-fold “language barrier”, a) identifying common causes in different modalities which may yield complementary insights and b) a language barrier within the interdisciplinary field of computational imaging, since approaches in different disciplines using their specific technical terms need to be merged.

Potential Impact: Improved simulation techniques will be crucial for the improvement of all imaging techniques based on inverse formulations. Arguably, the term “computational microscopy” implies this dependence. Progress will be made along two main lines. First, simulation speed improvements, which may accelerate methods to a point that simulation techniques today restricted to offline use can be employed in an online fashion in inverse

problems and online learning. Second, the quality of the simulation will be improved by including increasing amounts of details that are partially still to be extracted from observation. This will be facilitated by programmable microscopes that adapt to sample structures, avoid sample damage, and choose appropriate imaging modes to best highlight structures of interest to an observer. These microscopes may be autonomous, carrying out large protocols of different acquisition steps in sequence with automated analysis steps to provide the high-level information scientists seek.

The road towards this future should exploit the input from a growing community of computational microscopists. Growing this community is best done by sharing the results of our work: open databases for learning, open code for reproducing and extending work by others, and the creation of community resources will benefit the development of the field.

3. Inverse problems in cryo-EM and phase retrieval

3.1 Inverse problems in cryo-EM

3.1.1 Introduction

Electron cryo-microscopy (cryo-EM) is a technique capable of resolving high-resolution structures of biological macromolecules and ultrastructures of whole cells or cell sections which are flash frozen in vitreous ice. The past decade has seen an exponential growth in the number of deposited structures in public databases from cryo-EM due to major advances in hardware and software. Contributions from cryo-EM have been recognized by Nature Methods as the “Method of the Year” in 2015 and with the Nobel Prize in Chemistry in 2017. Current and ongoing impacts from cryo-EM include furthering our understanding of fundamental biology which can have direct benefits to human health. In particular, cryo-EM has recently played a vital role in our understanding of SARS-CoV2.

Cryo-EM can broadly be separated into two main approaches: single particle analysis (SPA) and electron cryo-tomography (cryo-ET). The former aims to recover the high-resolution ($<3 \text{ \AA}$) 3D structure of a target macromolecule from many projection images of the purified target, whereas the latter aims to recover the near-native structure of an entire cell or cell section and its contents. Each method has unique challenges and opportunities which were the focus of Workshop III – Cryo-Electron Microscopy and Beyond. The emergent themes from Workshop III can be categorized into four groups: 1) data automation, validation, and accessibility, 2) heterogeneity analysis and free-energy landscapes in SPA, 3) physical limitations and improvements to hardware, and 4) novel workflows and analysis in cryo-ET.

The specific challenge explored during the long program was heterogeneity analysis in SPA. Below we discuss this challenging inverse problem in detail and describe the work done in topics on cryo-EM while at IPAM.

3.1.2 Heterogeneity analysis

Due to the sample preparation technique employed in cryo-EM, the 3D structures of molecules are closer to their natural states as compared to other microscopy techniques such as crystallography. As a consequence, a unique opportunity presented by cryo-EM is to resolve not only a single structure of a protein but also the distribution of different shapes (called conformations) it can adopt. This capability can reveal unique biological insight; e.g., how proteins achieve their function by deforming into different conformations. However, the reconstruction of such conformations presents a challenging inverse problem since for each highly noisy, corrupted particle image, one needs to infer both the pose and the conformation present in the image. This problem is of unique scientific importance as recognized by being named Nature's Method to Watch in 2022 (A. Doerr, *Nat. Methods*, 2022), and it presents one of the best opportunities for mathematicians to have a large impact in the life sciences. This fact was particularly highlighted in Workshop III, which assembled several practitioners and leading methods developers in the field.

There exist two types of heterogeneity in cryo-EM datasets: discrete heterogeneity, often caused by the presence of protein complexes in different assembly states, and continuous heterogeneity, usually the result of a molecule deforming. Several methods have been proposed to resolve heterogeneity in cryo-EM, including:

- clustering-type methods used for discrete heterogeneity (e.g. in S. Scheres, *J Struct Biol.*, 2012),
- methods based on molecular dynamics (e.g. M. Harastan *et al.*, *Front. Mol. Biosci.*, 2022),
- linear subspace methods (e.g. E. Katsevich *et al.*, *SIAM Imaging*, 2014), and
- deep-learning methods (e.g. E. Zhong *et al.*, *Nat. Methods*, 2021).

Several major challenges were pointed out during the IPAM program. First, there is a lack of theoretical understanding in all of these methods; e.g., there exist no bounds characterizing which signal-to-noise ratios (SNRs) actually permit heterogeneous reconstruction. A second major and compounding obstacle to the understanding of these methods is the lack of a metric on the accuracy of a particular method on a particular dataset. Indeed, while each class of methods provides qualitatively good results for certain datasets, and while these have helped yield insight into biological phenomena, the accuracy has, for the most part, been measured by the “eyeball norm” or prior knowledge. This is, in part, due to the complicated nature of the output of heterogeneous reconstruction algorithms. In most cases, the results take the following form:

- Each image $x_i \in R^{N^2}$ is assigned a “latent coordinate” $z_i \in R^d$ by the algorithm.
- The algorithm learns a function $V(z)$ which maps this latent space to a scattering potential in R^{N^3} .

Hence, a heterogeneity metric must be able to quantify:

- The accuracy of two sets of points in different latent spaces $\{z_i^1\}_{i=1,\dots,n}$ and $\{z_i^2\}_{i=1,\dots,n}$ corresponding to two embeddings of the dataset. These two embeddings produced by two methods (or even the same method) may have different dimensions, and they do not necessarily conserve properties such as pairwise distances.
- The quality of a particular reconstruction $V(z)$. Unfortunately, the Fourier Shell Correlation (FSC), used widely as a metric in homogeneous reconstruction, is uninformative in this situation. This is because different conformations present in the same dataset are often highly correlated; thus the FSC score obtained often reflects the accuracy of the mean conformation rather than the accuracy of a particular conformation.

The invention of accurate metrics in cryo-EM has often led to novel regularization schemes (e.g., the use of the gold standard FSC in S. Scheres, *J Struct Biol.*, 2012); thus a novel metric may not only help interpret the results of a particular heterogeneous reconstruction but also enhance it.

