A Computational Framework for DNA Sequencing Microscopy

Ian T. Hoffecker[a], **Yunshi Yang**[a], **Giulio Bernardinelli**[a], **Pekka Orponen**[b], **and Björn Högberg**[a,2]

[a]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden; [b]Department of Computer Science, Aalto University, FI-00076 Aalto, Finland

This manuscript was compiled on August 16, 2019

We describe a method whereby micro-scale spatial information such as the relative positions of biomolecules on a surface can be transferred to a sequence-based format and reconstructed into images without conventional optics. Barcoded DNA polony amplification techniques enable one to distinguish specific locations of a surface by their sequence. Image formation is based on pairwise fusion of uniquely tagged and spatially adjacent polonies. The network of polonies connected by shared borders forms a graph whose topology can be reconstructed from pairs of barcodes fused during a polony crosslinking phase, the sequences of which are determined by recovery from the surface and next-gen sequencing. We developed a mathematical and computational framework for this principle called Polony Adjacency Reconstruction for Spatial Inference and Topology and show that Euclidean spatial data may be stored and transmitted in the form of graph topology. Images are formed by transferring molecular information from a surface of interest, which we demonstrated *in silico* by reconstructing images formed from stochastic transfer of hypothetical molecular markers. The theory developed here could serve as a basis for an automated, multiplexable, and potentially super resolution imaging method based purely on molecular information.

next gen sequencing | DNA microscopy | polonies | DNA computing | graph theory |

Microscopic imaging has traditionally relied on optics to amplify signals derived from initially confined spatial regions. Exceptions include atomic force microscopy which images by utilizing a probe to interact with the sample. DNA has a high information density, with storage levels of 5.5 petabits per cubic millimeter achieved (1), making it an attractive medium for encoding spatial information at microscales. In this paper, we present a theoretical foundation for a spatial information encoding approach that utilizes DNA sequencing and graph theory that could be used to generate whole images.

DNA-driven reactions can be coupled to optically-acquired spatial information such as with proximity ligation assay (PLA) (2), and DNA-PAINT (3) where molecular interactions mediated by DNA are discovered using fluorescence. There is also a family of techniques for connecting spatial locations with single cell RNA sequencing data: using *a priori* knowledge of spatial marker genes to associate unknown genes to approximate locations, the *a priori* data being in most cases obtained by microscopy such as with *in situ* hybridization or modelling of spatial expression patterns to retrieve locations of associated genes (4–9). Alternatively, direct microscopy-based *in situ* sequencing methods achieve precise context-sensitive spatial transcriptomic information without needing to scramble spatial data by dissociation prior to sequencing (10, 11).

Encoding spatial information in a way that is preserved in the scrambling during isolation and recovery from *in situ* contexts that can then be read and recovered with sequencing is a major challenge. A few techniques achieve this by encoding spatial information directly into a molecular format, e.g. in the form of DNA read during sequencing along with transcriptomic data. These methods are based on artificial generation of an addressable surface using printing or lithography (12–14).

Herein, we describe a computational framework for a method called Polony Adjacency Reconstruction for Spatial Inference and Topology (PARSIFT), for the purpose of encoding images, for example of the positions of specific molecules relative to others on a 2D plane, directly into a DNA-based format without transduction of information through any other medium without *a priori* surface addressing. PARSIFT utilizes the connectivity of vertices in a graph of paired DNA sequences to infer Euclidean spatial adjacency and next-gen sequencing to recover that information *a posteriori.*

Encoding of topological data in DNA sequence format is possible by using DNA barcodes (unique molecular identifiers), i.e. randomized stretches of bases within a sequence of synthetic DNA. Barcodes associated with spatial patches can establish an identity for those locations, each patch distinguishable from another by sequence. A DNA barcode with 10 bases has over a million possible sequences, and larger barcodes can be used to create effectively unique labels in a system. The basic unit of topological data is an edge or association between two adjacent patches by physically linking between their barcodes. Topological mapping with barcoding has been used to infer neural connectomes by building a network from cells sharing common barcodes left by cell-traversing viruses (15) as well as features of DNA origami (16).

We can barcode surface patches using polony generation methods like bridge amplification (17), a 2-primer rolling

Significance Statement

Traditional microscopy is based on the propagation of interactions between light and small scale objects up to larger scales. Such information may be encoded in DNA and transmitted with next-gen sequencing to be later reconstructed and visualized computationally. We provide a mathematical framework and computational proof of concept for a form of DNA-sequencing based microscopy that may be used to construct whole images without the use of optics. Such an approach can be automated in a parallel and multiplexable way that current optical and scanning-based techniques are unable to achieve.

ITH, GB, and BH conceived project. ITH and YY implemented the in silico proof of concept. ITH, YY, PO, and BH developed the mathematical theory. ITH, YY, GB, PO, and BH wrote the manuscript.

