Quantum-inspired techniques for learning the geometry of data

Akshat Kumar*

Instituto de Telecomunicações,
Instituto Superior Técnico, Lisboa

Physics of Information and Quantum Technologies Group

Mohan Sarovar

Sandia National Labs, Livermore, CA

QuantHEP 2023, Istituto Nazionale di Fisica Nucleare

September 26, 2023

- **Data analysis tasks are *often* geometric analysis of data**

- **Geometry of HEP data analysis**

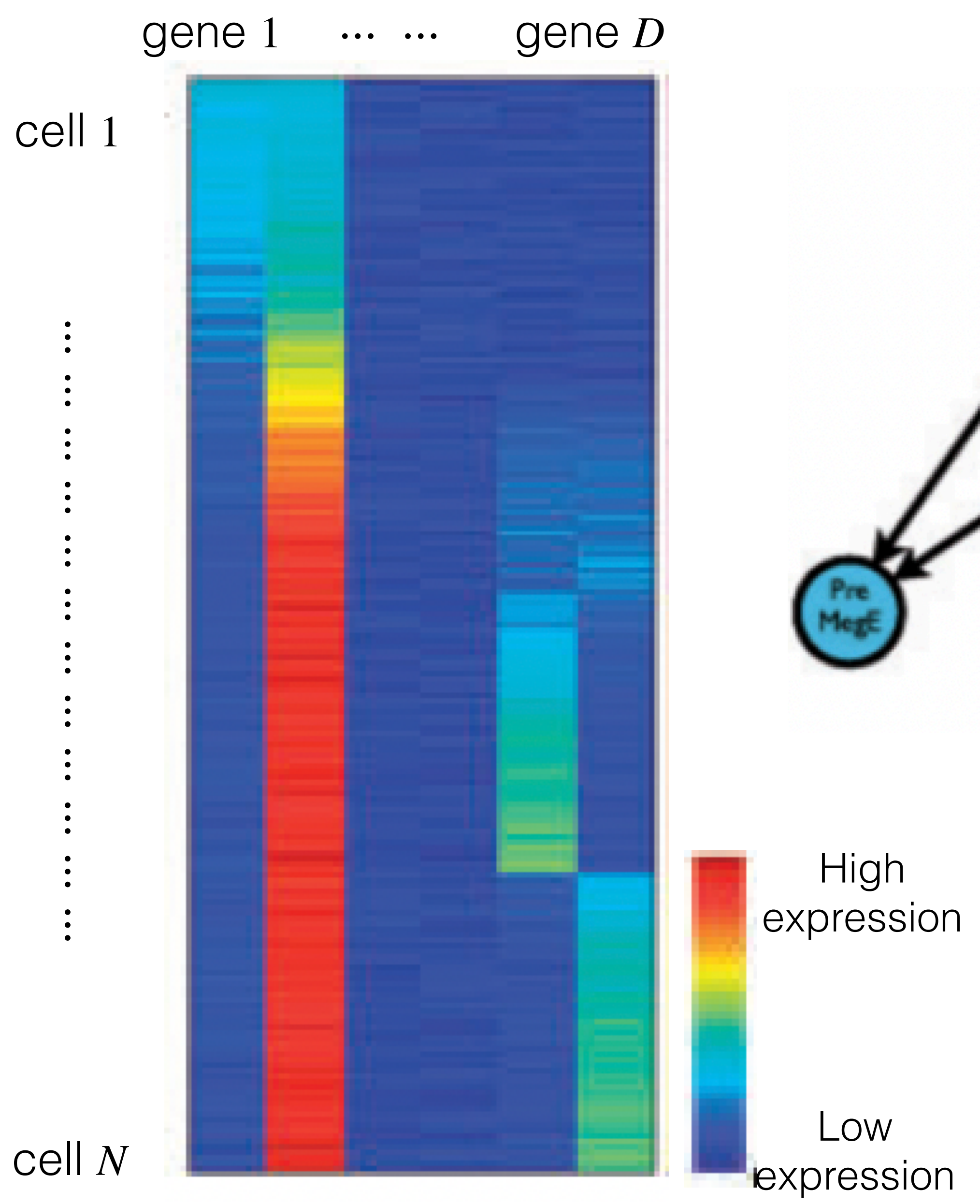
- **Quantum dynamics underlying the geometry of data**
  - **Example applications**

