Chapter 1

Optics-Free Imaging with DNA Microscopy: An Overview

Xin Song\textsuperscript{1,2,3} and John Reif\textsuperscript{1,3}

\textsuperscript{1}Department of Electrical and Computer Engineering
\textsuperscript{2}Department of Biomedical Engineering
\textsuperscript{3}Department of Computer Science

Duke University, Durham, NC 27708, USA

Introduction

The specificity and programmability of Watson-Crick base pairings enable scientists to design networks of synthetic DNA oligonucleotides (oligos) that react in predictable ways to realize systems of DNA-based computation,\textsuperscript{1} data storage,\textsuperscript{2} and molecular imaging.\textsuperscript{3} The ability to resolve details at small scales is crucial to the development of many fields such as biological sciences and materials engineering. Conventional microscopy relies on photons (e.g., in optical microscopy), electrons (e.g., in electron microscopy), or scanning probes (e.g., in atomic force microscopy) to directly interrogate the spatial arrangement of surface/volumetric features of a given sample. However, these techniques have various shortcomings such as the diffraction-limited resolution in optical imaging, expensive instrumentation in electron beam imaging, and low throughput in atomic force imaging. Over the past few decades, DNA nanotechnology has contributed a wealth of techniques and tools including super-resolution
Alternative Computing

imaging, which can leverage the stochastic binding kinetics of short dye-labeled DNA probes to visualize fine features on molecular-scale targets such as DNA origami, improving the imaging multiplexity and localization precision. Recently, the concept of optics-free imaging\(^5\) has offered an alternative route to use DNA as a molecular-scale imaging medium without a priori spatial addressing and to leverage next-generation sequencing for high-throughput data readout. Such a new modality of DNA-based microscopy completely obviates the use of optics for molecular image acquisition and reconstruction and may offer a scalable and economical solution to address prior limitations faced by conventional microscopy.

This chapter provides a short survey of recent achievement in both theoretical and experimental implementations of DNA-based optics-free imaging, which we combinedly refer to as DNA microscopy. As shown in Figure 1, the typical pipeline of DNA microscopy involves a few steps including (i) DNA barcoding of molecular targets, (ii) pairwise concatenation of adjacent barcodes, (iii) sequencing of concatenated barcodes, and (iv) image reconstruction based on barcodes proximity information. Unlike conventional imaging systems that generate direct visualizations of a sample in the microscope’s field of view, DNA microscopy investigates the local proximities between individual targets of interest to map out the relative positions among all the targets (e.g., a population of molecules). This set of pairwise adjacencies can be algorithmically analyzed to reconstruct a global image that approximates the original positions of the molecular targets on a 2D surface or in a 3D volume without any a priori knowledge of their spatial localities.

![Figure 1. Typical pipeline of optics-free DNA microscopy. (a) Original spatial arrangement of molecular targets. (b) Tagging the targets of interest with unique DNA barcodes. (c) Pairwise concatenation of adjacent barcodes. (d) Recovery of concatemers by DNA amplification and sequencing. (e) Reconstruction of a spatial image based on pairwise proximities. The estimated image may not preserve the original scale, rotation, and chirality of sample without additional information.](image-url)
DNA Microscopy for Surface 2D Imaging

For DNA microscopy on flat surfaces (including applications that investigate the spatial arrangement of 3D objects through their projections onto a 2D surface), the process of image reconstruction is analogous to forming a 2D graph, where each node corresponds to an individual target on the sample and each edge connecting two nodes reflects the juxtaposition between a pair of neighboring targets. In other words, the nodes of the graph can be viewed as the pixels of an underlying sample image. The first step to form such a graph is to confer each node with a unique identity. Typically, DNA microscopy achieves this by tagging each target (e.g., molecule) on the sample with a unique DNA barcode (i.e., short random sequence of nucleotides). Next, barcodes of adjacent targets undergo enzymatic reactions to record the pairwise proximity between each two neighboring nodes. In large and complex graphs, each node may be in close proximity to multiple other nodes. Therefore, practical DNA microscopy based on the recovery of local proximities needs to account for this potential “one-to-many” relationship between any given node and its immediately adjacent neighbors. Once an adequate set of pairwise proximities are recovered (typically by DNA amplification and sequencing), edges of the graph can be reconstructed and collectively determine the relative positions of nodes to form a graph that approximates the original spatial distribution of barcoded targets on surface. Because only local proximity information is used to reconstruct the graph, the resulting global image may not well preserve the original scale, rotation, and chirality of the actual sample image without additional information. However, the use of DNA barcodes and massively parallel sequencing gives DNA microscopy a unique advantage in terms of achieving remarkably high multiplexity and throughput for investigating complex and large population of molecular targets.