3.1.3 Robust estimation

A simple case of heterogeneity in cryo-EM images arises due to unknown contaminations, such as ice, which typically need a preprocessing procedure (e.g., particle picking) to remove. The unknown contamination might cause traditional moment methods like Kam’s method to fail since moments might not exist or be stably estimated due to the contamination. For instance, contaminations in general elliptical distributions do not even have a finite first moment. Robust estimation or learning with unknown contaminations has been a central topic in statistics since Richard Huber in the 1970s. The well-known Tukey’s median estimator is statistically optimal yet computationally intractable. Therefore robust learning under Huber’s contamination model has become an important field in both statistics and theoretical computer science. At IPAM, Yao talked about their recent discovery that generative adversarial networks (GANs) can be used for robust learning with arbitrary unknown contaminations, and scale up well in high dimensional problems such as cryo-EM robust image denoising. In this approach, a discriminator network plays a role of identifying contaminations and a generator network aims to generate uncontaminated noisy images with possibly continuous heterogeneity. This approach can be regarded as a generalized moment method with nonlinear feature matching. It avoids the

difficulty of possible moment instability and opens the door to robust reconstructions from cryo-EM images with unknown contamination.

3.1.4 Work done at IPAM

Kam’s method with data-driven prior. Kam’s method (Z. Kam, *J. Theor. Biol.*, 1980) is one of the earliest methods proposed for homogeneous reconstruction in cryo-EM. It is the classical statistical method-of-moments applied to the cryo-EM reconstruction problem. Kam’s method has helped push the theoretical understanding of the cryo-EM reconstruction problem, e.g., it is a provable algorithm for reconstruction and provides bounds for the quality of the reconstructions in terms of the noise level and the number of images. Kam’s method has several remarkable properties:

- it bypasses the need for angular assignment, typically a huge computational burden in competing methods,
- it is a streaming algorithm, thus theoretically much faster than iterative methods, and
- it can – in theory – break the detection limit of the minimal size of proteins that can be observed in cryo-EM (T. Bendory *et al.*, *SIAM imaging*, 2022).

While theoretically attractive, Kam’s method has not been able to yield high-resolution reconstructions. One direction that is currently being explored in the literature to resolve this issue is the development of better priors (e.g., based on the sparsity of the solution as in T. Bendory *et al.*, *Arxiv*, 2022) for Kam’s method.

Separately, following AlphaFold’s breakthrough in protein structure prediction, there has been considerable interest in leveraging the vast amount of solved structures stored in the PDB (protein data bank) to improve cryo-EM reconstructions. As a first step in this direction, we propose to craft a “PDB prior” for Kam’s method by matching moments obtained empirically from particle images to moments computed analytically from the known structure in the PDB. To do this, one must first compute the first and second moments for a large number of structures. During the time at IPAM, we started forming this pipeline, and we have created an embedding of 1000 second moments for a specific class of enzymes in the PDB. The next step is to compare these precomputed analytical moments to empirically computed moments, and to account for several nuisance variables including pose distribution, B-factor, and off-centering.

Fast Approximate Wasserstein distance for heterogeneity analysis. A method was recently proposed for manifold learning of conformations using the Wasserstein distance. This method leverages the approximation of the Wasserstein distance by the L_1 distance in a wavelet basis. This method involves computing the pairwise distance between n volumes of size m , thus having cost $O(mn^2)$. In cryo-EM, the size of the volume is typically $m \sim 10^6 - 10^8$, and the number of volumes (equal to the number of images) is $n \sim 10^4 - 10^6$, thus this $O(mn^2)$ cost is prohibitively

expensive. However, some algorithms for heterogeneity output structures that live in a low-dimensional space of dimension $k \ll m$ (often $k < 10$); it is thus of interest to design an algorithm that can leverage this structure to compute these distances faster. To address this problem; we propose a novel algorithm for computing these distances that runs in $O(\log(m)2^k)$, with $O(m)$ precompute based on a tree decomposition. The full pairwise distance could then be computed in $O(m + n^2 \log(m)2^k)$, which is a potentially large improvement over the naive computation for large volumes.

Contrast invariant deep-learned heterogeneity. A recurring theme in Workshop III was the development of equivariant machine learning methods to improve cryo-EM. Following discussions during the workshop, a simple modification of deep-learning networks was proposed that is invariant to the image contrast, a nuisance variable caused by the varying ice thickness in cryo-EM. This simple modification provided encouraging results on simulated datasets.

3.1.5 Outlook

After undergoing a “Resolution Revolution” (W. Kühlbrandt, *Science*, 2014), cryo-EM is on the verge of two new great advances: first, in understanding the motion of molecules through heterogeneity analysis and how it informs on biological function, and second, in observing molecules in their native environment in cells through high-resolution cryo-ET. Realizing these advances in cryo-EM will require developing new mathematical frameworks and adaptation of methods outside of the current paradigm.

3.2 Phase retrieval with TV-S1 regularizer

3.2.1 Introduction

Phase retrieval aims to reconstruct signals whose phases are lost during measurement. With high resolution cameras, oversampling is often employed successfully to recover the desired phase signal. While the signal magnitudes provide the scattering information, the phases deliver more meaningful figures about the specimen such as type of material, thickness, or phase delay. However, in the weak scattering case, the phases are weak; they can not be reconstructed faithfully and contain noisy artifacts. Phase retrieval was a central theme of Workshop I.

A phase retrieval algorithm that can regularize the phase will be beneficial to the analysis of such specimens. Specifically under the assumption of low total variation (TV) in the phase, a TV regularizer on the circle S1 can be applied to achieve a more accurate result. Constraining the minimization on a periodic manifold such as S1 is not trivial and needs to be explored carefully.

The successful investigation of minimization on a periodic domain or manifold will provide a new set of optimization tools for inverse problems.

3.2.2 Work done at IPAM

We are exploring solutions to the problem using two different approaches. The first one is a classical optimization approach with adaptation to the S^1 domain, and the second approach uses a lifting function, a powerful technique proposed by Strekalovskiy, *CVPR* 2011 that exploits the sub-graph function to extend the minimization to arbitrary differentiable sampled manifolds.