The authors declare no conflict of interest.

[1] To whom correspondence should be addressed. E-mail: bjorn.hogberg ki.se

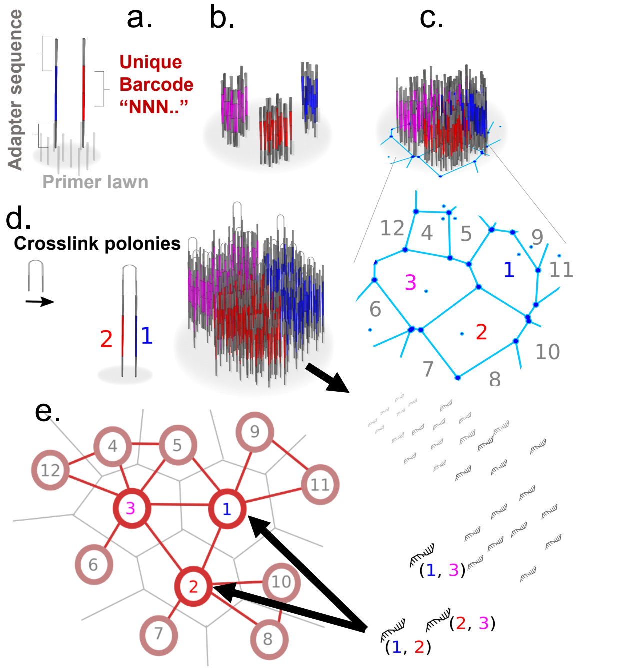

Fig. 1. Encoding and recovering metrics through polony adjacency. (a) Seed molecules with unique barcode sequences land randomly on a surface of primers. (b) Local amplification of seed molecules produces sequence-distinct polonies. (c) Saturation of polonies occurs when polonies are blocked from further growth by encountering adjacent polonies, forming a tessellated surface. (d) Random crosslinking of adjacent strands leads to pairwise association of nearby barcodes. (e) Recovery and sequencing of barcode pairs enables reconstruction of a network with similar relative positions of polonies as the original surface.

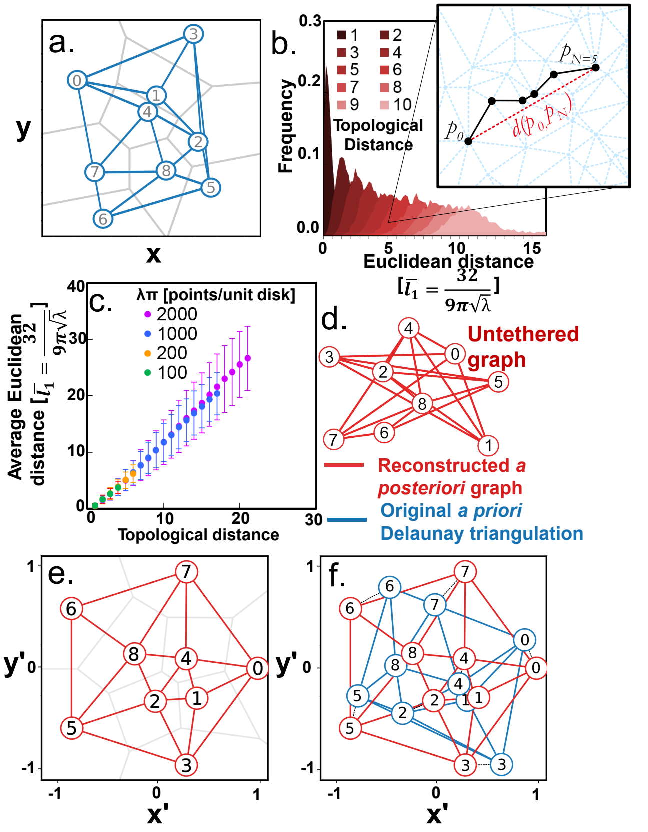

Fig. 2. Encoding and recovering metrics via topology. (a) Nine seed molecule points distributed randomly on a plane, the induced Voronoi tessellation T (gray lines), its Delaunay diagram D (blue lines), and the untethered graph G. (b) The distribution of Euclidean distances associated with a given topological distance (path with the fewest edges between two points) sampled for random Poisson Delaunay triangulations (5000 samples per topological distance value). (c) Euclidean distances normalized to the average length of a typical Poisson Delaunay edge (Equation 9.9 (21)) plotted versus topological distance for different Poisson intensities, exhibiting linearity between topological and Euclidean distance. (d) The untethered graph: a set of nodes (black) and edges (red) that constitutes the information preserved after dissociation from spatial context. (e) Reconstructed planar embedding of the initially untethered graph (red lines) using the Tutte embedding approach and corresponding Voronoi tessellation (gray lines). (f) Alignment of reconstructed embedding from e with the original Delaunay diagram from a.

circle amplification (18), template walking amplification (19), or packing of barcoded beads (20). Unique “seed” strands are captured by primer strands on the surface (Figure 1a) and locally amplified in the immediate vicinity where they landed. This generates numerous distinct patches, or “polonies”, of amplified DNA (Figure 1b). Within each, all DNA is derived from a single seed molecule. Any of the above techniques could be applied to our method, though we focus herein on the polony-amplification-by-surface-primers approach.