Experimentally, the proximity record between two adjacent DNA barcodes may be generated via biochemical colocalization techniques such as DNA proximity ligation\(^6\) (facilitated by a connector oligo that binds to two adjacent probes and joins them to form a template for ligation and subsequent PCR amplification) or DNA proximity extension\(^7\) (facilitated by hybridization between two adjacent probes with matching annealing sites and then followed by polymerase extension of the hybridizing oligos to form a template for PCR amplification). However, the barcoded probes
used in these techniques usually become depleted after recording single proximity pairs and thus do not work well in scenarios where a given node may have more than one neighbor in immediate proximity. To address this challenge, Schaus et al. proposed an “auto-cycling proximity recording” mechanism (Figure 2a) that enables continuous and repeated generations of proximity records between any adjacent pair of DNA-barcoded molecular targets. In their design, each target is tagged with a DNA hairpin probe that has a unique barcode and an open toehold (i.e., single-stranded overhang). DNA hairpin is a secondary structure of single-stranded DNA when it folds on itself and forms an unpaired loop with a duplex stem region. In this case, the toehold forms a sticky end that can hybridize to a primer. With strand-displacing polymerase and the energy from dNTPs, the double-stranded hairpin stem can be primed partially open until the polymerase encounters a stopper before the hairpin loop. The hairpin sequence is carefully designed to bias toward the reformation of a closed hairpin – a strategy that facilitates the newly synthesized strand (referred to as a “half-record”) to be isothermally displaced by DNA strand displacement (a hybridization-based process in which an incumbent strand is displaced by an invading strand via branch migration) which spontaneously reforms the hairpin stem. This single-stranded half-record contains the identity of the hairpin and exposes a palindromic domain that can hybridize to the palindromic domain of a nearby half-record produced by an adjacent probe. The hybridized palindromic domains are then extended by polymerase, resulting in the release of a double-stranded “full-record” that incorporates both barcodes from the two adjacent probes. The release of the “full-record” in turn regenerates the involved pair of probes to their initial states so that both of them can continue the cycles of transient binding with any adjacent probes. Such an isothermal “copy and release” mechanism enables a probe to produce multiple proximity records with any immediately adjacent probes in a nondestructive, autonomous, and continuous manner. The authors compared different probe lengths and attachment chemistries on DNA origami to characterize the generation rate of proximity records, which was found to be nonlinearly related to the probe-probe distance. In addition to experimentally interrogating complex nanoscale geometries such as a seven-probe geometry on DNA origami, the authors demonstrated their nanoscope’s reusability by repeatedly sampling the same probe undergoing different state changes. This work reports that each probe can generate ~30 records on average, offering high sensitivity for
potentially recording a complete set of local proximities needed for intact global image reconstruction.¹

Figure 2. Mechanisms for generating multiple proximity records between adjacent barcoded molecules. (a) Auto-cycling proximity recording. (Adapted from ref. 8 with permission). (b) Iterative reversible proximity ligation. (Adapted from ref. 9 with permission).