In the first approach, we design a primal-hybrid gradient descent (PDHG) algorithm that can work back and forth between the Euclidean space R^n and the cyclic domain S^1 . In the primal step, the algorithm needs to solve a backward Euler (BE) equation on the cyclic domain S^1 . We separate this step into two sub-steps: first solving the BE equation on the Euclidean domain, then projecting this solution onto S^1 . This separation technique yields an equivalent solution to the original primal step when the iterative solution is close enough to the minimum. In the dual step, we invented a two-way shrinkage operator to perform the TV on the cyclic domain S^1 . This method has been tested on synthetic data and performs much faster than the lifting technique. Memory efficiency is also a remarkable advantage of this technique.

While the first approach focuses on optimization, the second one pursues a more mathematical perspective. Via functional lifting, the periodic domain is interpreted as a linear constraint which can be handled easily by primal dual methods. The algorithm is guaranteed to have correct solutions on S^1 and converges monotonically to a unique solution. It has been tested to perform well on synthetic data despite extra memory usage and slow convergence. However, along with dimensional extension, the method requires re-evaluating the sub-graph function at every iteration; hence, the computational expense increases significantly and exposes a big challenge to optimization.

3.2.3 Outlook

In future work, we look forward to optimizing the functional lifting technique, reducing computational costs, and improving the convergence speed. We will then extend the technique to other minimization problems on periodic manifolds, such as synthetic aperture radar (SAR) imaging, and image registration and alignment.

3.3 A general variational framework for some inverse problems in microscopy

3.3.1 Introduction

In microscopy and other natural data processing, the measurement process may be described by operators that are not fully known. For instance, in dark-field electron microscopy, several observations of atoms with the same type may be obtained by scanning the samples line by line. However, the distortion operator applied to each image of an atom is unknown and different from atom to atom. In such settings, an inverse problem may be formulated to recover both the distortion operators and a clear image of the atomic structure. Similar problems appear in other research areas such as natural image processing. For example, an unknown distortion operator is applied when a camera is used to capture a natural image, and a variational model can be used to both restore the clear images and recover the distortion operator. In practice, however, heuristic algorithms are developed independently to solve each such problem. In this focus group, we have aimed to formulate such problems in a unified framework that enables analysis and the development of algorithms for new problem settings.

3.3.2 Work done at IPAM

Yankovich *et al.*, *Nat. Comm.* 2014 proposed a variational method for non-rigid (deformable) image registration. We have generalized this model so that it is suitable for more general problems. In our formulation, the inverse problems are formulated as a minimization problem whose unknown variables are the distortion operators and the clear image that we desire. The objective function of this minimization problem is a summation of data fidelity terms and regularization terms. The data fidelity term is the distance between the observation and the distorted clear image (i.e., the image obtained by applying the unknown distortion operator to the unknown clear image). The regularization terms are on the distortion operators as well as the unknown images. We formulated the minimization problem and derived formally the first order variations. We proposed an algorithm based on gradient descent type algorithms using those first order variations. Additional numerical experiments are planned in the future.

4. AI & Learning theory

4.1 Introduction

Thanks to the substantial improvement in computational power and investments in computer technology, artificial intelligence systems can perform remarkably well on tasks coming from various application areas like healthcare, visual recognition, text analytics, cybersecurity, and many more. Much of this success is due to deep learning, in which deep neural networks are

used as an underlying model for AI. A deep neural network is a parametric function whose optimal parameters are chosen through minimization of a user-determined loss function. This loss function can measure, for example, the distance between the network prediction and the target data. In particular, rather than deciding a priori a set of parameters, a deep neural network is trained on data to learn the optimal parameters for the task at hand.

Despite its success in multiple application areas, deep learning theory is still in its infancy and a mathematical explanation for many of the successful applications of deep learning is still missing. This is true in particular for computational microscopy and computer vision applications where often deep learning methods are used (successfully) as a substitute for traditional iterative algorithms at the expense of mathematical explainability.

Inspired by these challenges, during our time at IPAM we explored the topics of implicit regularization, sparsity, and unrolled networks.

4.2 Implicit regularization of stochastic gradient descent

Stochastic gradient descent (SGD) is a widely used algorithm in machine learning; one study of its behavior was presented in Workshop IV talk by Rachel Ward. A common application of SGD is in training neural networks. In practice, when we train a neural network, the optimization landscape often has many local minima or even many global minima. For reasons which are not fully understood, SGD seems to find a minimum that generalizes well. That is, using SGD results in implicit regularization or an implicit bias towards a “good” local minimum, which is essential for learning useful parameters for overdetermined models. In order to make rigorous choices about optimization methods, it is essential to understand this phenomenon, which creates a source of bias in the output of all neural network models trained using SGD. Recently, many works have focused on studying implicit regularization on simple models. Some analysis of the implicit spectral bias of gradient descent was also presented in Workshop IV by Mahdi Soltanolkotabi.

4.3 Sparsity

Sparsity is a ubiquitous structural constraint in computational imaging and machine learning. For instance, in imaging, the underlying reconstruction object is often sparse in certain representations, exhibiting spatial regularity in piecewise constant structure, smoothness, or spikes. In popular machine learning models, convolutional neural networks have sparse architectures where every neuron only receives local inputs while transformers have self-attentions which adaptively select sparse tokens in processing. In these scenarios, sparsity decides how many measurements one needs to reconstruct the object or learn the model, which is

a fundamental topic in compressive sensing and machine learning. Sparse regularization and TV minimization (sparse gradient regularization) were mentioned by numerous IPAM workshop speakers including Bryan Reed, Stan Osher, Benjamin Berkels, and Paul Weiss.

4.4 Network unrolling

Rather than developing a mathematical theory for existing architectures, network unrolling is a way to develop mathematically principled neural network architectures by using as a building block for the network existing iterative algorithms. In this way, algorithm unrolling provides a concrete and systematic connection between iterative algorithms and deep neural networks and inherits some of the theoretical guarantees of the underlying iterative method. In applications such as MRI, unrolled deep networks have recently seen an increase in popularity due to their potential in developing efficient, high-performance, and interpretable network architectures from reasonably sized training sets. Unrolled algorithms were mentioned by IPAM workshop speakers such as Mahdi Soltanolkotabi and Laura Waller.

4.5 Work done at IPAM

To build a concrete understanding of how implicit bias emerges from SGD in learning problems, we have focused on a simple model for a neural network that is linear in the input vector but nonlinear in the network parameters. This problem is overparameterized, but when data is generated according to a sparse linear model, using SGD to train the network can indeed reveal the correct sparse solution. Our initial discussions have helped us understand the role that sparsity plays in creating such implicit bias, as well as the importance of the initialization of the SGD algorithm. Formal theoretical analysis is ongoing.