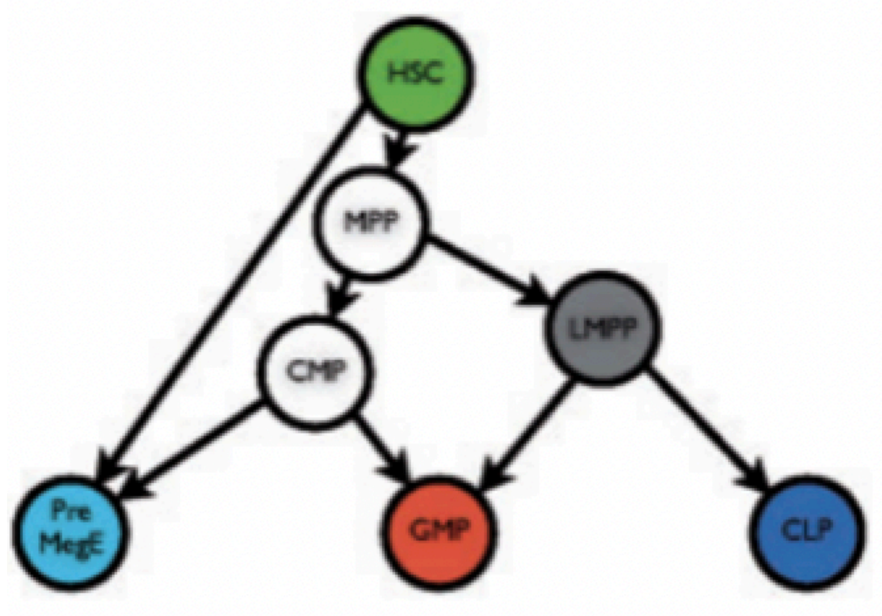
- **Future directions for HEP applications**

Data is often an *organized* point cloud

What is “data”?