By growing polonies on a surface of primers to saturation (Figure 1c), i.e. when growing polonies encounter the boundaries of other adjacent polonies, a tessellation of neighboring polonies forms. Each polony has a limited number of immediately adjacent neighboring polonies with their own respective barcodes. Though each patch is associated with a unique sequence according to its parent seed molecule, isolation of this DNA and subsequent sequencing would scramble information about the polony’s position and its neighboring polonies. Thus the critical step is to crosslink strands (SI Appendix Fig. S1) from each polony to strands from adjacent polonies (Figure 1d) in a way that enables both barcodes to be sequenced together in a single read. Recovery of the strands, i.e. stripping them from the surface followed by next-gen sequencing (by any means including non-optical approaches such as Oxford Nanopore) thus preserves topological association between neighboring polonies as pairs of barcodes — a complete set of which constitutes the whole topological network of adjacent polonies (Figure 1e). For random seed distributions we show that topological information alone, constrained by being a 2D planar network with known boundary geometry, retains significant spatial metrics of the original distribution. By generating such a mappable surface, we propose that localization of molecules bound to the surface can be done by covalent association with polonies, enabling inference of molecular spatial distributions and construction of images with polonies as pixels.

1. Results and Discussion

A. Voronoi Tessellation as a Model of Polony Saturation. The spatial distribution of polonies on a surface, the *a priori* Euclidean information that is not explicitly accessible after isolation, can be preserved by associations between adjacent polony sequences and recovered with sequencing. Information that is available after sequencing and subsequent transformations of that data are then referred to as *a posteriori*.

Assume that seed molecule amplification on a bounded 2D surface, say in the shape of a disk, takes the form of uniform circular growth. At the point of saturation, polonies have amplified to the extent that their expanding boundaries are restricted from further growth, having encountered neighboring polonies. The system of polonies then forms a planar *Voronoi tessellation* T (SI Appendix A), appearing as a characteristic mosaic of polygons with the property that every point within a given cell is closer to its parent seed point than any others. T can also be represented by its plane dual *Delaunay diagram* $D = (P, L)$ whose vertices P are the seed points of T and edges L are the line segments connecting the seed points of adjacent cells (polonies). By the geometric characteristics of T, all the faces of D are triangles (22)(Section 9).

We refer to the graph defined purely by its vertices and edges without spatial considerations as the *untethered graph.* Figure 2a presents a miniature Voronoi tessellation T formed from 9 seed points within a square and its Delaunay diagram D. The untethered graph $G = (V, E)$ (Figure 2d) is obtained from D by omitting all geometric information, retaining only topological characteristics of the Delaunay diagram D. This includes a topological distance function $t(i, j)$ defined as the fewest number of edges that must be traversed to get from one vertex i to another j, but no other information about the spatial origins of G is explicitly stored (e.g. no Euclidean coordinates of the original points).

B. Topological Metrics as a Proxy for Euclidean Metrics. Let $P = \{p_k \mid k = 1, \ldots, N\}$ be a planar placement of N seed points, resulting from a Poisson-distributed seeding with intensity (i.e. polony density) λ over an area A. Thus $N \approx \lambda A$, and an untethered graph representation $G = (V, E)$ of the true Delaunay diagram D can be obtained by:

$$
\begin{aligned}
V &= \{1, \ldots, N\}, \\
E &= \{\{i, j\} \mid \text{barcodes } w_i \text{ and } w_j \text{ co-occur}, i, j = 1, \ldots, N\}
\end{aligned}
$$

Since sufficiently long barcodes are with high probability unique (SI Appendix B), we treat pairs of barcodes as unique markers of polony adjacency. We postulate that with a sufficiently dense Poisson-distributed placement P, the topological metric on G (with an appropriate linear scaling) approximates well the actual Euclidean metric of the points in P (SI Appendix D-E). Figure 2b shows the Euclidean distance distributions for increasing topological distances from a reference vertex, for a large collection of Delaunay triangulations of Poisson random point sets. Figure 2c then plots the scaled (Equation 9.9 (21)) average Euclidean distances as a function of topological distances for Delaunay triangulations of random point sets generated by Poisson processes of increasing intensity λ, showing crucially that the two variables are proportional.