Our discussions as well as the IPAM Workshop II talk by Yonina Eldar on network unrolling as a way to design interpretable and mathematically principled architectures resulted in a collaboration with another focus group ([Data-driven information extraction from microscopy data](#)). During our time at IPAM we developed a novel algorithm for segmentation of images of regular lattices. Starting from the primal-dual hybrid gradient algorithm (PDHG) applied to a convex relaxation of the Mumford-Shah optimization problem, we trained a neural network where each layer performs a PDHG iteration and the proximal operators are learned from simulated data. See [Section 5](#) for more details on this project and outlook.

4.6 Outlook

The work done at IPAM has established two research programs: (1) developing an understanding of implicit bias caused by SGD, and (2) developing a novel image segmentation algorithm using

network unrolling. Both projects are under active development and have achieved promising results. Although the implicit bias of SGD and the concept of algorithm unrolling are quite powerful, it will be important to gain a better understanding of both phenomena so that the concepts can be generalized and their limitations can be characterized.

5. Data-driven information extraction from microscopy data

5.1 Introduction

Although there are a variety of imaging modalities (many of which were encountered during this long program), sample images often display similar characteristics. For example, many experiments in materials science involve samples with regions of regular lattice structures separated by disordered regions that reflect the properties of the material. Therefore, developing a computational technique that captures this common characteristic and does not depend on features of a specific imaging modality would be of great use to experimentalists across disciplines.

An illustrative application where such work would show its usefulness is the processing of scanning probe microscopy (SPM) data. Researchers have used SPM techniques to image individual atoms and molecules organized on a variety of substrates (G. Binnig *et al.*, *Angew. Chem. Int. Edit.*, 1987; G. Binnig *et al.*, *Phys. Rev. Lett.*, 1983; L. Gross *et al.*, *Science*, 2009; D. Yugay *et al.*, *Nano Lett.*, 2016). As part of this group, an unrolled network approach such as the one discussed in the [AI and Learning Theory](#) group has been developed for the problem of segmentation of data with regular lattice patterns.

We applied these segmentation techniques on scanning tunneling microscopy (STM) images as they contain well-ordered regions of surfaces with repeating features. However, the most reactive region of the surfaces occurs at the defect sites, (P. Weiss, *Accounts Chem. Res.*, 2008) so the segmentation algorithm needs to distinguish those regions efficiently.

Another SPM technique is atomic force microscopy (AFM), which relies on signal processing to control the feedback controller. While imaging surfaces, one needs to minimize the distortions by adjusting the tip in the vertical direction to keep a constant deflection or amplitude setpoint between the tip and sample. The most widely used controller in AFM is the proportional-integral-derivative (PID) controller (D. Abramovitch *et al.*, *American Control Conference*, 2008).

In the early days, the PID values were fixed in each image and hand-tuned based on the discrepancy between trace and retrace lines and final images. In this case, large errors may arise from the biased judgment from different users and the non-dynamic process is not able to adapt to the varying features in different regions. To optimize the images and expand the bandwidths, different algorithms have been developed (M. Dukic *et al.*, *Rev. Sci. Instrum.*, 2017; H. Liu *et al.*, *Micron.*, 2018). The optimized PID values also make it possible to study the dynamics of images enabling us to use time as an extra degree of freedom.

In this section, we will describe the workflow that can be applied to a multitude of datasets. To show an application of these techniques, we will show how we used algorithms and neural networks to perform image segmentation on SPM images and on simulated data.

5.2 An approach to segmentation of images of regular lattices

Standard image segmentation techniques work by finding the boundary between regions of an image based on the pixel values, i.e., pixels with similar values are part of the same region. In addition, regularization such as total variation (TV) is often imposed on the boundary between regions as part of an optimization procedure. However, this procedure is not applicable when the image domains differ not by pixel values (which are often the same in both regions), but by specific features of the pattern contained within them (e.g., the pattern orientation). Therefore, a general workflow for segmenting such images requires first a feature extraction procedure followed by a segmentation step that uses these features. In the following sections, we describe the main steps and the resulting segmentation method that we considered.

5.2.1 Feature extraction

While separating the different regions of the images based on pixel values leads to incorrect region assignments, a preferable approach is to consider local neighborhoods of each pixel, or transformations of the neighborhoods, as feature vectors to be used for segmentation.

One approach to extract features of regions with regular lattices is to take the absolute value of the Fourier transform of the local neighborhood of each pixel (Mevenkamp *et al.*, *IEEE Wint. Conf. Appl.*, 2016). This results in patches belonging to the same region having similar feature vectors, with the additional benefit of obtaining similar feature vectors that correspond to the boundary between regions (dependent on the size of the neighborhood). Similar ideas involve performing a similar operation at different scales (Staudacher *et al.*, *Physica A* 2005). Furthermore, one can use a data-driven approach to select the most appropriate basis to represent the image, such as the empirical wavelets used in K. Bui *et al.*, *Pattern Anal. Appl.*, 2020.

5.2.2 Segmentation

A standard approach to image segmentation involves the minimization of the Mumford-Shah model. In the binary case, where two regions are sought, this is a function over the boundary between regions and the two average values of the features in each region. Often the average values are considered to be known (they can be obtained, for example, by performing K-means clustering on the feature vectors).

A convex formulation of the piecewise constant binary Mumford-Shah model is described in Chambolle *et al.*, *Int. J. Comput. Vision.*, 2009, where the optimization problem results in a “soft segmentation” function whose superlevel set for a specific value gives the boundary between regions. This function is strongly convex and benefits from theoretical guarantees.

The convex formulation of the Mumford-Shah optimization problem can be solved with the primal-dual hybrid gradient (PDHG) algorithm, an algorithm that has been translated to an unrolled neural network in the context of tomography (Adler *et al.*, *IEEE T. Med. Imaging*, 2018). Following this approach, we implemented an unrolled network (discussed in [AI & Learning Theory](#)) where the two proximal operators are learned from the training data and the feature extraction process is performed on the fly by one of the trained networks.

The iterative approach based on the Mumford-Shah model can be generalized to multiple regions, and we expect that the unrolled approach will easily generalize as well, although this work is currently in progress.