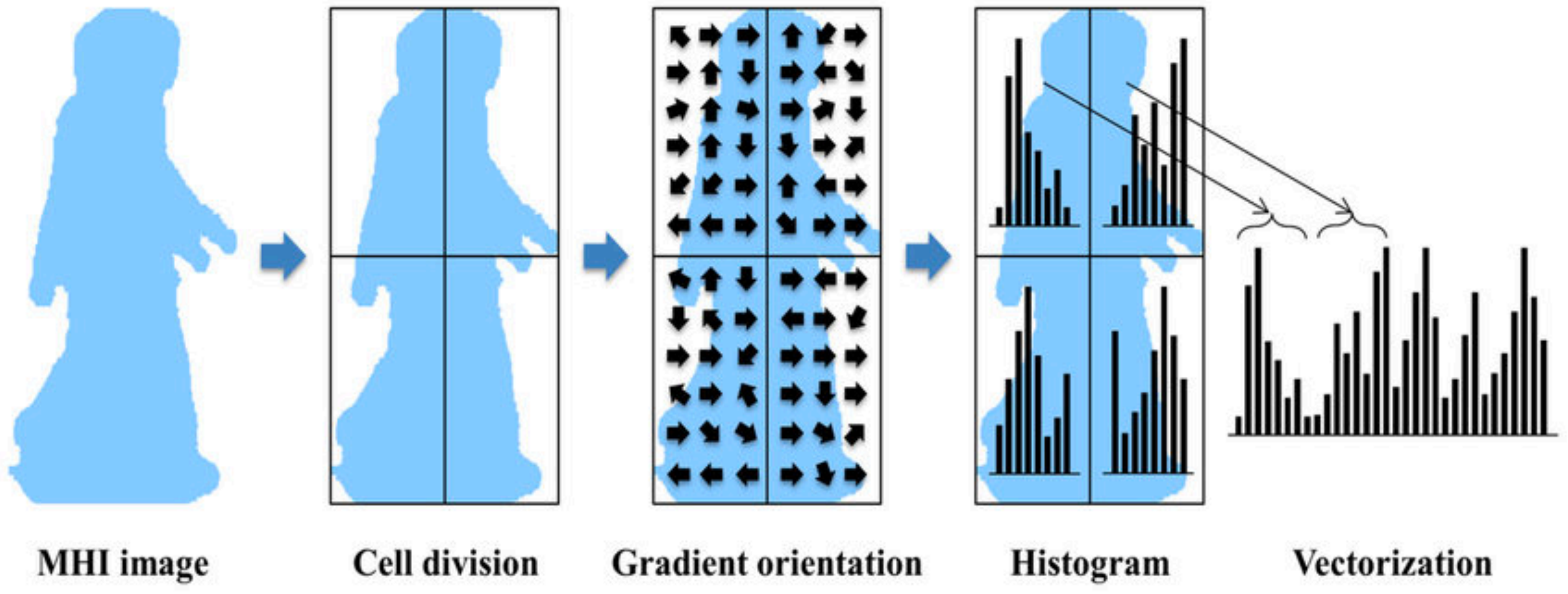
N vectors in $\mathbb{R}^D$



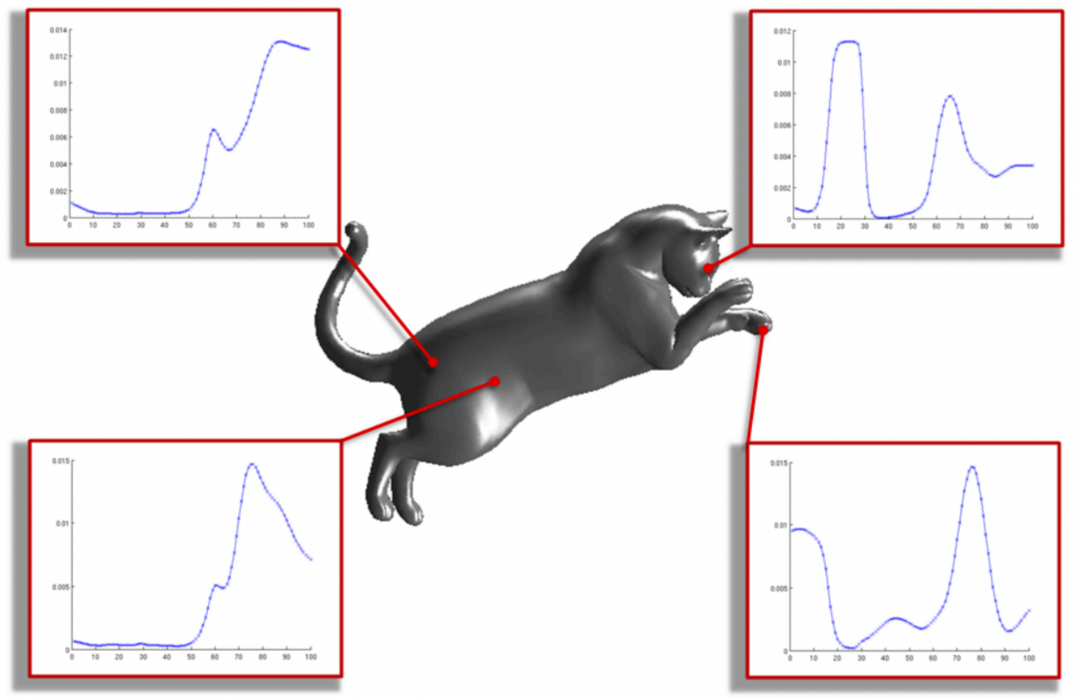


Hierarchy of cell types

Haghverdi et al., *Bioinformatics* (2015)