On this basis we propose that by finding a proper straight-line planar embedding of the untethered graph G we approximate also the metric properties of the underlying Delaunay diagram D and the corresponding Voronoi tessellation T. A straight-line embedding of G in a plane is determined by the placement P' of its vertices, from which the line segments L' corresponding to the edges can be deduced, thus denoted as $\langle G, P' \rangle$. Our hidden *a priori* embedding is the Delaunay diagram $D = \langle G, P \rangle$, and the goal is to approximate this with a good *a posteriori* embedding $\langle G, P' \rangle$.

One constraint on our candidate $\langle G, P' \rangle$ is that it must be *planar*, i.e. no two edges may cross each other. This is due to the physical assumption that the barcode-pairings correspond to polony adjacencies and thus cannot bridge non-neighboring polonies. There are several efficient algorithms for finding a plane embedding of a planar graph, one of which is the *Tutte* or *barycentric embedding* (23), applicable to Delaunay-diagram type graphs. Another quality constraint is that an average spatial density of the *a posteriori* vertex positions λ' should be obtained from the final distribution with no systematic variation across the reconstructed area. Finally, if we were to generate a new Delaunay triangulation from the reconstructed points (as can be done from any arbitrary set of points), this should produce a similar set of edges as the original untethered graph that was the basis for reconstruction.

Our reconstruction approach (flow diagram SI Appendix Fig. S2) starts by determining the *outer* or *boundary face* of the Delaunay diagram D underlying the untethered graph G. This can practically be done by finding the face in any planar embedding of G that has the most vertices with an intermediate planar embedding, because in D all faces except the boundary face are few-vertex triangles. Fixing the placement of the vertices on the boundary face, we then compute positions for the other vertices of G by Tutte's algorithm, which simply places each vertex at the average (barycenter) of its neighbors' positions. In the case of a Delaunay-diagram type graph with the boundary face a convex polygon, this system is guaranteed to be non-degenerate (23), and the result will be a crossing-free straight-line embedding of G.

If spatial characteristics of the original Euclidean boundary are known — for instance if we specify that all boundary points must lie on a circle of known radius — then the embedding may also be scaled to match the original Euclidean metrics. Figure 2e shows the Tutte embedding of the untethered graph (Figure 2d) with boundary points arranged uniformly around the unit disk. For comparison, we have aligned the reconstructed graph with the original Delaunay diagram (Figure 2f) by linearly transforming the planar graph to minimize the distance between corresponding vertices. We can see that relative positions are preserved albeit with local distortion that leads to slight displacement of each reconstructed vertex relative to its original seed counterpart. The algorithm thus returns approximate relative spatial positions of polonies from an input of paired polonies.

C. Simulation and Reconstruction by Embedding. We simulate the primer lawn as a hexagonally packed disk of area A with M primer sites as the region of interest (ROI) (Figure 3a). We simulate a random seeding at a polony density λ by selecting $N = \lambda A$ random sites, followed by pairing of adjacent polony primer sites and scrambling of edge data prior to reconstruction. Figure 3b shows how crosslinking leads to random pairing of adjacent sites, some of which are self-pairing events (providing no additional pairing information) and some of which are cross-polony sites that can be used to deduce the presence of a spatial boundary, with the fraction of information-bearing cross-pairs diminishing with the relative site density $\rho \overset{\text{def}}{=} M/(A\lambda)$ or average number of sites per polony (SI Appendix Fig. S3). The probabilistic nature of the pairing opens up the possibility to miss an existing boundary, particularly when the boundary is small or when ρ is low. A 2000 polony-simulated surface is shown in Figure 3c, and SI Appendix Fig. S4 shows site-linking and the corresponding

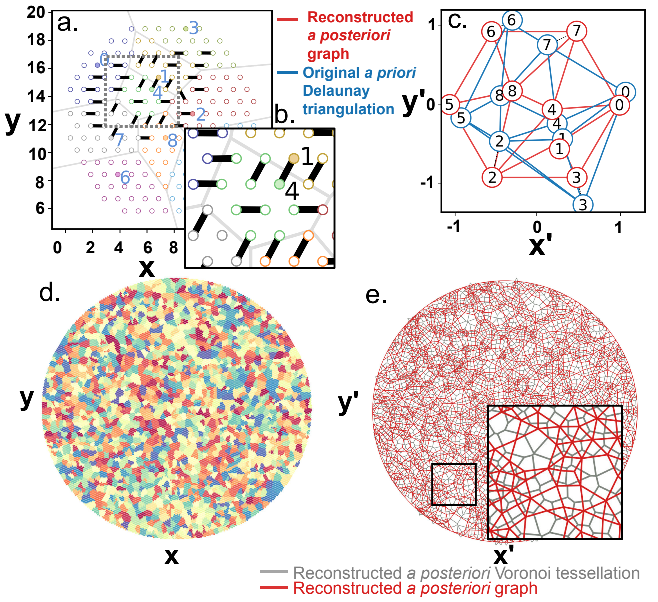
a.
y
20
18
16
14
12
10
8
6
0
2
4
6
8
x
Reconstructed *a posteriori* graph
Original *a priori* Delaunay triangulation
b.
c.
y'
1
0
-1
-1
0
1
x'
d.
y
x
e.
y'
x'
Reconstructed *a posteriori* Voronoi tessellation
Reconstructed *a posteriori* graph