5.2.3 Synthetic image generation

In order to train the unrolled neural network, an efficient way of generating training data is required. Obtaining the number of STM images required and then manually labeling each would be time consuming and partly defeat the purpose of the proposed work.

A better method is to generate synthetic training images that capture the main features of the data, namely the periodic patterns and the disordered regions at the boundaries between them, as well as the masks required by the training. Therefore, we have implemented a function that generates images with multiple regions with random orientations separated by a randomly generated boundary and the corresponding masks.

5.2.4 Workflow

By combining particular elements of the pipeline described in Sections [5.2.1-5.2.3](#), we obtain a promising approach to segmentation of SPM data. The first step after acquiring the data is to

generate a synthetic dataset that reflects the characteristics of the true images, such as regular lattices and disordered regions at the boundary between them. Then, feature extraction is performed using one of the methods described, followed by training the unrolled neural network to perform the segmentation. Once the network is trained, one can use it to produce the segmentation of (previously unseen) microscopy images. Lastly, the generated masks are used for further analysis.

5.3 Work done at IPAM

The workshops brought together mathematicians and scientists from other fields to tackle scientific interdisciplinary problems. This focus group consisted of people with varied academic backgrounds, which helped us view the problem through different lenses. We benefited from close ties with Paul S. Weiss's group in the chemistry department at UCLA, specifically with Kristopher K. Barr and Tianhan Liu, an adjunct professor and a postdoc respectively working in the lab. The group took advantage of the in-person modality of the long program and visited the lab. Moreover, talking to the lab members working with the STM on a daily basis gave us a better intuition about the hardware and the data acquisition process, which was essential for better understanding of the data.

To learn about the potential approaches to image segmentation, we invited speakers to give seminars on related topics: Jérôme Gilles, Philippe Sautet, and Philip Cho. Jérôme Gilles described how to use a combination of image decomposition, a variational approach, and empirical wavelets for image segmentation of STM images that helped us build our workflow. We used a similar pipeline to separate components of images and obtain a low memory, fast method to extract regions of different directionality. In addition, Workshop IV was directly relevant to this topic, as both talks and a number of posters touched on segmentation and SPM topics: *Supervised and Unsupervised approaches for Electron Microscopy Data Analysis*, Mary Scott, *Autonomous Hyper Scanning Tunneling Spectroscopy*, John C. Thomas, and *Computer Vision for Atomic Resolution Liquid Cell STEM*, William Thornley.

More specifically, we have followed the theoretical paradigm introduced in [Section 5.2](#) to implement an unrolled neural network approach to segmentation for specific types of images with regular patterns, namely SPM data. This approach is general and can be extended to other imaging modalities.

5.4 Outlook

A number of possible research directions have been discussed as part of this working group, which would enhance the study of physical and chemical phenomena of imaging modalities such as SPM.

Firstly, one potential direction is to increase the complexity of the synthetic data by simulating the noise more accurately. Moreover, the recent applications of diffusion models in deep learning are a promising direction for generating training data that reflects the imaging conditions.

Secondly, segmentation of data displaying regular patterns is a general problem. Therefore, extending the approach presented in this section to other types of features present in different imaging modalities would be a fruitful direction.

Thirdly, the approach to segmentation described above can easily extend to 3D datasets such as those obtained in materials science and valuable insights can be drawn from 3D spatial data that are not always accessible in 2D images.

Lastly, there are other potential issues that would benefit from computational approaches. For example, the tuning of the parameters of the PID feedback system ([Section 5.1](#)) can be done at the acquisition stage. This has been commercially realized in Bruker AFM systems and implemented in the ScanAsyst mode, which automatically adjusts all critical parameters. However, this has not been realized in most custom STM systems. In STM imaging, either constant current mode or constant height mode are applied for topography or tunneling current imaging, respectively. The implementation of automatic tuning PID values would improve the quality of the STM images.

6. Multimodal data processing and acquisition

6.1 Introduction

Multimodal microscopy that combines complementary nano- and atomic-scale imaging techniques is critical for extracting comprehensive chemical, structural, magnetic, and functional information, particularly for heterogeneous samples. Experiments from correlative electron, X-ray, optical, and scanning probe microscopes have generated very large data sets, and the scientific community desperately needs more efficient methods to deal with those. Applications of computational modeling in various microscopy techniques were demonstrated during this long program. Below we provide two specific examples on how to implement multimodal data

processing and acquisition using scanning transmission electron microscope (STEM) and atomic force microscopy (AFM).

6.2 Integration of STEM imaging, diffraction, and spectroscopy

A perfect testbed for exploring multimodal processing is provided by STEM. We consider three different types of data that can be gathered at the same time by the microscope. The first type encompasses the intensity measurements including the bright field (BF), annular dark field (ADF), and high-angle ADF (HAADF). The second type encompasses the hyperspectral data represented by energy dispersive X-ray spectroscopy (EDX) and electron energy loss spectroscopy (EELS). The third type encompasses the high-dimensional data consisting of diffraction patterns for each particular location represented by ptychography. The discussions in the focus group were based on the pairs of modes that can be processed simultaneously to obtain additional information. Some combinations of the above modes were also considered and reported during Workshop IV. During the Long Program the discussions circled around particular multimodal models and the general problems arising when fusing the models developed for two different modalities.

One example is the simultaneous processing of EDX and HAADF data. While EDX provides information about the element composition of the specimen along the path of the electron beam, this information is carried by very few X-ray counts and as a result the obtained relative concentrations are not very reliable. The Z-square effect observed for the much more reliable HAADF measurements can be used to evaluate and adjust the concentration information. It also can be a factor in reducing the effects from X-ray absorption affecting the concentration estimation from EDX. Additionally, the information from other modes like BF and EELS can increase the reliability of the estimates. Other multimodal settings of interest include combining the HAADF and ptychography data. One benefit from such a combination is the possibility of estimating the position noise from the HAADF data when taken several times from the same region and applying it to correct the positions of simultaneously gathered ptychography data.