To allow the generation of more than one proximity record per probe, Boulgakov et al.⁹ extended the conventional technique of proximity ligation to implement iterative cycles of ligation, cleavage, and re-ligation between adjacent pair of DNA probes (Figure 2b). Each probe contains a primer site, a unique barcode sequence, and a half restriction site. With the presence of a short bridging oligo complementary to the restriction half-sites, two adjacent probes can be ligated to form a concatenated record, which is subsequently duplicated by primer extension and read out by real-time polymerase chain reaction (qPCR). The sequences of probes are designed such that a restriction cut can be made at the ligation site to revert the probes to their original un-ligated state. This enables each probe to participate in further cycles of reversible ligations to record its adjacency to different probes located in close proximity. The overall technique is conceptually simple however it requires that two neighboring DNA probes must have opposite polarities in order to form a pairwise concatenated record. Such a criterion may limit the achievable robustness and resolution of image acquisition and reconstruction. As a proof of concept, the authors tested the feasibility of reversible ligations using probes attached to magnetic beads. Although the technique was not yet experimentally demonstrated to interrogate complex spatial colocalizations involving multiple targets, this work proposed the application of graph theory to
interpret the outputs from iterative proximity ligations and subsequently recover complex 2D geometries using spring layout algorithms. According to simulations, the algorithms were able to reconstruct various 2D topologies with relatively small errors despite just a few rounds of iterative ligations and low ligation efficiencies.\(^9\)

Recently, Hoffecker et al.\(^{10}\) developed a computational framework that also explores the use of graph theory to facilitate optics-free image reconstruction for DNA microscopy. Their idea is to convert the surface of interest into a tessellation of barcoded patches and then form an untethered graph by probing the pairwise adjacencies between patches with shared borders. The tessellated surface offers several characteristic properties that can be leveraged mathematically to reconstruct planar embeddings of the untethered graph. Different algorithms may be used to compute a proper embedding that approximates the original Euclidean spatial information of the underlying image. In order to form such a tessellated surface, the authors propose to randomly seed barcoded DNA probes on a surface full of primers, and then each seeded probe is locally amplified to form a colony of probes bearing the same barcode from the seed probe. As the colonies expand and saturate (i.e., when adjacent colonies share boundaries), a Voronoi tessellation is formed with each patch uniquely barcoded. After that, the pairwise proximities between immediately adjacent colonies can be biochemically recorded and read out by sequencing to facilitate subsequent graph recovery and image reconstruction. According to simulations, higher colony density (i.e., average number of colonies per unit area) reduces the average distortion and improves the resolution of image reconstruction, whereas variations in site density (i.e., average number of duplicated probes within a colony) has a minimal effect on distortion. It is worth noting that each colony ultimately represents a single pixel on the estimated image, as a result, self-pairing events within individual colonies does not provide additional information for graph reconstruction and may be avoided via techniques such as bipartite networks or series colony generation.\(^{10}\)

**DNA Microscopy for Volumetric 3D Imaging**

As image reconstruction in DNA microscopy depends primarily on investigating local proximities, the technique can be applied to reconstruct a sample’s spatial information post hoc even if the sample’s original
spatial arrangement is scrambled or lost. Glaser et al.\textsuperscript{11} views the basic principle behind DNA microscopy as analogous to putting together a large puzzle based on only the relative positions between adjacent puzzle pieces (Figure 3a). Specifically, the set of relative positions can be collectively represented by an $N \times N$ similarity matrix for a puzzle containing $N$ pieces, where each entry in the matrix describes the proximity between two given puzzle pieces. Then the goal of image reconstruction can be viewed as mapping such a high-dimensional matrix to a lower-dimensional image that reflects the puzzle pieces’ original spatial coordinates (typically in 2D or 3D). To facilitate high-resolution image reconstruction for large-scale “puzzle imaging”, the authors developed two dimensionality reduction algorithms – sparse diffusion maps and unweighted landmark isomap – and evaluated the algorithms’ performance in three hypothetical applications, including (i) neural voxel puzzling, (ii) neural connectomics puzzling, and (iii) chemical puzzling. The first application explores 3D mapping of neurons in a tissue sample by labeling each neuron with a unique barcode and then shattering the tissue sample into numerous tiny voxels. Each voxel may contain several barcoded neurons whose identities can be inferred by sequencing. Based on the fact that adjacent voxels are likely to share more common neurons, a similarity matrix can be constructed and analyzed to recover the relative coordinates of each voxel’s placement in the original 3D space. According to simulated data, both algorithms reconstructed the image faithfully with impressive robustness against missing voxels. In the second application, information on neural connections is used to recover the spatial locations of individual neurons within a 3D specimen. First, neurons forming synaptic connections link their DNA barcodes to form pairwise proximity records. Next, the algorithm takes these proximity records as input to generate a connectivity matrix that can be treated as the similarity matrix to puzzle back the individual neurons into their respective positions in the original 3D space. For this particular application, it is worth noting that neurons may form both short-range and long-range connections, therefore, algorithms such as the unweighted landmark isomap that expects only short-range connections may lead to incorrect image reconstructions. In the third application, puzzle imaging was used to map out complex gradients of chemical concentrations by recovering the locations of chemical-sensing bacterial cells growing on a surface. Recovery of spatial localities relies on the recording of pairwise adjacencies between neighboring bacterial colonies via conjugative transfer of plasmids
between cells. Overall, this work provides thorough evaluations of the proposed algorithms and offers important insights on the computational complexities and image reconstruction accuracies of DNA microscopy in different use cases, which may be generalizable to various other applications.