Fig. 3. Simulation of polony adjacency reconstruction. (a) Lattice diagram of primer lawn and polonies denoted with color and Voronoi cell boundaries. Filled circles indicate seed locations. (b) Illustration of random site pairing between adjacent primer sites. (c) Alignment between *a priori* and *a posteriori* points from a. (d) Larger simulated surface with a polony density λ = 2000 polonies/unit area and a relative site density ρ = 50 sites per polony on average. (e) Reconstructed graph (red lines) and corresponding Voronoi tessellation (gray lines) computed using the Tutte embedding approach from scrambled edges derived from the simulated surface in d.

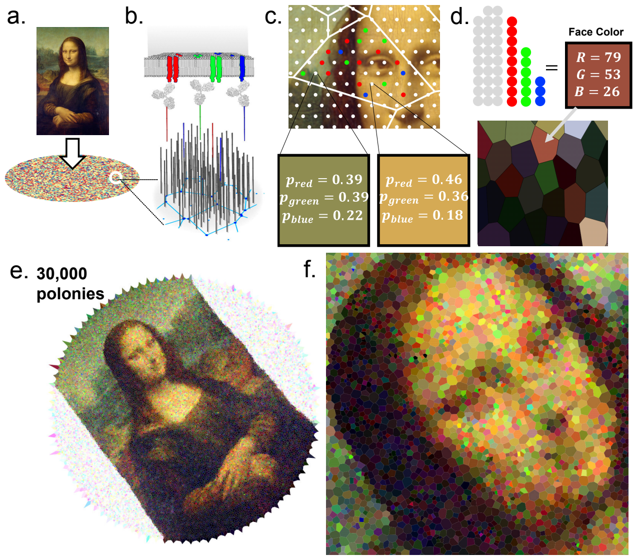

Fig. 4. Voronoi image formation. (a) An image is overlaid on a surface of primer sites. (b) Molecular markers representing different targets (R, G, and B) contact-transferred to the polony surface and each covalently linked to a polony barcode. (c) Monte Carlo sampling to determine if a marker is associated with a given site and if so which target by taking the probability from the RGB value normalized to 1 at the corresponding position in the image. (d) Tallying of markers and empty sites within a polony/Voronoi cell determines the color and brightness of that "pixel". A subsequent image (lower pane) is formed by coloring each cell accordingly. (e) Larger scale reconstruction from scrambled edge data using the Tutte embedding approach with 30,000 polonies. (f) Closeup of e revealing individual Voronoi "pixels"

Delaunay triangulation of a 500 polony example.

We reconstructed the topological network from the scrambled edges and performed intermediate embedding, boundary face determination, and Tutte embedding (Figure 3c). For this larger reconstruction, spatial uniformity is more apparent, we see that Voronoi cells take on the approximate size of polonies in the *a priori* surface, observe no obvious systematic changes in mesh density across the length of the ROI, and note the absence of crossed edges. Besides the Tutte embedding strategy, we developed 2 additional approaches for approximating Euclidean metrics from the untethered graph. One is a non-deterministic spring relaxation (24). This approach does not strictly require a crossing-free planar embedding, and can thus lead to provably false positions involving non-planar adjacency, however this feature could also be advantageous if natural interpenetration of adjacent polonies leads to such topology. The last approach (SI Appendix F) is based on the notion of topological distance $t(i, j)$ and its role as a proxy for Euclidean distance. We extend the principle of geometric triangulation, whereby the set of distances of a point to other points in a plane can be converted to Cartesian coordinates, to incorporate $t(i, j)$ as a surrogate for Euclidean distance. In one variant of this method, a total topological distance matrix is reduced to two principal component vectors approximating the x and y coordinate vectors. In the alternative variant, $t(i, j)$ of each vertex are only measured out to peripheral vertices, reducing systematic distortions. A comparison of reconstructed meshes from the different approaches is shown in SI Appendix Fig. S5-S6.