Another emerging imaging model that could benefit from multimodal data processing is cryo-electron ptychography. For example, recent work from L. Zhou, *Nat Comms* 2020 proposes a phase contrast imaging ptychography technique in a setting similar to cryo-EM. This method shows a significant improvement for the reconstruction at low spatial frequency components and it has a great potential in further dose reduction. It is reported that only $5.7\text{e}/\text{\AA}$ electron dose was required to reveal high quality features of rotavirus particles. In comparison to standard phase contrast imaging in cryo-TEM, ptychographic reconstruction achieved a better SNR which can lead to the further improvement of electron tomography. At Workshop II, P. Wang also introduced a multi-modal adaption for ptychography, which can ideally achieve the

reconstruction of a 5D STEM image. Alongside the standard ptychographic quantitative phase recovery but with a “hollow detector”, EELS and HAADF analysis can be conducted simultaneously, which can also be further adapted for cryo-electron ptychography.

6.3 Multimodal approach with AFM data

In AFM, the microscope images the sample surface in separate modes. In each mode, we can collect distinct information about the sample surface. In the tapping (intermittent contact) mode, the oscillation amplitude can change depending on the distance between the tip and surface. Thus, by setting a fixed average amplitude, it is possible to collect the topographic information simultaneously with the phase difference (which depends on the mechanical properties and the topography of the sample) between the driver and the cantilever oscillations.

In the force spectroscopy mode, the force curve is recorded (force interaction between the tip and the sample as function as the tip-surface distance) for each pixel of the image. From the force curve, the mechanical properties (e.g. Young's/elastic modulus and the maximum adhesion force between the tip and the sample) of the sample for each pixel can be obtained. While the topographic and phase information are available at high resolution (e.g., 512×512 pixels) and allow fast collection, other information about the mechanical properties extracted from the force curve is collected slowly at low resolution (e.g., 64×64 pixels). Therefore, a challenge in AFM data processing is enhancing the low-resolution data to high-resolution in order to learn the properties of the sample in detail.

One option to achieve this would be by modeling the physics of the interaction between the probe and the sample which relate the physical properties of the material and the measurements collected by the microscope. By calibrating such a model to the collected data, it would be possible to expand the low-resolution information on a finer grid. However, describing all the interactions is an open problem. Alternatively, once a large enough dataset is collected, the latest approaches in machine learning could be leveraged to find the correlations of the different types of collected data within certain regions of the sample. One possible way is fitting the parameters of a neural network that describes how the physical properties of the samples relate depending on the position in the region of interest.

6.4 Work done at IPAM

The focus group had a number of discussions on topics related to combining different imaging modalities to improve the reliability and quality of the results. These focused around the activities in Workshop II initially and crystallized further during Workshop IV, which highlighted the application of deep learning and tensor methods to the processing of multimodal data. In

particular, we discussed several modalities where data from different sources can be fused, such as STEM and AFM as described in Sections [6.2](#) and [6.3](#).

Based on these discussions, a number of research opportunities have been identified which will result in fruitful collaborations between researchers from varied disciplines such as mathematics, materials science, biology, physics, and chemistry. Identifying specific mathematical and computational approaches, as well as the gathering of the appropriate datasets have been enabled by this long program, which are essential for kick-starting these collaborations in the future. In particular, the datasets have already been made available to the participants involved in this focus group.

Lastly, we have also considered creating a multimodal learning framework, which would allow one to combine individual neural network models. Neural networks are often used to process individual types of data. However, merging such models into an overarching model allows information to flow between the individual types of data, which has the potential to improve the reliability and precision of the results.

6.5 Outlook

When multimodal data is collected, the data processing of each modality usually is separate and the combined results are checked for consistency. One important future direction is to implement the simultaneous data processing, which has the advantage of requiring less data for extracting the same amount of information and could substantially reduce the beam damage to the specimen. To achieve this goal, we will collaborate with domain scientists to have consistent imaging modalities for each detector and stable mathematical learning procedures to fuse the data in reliable and reproducible ways.

Another important future direction is to efficiently acquire and reconstruct signals from few measurements under an assumption of model sparsity. Recently, it has been shown both empirically and theoretically that the sparsity assumption can be replaced by an appropriate generative prior. Such generative priors have been widely used for solving inverse problems like denoising, inpainting, and reconstruction from a small number of noisy measurements. This not only broadens the class of signals to which these techniques can be applied, but also results in fewer required measurements than the classical sparsity prior. Typically, these generative models stem from learned deep neural nets and offer numerous practical advantages.

Finally, distinct imaging modalities are used to analyze similar types of samples, each with its own data processing pipeline. For example, cryo-EM, cryo-ptychography, and cryo-4D-STEM are all able to process biological samples at similar length scales. Therefore, the multimodal framework can be extended to processing the same set of samples with all these modalities. Such

an endeavor would require the development of a novel paradigm that is able to combine these types of data and process them in parallel, resulting in improved results with high confidence.

7. Space-time models

7.1. Introduction

Tomography has had a radical impact on diverse fields ranging from the 3D determination of the individual atoms in matter to the diagnosis of disease in medicine. In the physical sciences, atomic electron tomography (AET) has been developed to resolve the atomic structure of crystal defects and amorphous materials in unprecedented 3D detail. In the biological sciences, cryo-electron microscopy has become a powerful tool for 3D structural determination of macromolecules with identical or similar conformations at near-atomic resolution. In medicine, computed tomography has been routinely used as a diagnostic imaging procedure. Although the applications of tomography are wide and diverse, there is a central problem associated with its mathematical implementation : accurately reconstructing a 3D structure from noisy and incomplete projection data. Over the years, several advanced tomography algorithms have been developed to address this challenge. However, an even more challenging and more impactful problem would be to reconstruct the 3D structure and dynamics (i.e., 3D space and time) of an object from a series of tomographic tilt series. Formulating an inverse problem to account for both the spatial and temporal changes is a major computational challenge: not only does the number of optimization variables increase, but the ratio between measurements and unknowns shrinks dramatically.

This group has focused on space-time models, which can be used to regularize inverse problems with spatial and temporal dimensions. Spatial regularizers such as sparsity and nonnegativity are commonly used in 3D reconstruction. Solving 4D inverse problems requires integrating these regularizers with constraints on the changes of the 3D structure through time. To date, many space-time regularization techniques have been proposed in various computational imaging systems. Algorithms for compressive sensing of video, for example, have employed sparse regularization in multiscale spatial-temporal transform domains, motion estimation and compensation, optical flow, dynamical system models, and deep neural networks. Similar techniques have been employed for dynamic MRI. Variational methods for solving dynamic inverse problems have also been proposed that allow for nonparametric estimation of time-varying distortions; see also [Geometry in Microscopy](#).