Weinstein and coauthors\(^\text{12}\) recently developed an experimental platform for spatio-genetic DNA microscopy based on amplicon diffusion dynamics (Figure 3b). In their system, transcript molecules within a biological specimen are tagged with unique DNA barcodes, and then each barcoded molecule is amplified in situ to form a molecular diffusion cloud. As the PCR products slowly diffuse across the cell, adjacent diffusion...
clouds start to encounter and overlap at various degrees. During this process, amplicons located within the overlapping clouds can concatenate with one another via overlap-extension PCR reactions. The pairwise concatenation reaction is designed to not allow self-reactions among amplicons that have identical barcodes. As a result, each pairwise concatemer records the identities of two adjoining amplicons, the biological sequences of the two associated transcripts, as well as a unique identifier generated for the single concatenation event. Because each concatemer is uniquely labeled by an event identifier, one could count the total number of concatenation events that have occurred between two overlapping clouds. From a probabilistic point of view, this number gives an estimate of the degree of the cloud overlap and in turn reflects the original spatial proximity between the corresponding barcoded transcripts from which the clouds were formed initially. By inferring the counts of concatenations between different pairs of diffusion clouds in a matrix form, one could apply matrix algebra and spectral graph theory to decode the information into a high-resolution image that visualizes the original dimensionality and the spatial arrangement of target molecules in the specimen. Using the protocol, the authors experimentally demonstrated large-scale optics-free imaging of transcripts in mixed cell populations and preserved remarkable cellular resolution without systematic distortions.

**Conclusions and Outlook**

DNA microscopy enables an exciting new form of molecular imaging that completely sidesteps the use of optics from initial image acquisition to final image reconstruction and visualization. Taking advantage of DNA barcoding and next-generation sequencing, recent works have demonstrated the technology’s remarkable potential for achieving high resolution, multiplexity, and scalability without specialized equipment as required by conventional imaging platforms. As a nascent technology, DNA microscopy still faces various limits and challenges before it can be standardized and widely adopted. In particular, the reliance on local proximities to estimate global patterns makes it difficult to reliably recover images that have empty spaces or disjoint features. To address challenges like this, improvements may be needed for both the biochemistries (e.g., enhancing the efficiency and yield of proximity recording) and the image reconstruction algorithms (e.g., reducing the computation complexity and increasing the robustness against noisy and missing data). With continuing
development and refinement, DNA microscopy may become a leading implementation for optics-free molecular imaging that is both robust and cost-effective.

Acknowledgments

This work was sponsored by NSF grant no. CCF 1813805 and CCF 1909848.

References


(10) Hoffecker, I. T.; Yang, Y.; Bernardino, G.; Orponen, P.; Högberg, B. A