D. Stamping and Image Formation. Knowledge of polony locations could be exploited to provide spatial information about objects of interest. We devised a basic model of image reconstruction from the principle of contact or diffusion-based transfer of molecules of interest to the mapped surface, i.e. a kind of molecular stamp. As proof of concept, we use an image (Figure 4a) as a representation of a hypothetical probability distribution of 3 types of molecular markers labeled with identifying sequences called "red", "green", and "blue". The image represents a surface of interest that we would like to sample from, for example a cell surface covered in oligo-tagged antibodies, each of which would be coupled enzymatically to a given polony upon contact (Figure 4b) or diffusing RNA molecules like in (14). The color of the image corresponds to the density of such markers and thus the probability that a marker of a particular color is placed on the polony surface. To simulate molecule transfer, the overlaid lattice of primer sites denotes points where a Monte Carlo sampling will occur in the corresponding position in the image. If the image pixel at a given primer site location has an RGB value dominated by red and green for example, then there is a higher probability of that site being occupied by either a green or red marker (Figure 4c). Realistically, molecular transfer introduces distortion, e.g. from curvature of cell membranes or lateral diffusion of mRNAs.

According to the reconstruction procedure, a Voronoi tessellation is produced from the final set of vertex positions - each cell of which constitutes a pixel that can be used to form an image. The final RGB value of the cell can be determined by tallying the markers that have associated with the primer sites in the polony as well as the number of un-associated sites (Figure 4d). The Voronoi-images shown in Figure 4e and f were generated with the scrambling step that removes any spatial information of the original image and reconstructed using our algorithm. Note that global rotation and chirality are not explicitly preserved from the original image. To place this 30,000 pixel image in experimentally relevant terms, we point to a recent spatial transcriptomics manuscript (20) where circular discs of barcoded 10 μm beads (in their case sequenced optically *in situ* to obtain sequence addresses) are used to capture transcriptomic data from tissue slices. Rodriques *et al* report a typical size of 70,000 10 μm beads per 3 mm disk and obtain approximate single-cell resolution (see also SI Appendix H). Image reconstructions from the four approximation approaches are compared in SI Appendix Fig. S5-S6.

E. Assessment of Distortion and Precision. We may characterize reconstruction quality by defining a distortion metric. The *a priori* seed distribution points have a 1-to-1 correspondence with points in the *a posteriori* reconstruction, and since we generated the *a priori* points ourselves, we can directly compare corresponding original and inferred positions by applying a linear transform $Tx(P)$ (rotation, mirroring, scaling, and translation) to the set of reconstructed points that minimizes net displacement between the two distributions. Distortion is thus defined as the set of displacements: $Df = (Df_i, ...Df_N) \overset{\text{def}}{=} d(P, Tx(P')) \mid min(\sum_{i=1}^{N} d(p_i, Tx(p'_i)))$. Averaged over multiple runs, we obtain 2D histograms (Figure 5a and SI Appendix Fig. S7-S8) of distortion as a function of position in the ROI. Increasing the polony density (λ) reduces average distortion $\overline{Df} = \frac{1}{N}\sum_{i=1}^{N} Df_i$(Figure 5d and SI Appendix Fig. S10) whereas changes in the site density ρ (Figure 5f and SI Appendix Fig. S11) has a negligible effect on $\overline{Df}$ except at $\rho < 100$ sites per polony near the point of network disconnection from absent edges. Examining a single simulation (Figure 5b) we can visualize typical distortions, persistent over limited local scales and occurring with greater probability near the boundaries. Analysis of the radial distribution of this instance (Figure 5c) reveals this as a mild systematic worsening near the boundary, an artifact introduced by the algorithm's treatment of vertices on the boundary. SI Appendix Fig. S9 compares single instance distortions for the different reconstruction approaches.

We also characterize reconstruction quality with Levenshtein distance ($lev_{G,G'}$), the number of edits needed to make two graphs identical, between the untethered graph and set of edges derived from a Delaunay triangulation D' generated from the final reconstructed coordinates. Importantly, this metric is based only on *a posteriori* information, so it can be used in an experimental context where knowledge of the underlying distribution is unavailable. It weakly but positively correlates with distortion for a given λ (SI Appendix Fig. S13). $lev_{G,G'}$ grows linearly with λ (Figure 5e and SI Appendix Fig. S10), and like distortion is relatively constant as a function of ρ with a transient catastrophic breakdown at low ρ (Figure 5g SI Appendix Fig. S11). We also measured a classical resolution, the full width half maximum (FWHM) of a point spread function (Figure 5h), by sampling the inferred position of a single site (taking its position to be the centroid of whatever Voronoi cell it lands in). Like distortion, FWHM is approximately $\propto 1/\sqrt{\lambda}$ (Figure 5i) indicating that to halve the minimum size of distinguishable features, one should quadruple λ (SI Appendix Fig. S12). In experimental terms, polonies generated from techniques like template walking amplification, which forms polonies from sites that must be near the packing limit of oligo surface immobilization, can be on the order of nanometers (19) (SI Appendix G).