Further work is needed to push the frontiers of dynamic imaging. We have focused on one problem in this domain, which is 4D atomic electron tomography. This problem was suggested

by John Miao in a tutorial talk at IPAM and is also closely related to Workshop II on Mathematical Advances for Multi-Dimensional Microscopy.

7.2. 4D atomic electron tomography (AET)

In 1959, Richard Feynman challenged the electron microscopy community to build a microscope that could identify all of the atoms in a chemical structure. As Feynman's challenge has now been addressed in several systems, a bigger challenge is to develop multi-dimensional (D) electron microscopy (with $D > 3$) to locate and identify all atomic species in complex systems, with additional dimensions such as time, energy, temperature, and external bias. Such a powerful tool would have a transformative impact in physics, chemistry, materials science, nanoscience, and other fields. Addressing this major challenge requires the advancement and integration of state-of-the-art electron microscopy methods such as AET, ptychography, and 4D scanning transmission electron microscopy as well as powerful computational algorithms and mathematical modeling. In this subtopic, we will develop 4D AET with advanced computational algorithms to capture atomic motion in 3D space and in time. We will take advantage of the fact that only a small percentage of atoms in a sample are dynamic, while the majority of the atoms are static. We will also use the atomicity constraint, that is, we are only interested in the atomic coordinates and species, but not the atomic shapes. Based on these facts, we will first divide a tomography tilt series into a number of temporal sequences and then develop computational algorithms to reconstruct the individual temporal sequences. One challenge associated with the reconstruction of each temporal sequence is the increase of the missing wedge, that is, when a sample is tilted around a small angular range, there is significant missing data in Fourier space. One approach to reduce the missing wedge effects is to incorporate neural networks into iterative algorithms; this technique was discussed by Yongsoo Yang in a talk at Workshop II, and it has been demonstrated to be successful in both simulations and experiments. Another idea we will pursue is to combine advanced tomographic algorithms with the Riemannian manifold method to achieve 4D AET (see [Section 8](#) in detail).

To test our computational algorithms, we will create several atomic models, of which a small fraction of atoms move in three dimensions as a function of time. We will compute projections from each model at various angles and add Poisson noise to each projection. We will then use the models to test and optimize different algorithms. The optimized algorithm will be used for experimental data, which will be performed at the National Center for Electron Microscopy, Lawrence Berkeley National Lab. For the initial experiment, we will acquire tomography tilt series from FePt nanoparticles at varying temperatures, where Pt atoms slowly diffuse from the center core to the shell. The goal is to capture the structure and dynamics of the atomic diffusion at the single-atom level. The success of this experiment will not only enable us to observe atomic

diffusion for the first time, but also paves a way to imaging a wide range of materials at atomic resolution in four dimensions.

7.3. Work done at IPAM

We have created a testbed for simulating the 4D AET problem and prototyped an initial 4D reconstruction algorithm. Our testbed allows for specification of atom positions and motions and generates a tomographic projection at each time step. As a baseline, we computed the 3D reconstruction by running projected gradient descent (PGD) to minimize the residual error in the tomographic measurements while ensuring nonnegativity and/or sparsity of the 3D volume. In this baseline, we noticed two types of artifacts: one blurring effect due to specimen motion and another due to the missing wedge.

For 4D reconstruction we explored the use of a small number of “anchor” volumes to reduce the complexity of the inverse problem. Each anchor volume is a 3D volume representing the specimen at a given time, but the number of such anchor volumes can be fewer than the number of tomographic projections. By assuming that the specimen at intermediate times can be interpolated from the anchor volumes, all tomographic projections can be expressed as a linear function of only the anchor volumes. To reconstruct the anchor volumes, we have again employed PGD and included regularization of both the spatial and temporal gradients of the 4D concatenation of the 3D anchor volumes. Using a full set of anchor volumes yields a highly under-determined system, and the 3D reconstruction baseline must be used as an initial input to avoid getting trapped in a local minimum. Exploiting the interconnection between all anchor volumes, we propose Gaussian weights to enforce the smooth temporal transformation among these anchor volumes. The smooth temporal interpolation has proved successful so far. The 4D reconstruction matches better to the ground truth compared to the static 3D reconstruction while the blurring artifacts and the missing wedge effects are reduced. Further investigation and additional constraints will be used to improve the 4D AET reconstruction while reducing the blurring and missing wedge artifacts.

7.4. Outlook

While our initial experiments are promising, many possibilities remain for improving the 4D AET algorithm. Our goal is to weigh the possible benefits of incorporating multiscale transforms, manifold-based deformation models, optical flow, parametric atomic models, and deep neural networks for solving this inverse problem. Specifically, implicit neural networks or coordinate-based neural networks are an emerging method to capture multi-dimensional data compressively. In the context of space-time models, a dynamic scene can be represented by such a neural network. To create such a neural space-time model, a dynamic scene can be split into

two parts: motion kernels corresponding to different timepoints can be represented by a motion multi-layer perceptron (MLP) and a time-independent scene can be represented by a scene MLP. Both MLPs are coordinate-based, meaning that they take the volume's spatial coordinates (x, y, z) and timestamp t as the input to retrieve its corresponding voxel value. Neural space-time models have shown great promise in dynamic imaging and should be evaluated for feasibility in 4D AET.

Parametric modeling of the atomic centers, intensities, and motion is another option worth considering for the 4D AET problem. One great benefit of the parametric model is that it eliminates all above artifacts in addition to identifying the atom types and positions. The challenge to this technique is that it requires a good initialization, such as the number of atoms and their approximated positions. With the help of the baseline 3D reconstruction and other techniques, we will aim to provide a more robust and faithful 4D reconstruction of a wide range of amorphous materials that are relevant to modern science and technology.

If the expected atomic motions are smooth on the timescale of the acquisition, another suitable strategy may be virtual time synchronization (Bradley et al., CVPR Workshops, 2009). The technique requires at least two time frames to track the motion of the image content, assuming a smooth displacement in between frames. In the 4D AET context, this could be realized by angularly scanning the sample at least twice. In this way, different time instances with the same viewing angle become available. Tracking the motion in a constant angular view and interpolating the images of all viewpoints to a common virtual time, a sharp 3D volume may be reconstructed with minimized effects of motion blur.