2. Discussion and Conclusion

The three reconstruction methods (Tutte embedding, spring relaxation, and topological distance matrix) succeed in producing approximations of the original seed distributions that can be used to generate images. Tutte embedding exhibited the best estimated algorithmic complexity (based on run time scaling with λ, SI Appendix Fig. S14) making it the fastest technique which becomes significant for large reconstruction problems ($\lambda > 10,000$ polonies/unit area). Both Tutte embedding and spring relaxation had the lowest distortion levels, with Tutte embedding exhibiting slightly better Df and $lev_{G,G'}$ scaling with λ. Tutte embedding was sensitive to catastrophic failure at low ρ, with singly-connected edges crashing the reconstruction, and all four approaches were sensitive to disjoint subgraphs - making noisy and unconnected graph data a likely challenge for experimental scenarios. SI Appendix Fig. S13 and SI Appendix I discuss our attempts to move towards an algorithm that optimally exploits the available information, and future research should seek to establish a provably maximum-entropy reconstruction that is efficient and deterministic.

Along these lines, utilizing information such as the number of self-pairing events could be useful to extract more information and weight edges according to estimated polony size and better control point placement. Alternatively, low-information content self-pairing events could be prohibited through a bipartite network approach whereby only pairings between A-type and B-type polonies would be allowed (SI Appendix Fig. S15). The bridge amplification approach to polony generation leaves the possibility of doing this with two species of independent primers on the surface and two interpenetrating/overlapping and independently saturated polony surfaces. Another possible approach is series growth of polonies. In the basic concept presented in previous sections, a primer of uniform sequence is assumed, however generation of a saturated layer of polonies that could then be used as primers for a subsequent polony generation step would then result in an overlapping of every 2nd-layer polony with multiple 1st-layer polonies. This would result in efficient pairing of barcodes without the need for subsequent crosslinking steps.

At the time of publication, we are aware of immediately prior works whose contributions are complementary to ours on development of DNA-sequencing based microscopy (25, 26). The former work experimentally demonstrates DNA microscopy with images of mRNA in cells using locally confined cDNA amplifications and polymerase extension-based fusion of barcodes to connect spatial patches. Their approach differs from ours through the fact that fusion events are used as a direct distance metric, whereas our data instead relies on topology as a proxy for Euclidean metrics. The latter work uses series proximity ligation to associate planar spatial patches and form a network, utilizing a spring relaxation approach for reconstruction.

A. Conclusion. PARSIFT is a concept for microscopic image reconstruction using spatial information encoding in DNA base format. We showed an *in silico* proof of concept by constructing a pipeline for taking decoupled edge data, generated from simulated polony distributions, that are then reassembled into a topological network and embedded in a Euclidean plane, resuming spatial characteristics of the original seed distribution. We saw that global distortions are low enough to resolve whole images. We hold that this framework and pipeline for reconstruction could be exploited for image acquisition of micro- and nano-scale surfaces with molecular libraries of potentially very high multiplicity and with throughput automated in a way that would not be possible with most optical approaches.

Supporting Information (SI). The code is available at https://github.com/Intertangler/parsift

ACKNOWLEDGMENTS. The authors thank Ferenc Fördős for insightful discussions. This work was supported by Åke Wiberg Stiftelsen grant for medical research M17-0214 to ITH., the Knut and Alice Wallenberg project grant 2017.0114, the Knut and Alice Wallenberg Academy Fellow grant KAW2014.0241 to BH, and by Academy of Finland grant 311639 to PO.

1. Church GM, Gao Y, Kosuri S (2012) Next-generation digital information storage in DNA. *Science* p. 1226355.
2. Söderberg O, et al. (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature Methods* 3(12):995.
3. Jungmann R, et al. (2010) Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano letters* 10(11):4756–4761.
4. Wang G, Moffitt JR, Zhuang X (2018) Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Scientific Reports* 8(1):4847.
5. Karaiskos N, et al. (2017) The Drosophila embryo at single-cell transcriptome resolution. *Science* 358(6360):194–199.
6. Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* 33(5):495.
7. Achim K, et al. (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nature Biotechnology* 33(5):503.
8. Halpern KB, et al. (2017) Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 542(7641):352.
9. Lein E, Borm LE, Linnarsson S (2017) The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* 358(6359):64–69.
10. Wang X, et al. (2018) Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* p. eaat5691.
11. Ke R, et al. (2013) In situ sequencing for rna analysis in preserved tissue and cells. *Nature Methods* 10(9):857.
12. Lee JH, et al. (2015) Fluorescent in situ sequencing (fisseq) of rna for gene expression profiling in intact cells and tissues. *Nature Protocols* 10(3):442.
13. Crosetto N, Bienko M, Van Oudenaarden A (2015) Spatially resolved transcriptomics and beyond. *Nature Reviews Genetics* 16(1):57.
14. Ståhl PL, et al. (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353(6294):78–82.
15. Peikon ID, et al. (2017) Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Research* 45(12):e115–e115.
16. Schaus TE, Woo S, Xuan F, Chen X, Yin P (2017) A DNA nanoscope via auto-cycling proximity recording. *Nature communications* 8(1):696.
17. Adessi C, et al. (2000) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research* 28(20):e87–e87.
18. Korfhage C, et al. (2017) Clonal rolling circle amplification for on-chip DNA cluster generation. *Biology Methods and Protocols* 2(1).
19. Ma Z, et al. (2013) Isothermal amplification method for next-generation sequencing. *Proceedings of the National Academy of Sciences* 110(35):14320–14323.
20. Rodriques SG, et al. (2019) Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363(6434):1463–1467.
21. Miles RE (1970) On the homogeneous planar Poisson point process. *Mathematical Biosciences* 6:85–127.
22. de Berg M, van Krefeld M, Overmars M, Cheong O (2008) *Computational Geometry: Algorithms and Applications, 3rd Rev. Ed.* (Springer-Verlag).
23. Tutte WT (1963) How to draw a graph. *Proceedings of the London Mathematical Society* 3(1):743–767.
24. Kamada T, Kawai S, , et al. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters* 31(1):7–15.
25. Weinstein JA, Regev A, Zhang F (2019) DNA microscopy: Optics-free spatio-genetic imaging by a stand-alone chemical reaction. *Cell.*
26. Boulgakov A, Xiong E, Bhadra S, Ellington AD, Marcotte EM (2018) From space to sequence and back again: Iterative DNA proximity ligation and its applications to DNA-based imaging. *bioRxiv.*

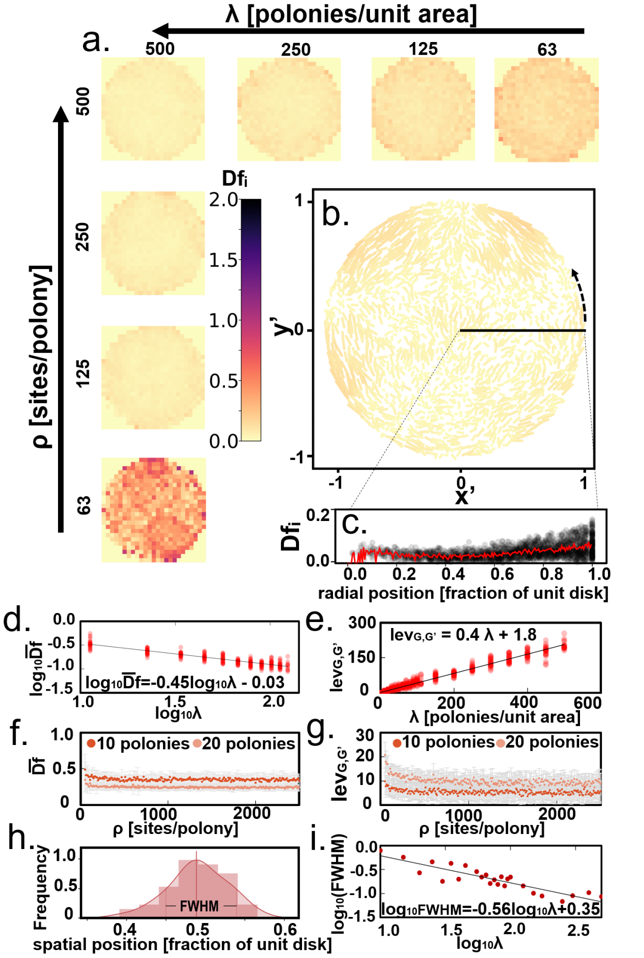

Fig. 5. Reconstruction quality. (a) 2D histograms of average displacement values binned by relative position in the unit disk (n = 5000/λ simulations per histogram) for varied parameters (λ and ρ). (b) Distortion in a single 2000 polony Tutte embedding with lines connecting *a priori* and *a posteriori* vertex locations. Color map indicates line length (max = unit disk diameter 2.0). (c) Radial profile of distortion in b and 5 point moving average (red line). (d) Log-log plot of average displacement versus λ (points single individual simulatoins reconstructions) and fixed $\rho = 500$ sites per polony showing displacement approximately $\propto 1/\sqrt{\lambda}$. (e) Linear plot of Levenshtein distance ($lev_{G,G'}$) between untethered and *a posteriori* Delaunay graphs as function of polony. (d and e: n = 25 simulations per λ value) (f) Plot of average displacement and (g) plot of $lev_{G,G'}$ each as a function of ρ for two values of λ, error bars represent standard deviation (f and g: n = 25 simulations per point, error bars: standard dev.) (h) Single instance of full width half maximum (FWHM) of *a posteriori* point spread function of a single site. (i) Log-log plot of FWHM versus λ, scaling approximately according to the negative square root of polony density.