Beyond 4D AET, novel hardware may also spark new possibilities for space-time models in the future. One example is the imaging sensor. Most of the current imaging algorithms take frame-based input images captured by complementary metal-oxide semiconductor (CMOS) or charge-coupled device (CCD) camera sensors, while event-based neuromorphic camera sensors are gaining traction in recent years. Unlike frame-based cameras that capture a 2D matrix of light intensity after each exposure, event-based cameras receive a stream of events continuously in response to the intensity changes. Event-based technology can observe very rapid motion beyond the temporal resolution of conventional frame-based cameras, which may enable new findings. Meanwhile, the continuous nature of event data poses challenges for existing discretized methods; new engineering tools, models, denoising algorithms, and compression techniques for continuous data streams are needed to unleash the full potential of event cameras.

8. Geometry in data processing for microscopy

8.1 Introduction

In microscopy, there are usually very clear underlying “objects” that one actually cares about, e.g., the protein structure/motion in Cryo-EM or the atomic and meso-scale structure of nano-particles/amorphous materials in AET. In several cases, these structures can be modeled as points living in a smooth manifold. Given the smooth manifold, there is the freedom to choose a smooth metric tensor field and define a Riemannian manifold. Data processing on the Riemannian manifold directly has several potential upshots. One clear advantage is that the obtained reconstructions will always be elements in the set of interest, which can be a great help in the design of algorithms. Secondly, well-designed geometry allows for more meaningful interpretation of data, e.g., when clustering/interpolation or when computing summarizing statistics such as the “mean” of a data set. However, data processing on Riemannian manifolds poses several new challenges that need to be addressed. In addition, direct geometric constraints may be exploited advantageously in the processing of microscopic data, e.g. for pose estimation.

8.2 Geometry in microscopy

As mentioned above, there is often an underlying object of interest being imaged indirectly through some microscopy technique. One approach to acknowledging this object is to construct a smooth manifold and endow it with a Riemannian structure that tells “what other objects are close by”.

Outside of microscopy, this approach has been – and still is – very popular in the setting of shape spaces. Using the geometry of shapes is a widely applied approach in several communities in computer vision, e.g., in object matching and in inverse problems, especially spatial-temporal inverse problems. It is important to note that even with the arising non-convexity of the resulting optimization problem (due to the underlying space), the non-linear structure often solves more problems than it causes in terms of performance. However, there tends to be computational challenges due to the infinite-dimensional nature of the shape spaces. Both in the computer vision setting and in the spatio-temporal inverse problem setting unsupervised deep learning is a popular approach to attain better performance, as discussed by Daniel Cremers at Workshop IV.

In microscopy, the underlying structures are often finite-dimensional in nature, e.g., a protein consisting of a fixed and finite number of moving parts or a nano-particle with a fixed number of atoms. Finite dimensionality not only also allows for a more custom approach to the microscopy

task, but also makes computations easier/feasible while maintaining mathematical rigor and giving cheap access to important mappings such as the logarithmic and the exponential mapping. Approaches within this spirit are rare, but the few existing approaches yield promising results.

8.3 Challenges in data processing for manifold-valued data

Besides developing the space of interest, there is an ongoing effort to generalize popular data processing techniques from Euclidean space to manifolds. In several cases there is a “natural” extension. However, arguably there are even more instances where there are several candidate generalizations that are vastly different on a non-linear space, but reduce to the same notion in the Euclidean setting. Picking the right generalization is therefore important, but has not always been historically addressed well. One topic that has drawn the attention and effort of several communities is solving these generalized models, which often come down to solving some potentially non-convex optimization problem on the Riemannian manifold.

8.3.1 Work done at IPAM

In the spirit of the paragraph above, we considered current notions of low-rank approximation for manifold-valued data. In particular, we revisited the notion of principal component analysis for manifold-valued data and showed that the current notion is a poor generalization from the Euclidean case. That is, this (popular) notion has provably serious issues due to the curvature of the space, which vanishes in the Euclidean case. We propose a general framework for overcoming these issues in decomposition methods/low-rank approximation through including curvature information in a generic way, i.e., one that is easily adaptable for other decomposition methods. Our initial results seem promising.

8.4 Tracking tomography

A prominent issue in tomography is the missing wedge problem, where less than a full 180 degree coverage of projection directions adversely affects the reconstructions and introduces artifacts. Recently, one of the core participants has been developing an experimental technique to address this problem for optical imaging of biological specimens in a microfluidic chamber. The method is based on applying acoustic force fields to rotate the sample approximately about its center of mass. This way, a full revolution of projections can be acquired, for some samples even around two orthogonal axes, circumventing the missing wedge problem.

The challenging aspect of this technique is that the rotation axis and speed depend on the geometry of the sample and therefore can not be predicted a priori nor be controlled to the required precision for a calibration-agnostic tomography scheme to succeed. It is therefore

necessary to estimate the orientation of the sample, its rotation axis and angle for every frame of the acquired video sequence.

8.4.1 Work done at IPAM

The problem description is similar to a technique termed structure-from-motion (SfM) from the 3D computer vision (CV) literature. There, pose estimation of a moving camera located in a static scene has been studied for several decades. The underlying formulations exploit feature tracking and algebraic constraints between different camera poses that are present in projective geometry. In contrast to the CV problem, in microscopy, we are dealing with transparent samples and the assumptions of trackable surface features, as used in the CV context, are not strictly valid. In our initial work at IPAM, we have obtained initial results showing the promise of adapting these techniques to the pose estimation problem in microscopy. In future, the robustness of the techniques will be improved by explicitly modeling the image formation and the associated mechanical constraints of the setting.

8.5 Outlook

Besides continuing the work described in the above, another opportunity in the spirit of [Geometry in Microscopy](#) would lie in designing geometry for spatial-temporal reconstruction of moving particles as outlined by the [Space-Time Models](#) group. Furthermore, we also see great potential of Geometry in Data Processing for Microscopy in other applications that were discussed in several long program workshops. In particular, there could be applications – with genuine high impact – in discovering hidden meso-scale structure in amorphous materials (John Miao at IPAM Workshop I) and efficient computation of phase transitions in the absence of data in-between states (Naomi Ginsberg at IPAM Workshop IV).

From a broader and more ambitious point of view, developing geometry and data processing tools for problems in microscopy seems a promising line of work for developing new mathematics and new tools for general data-related problems.