Eum et al., *Sensors* (2015)

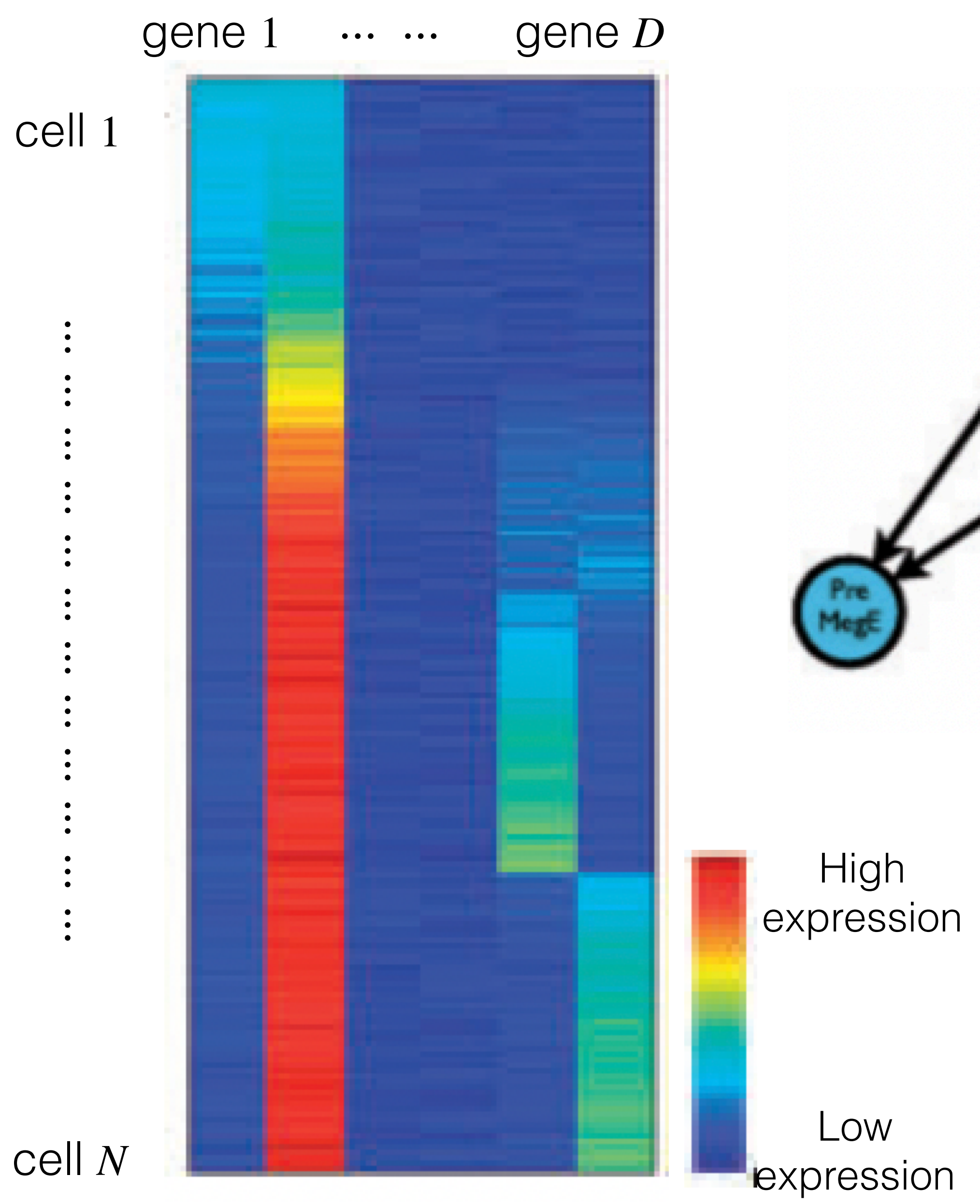


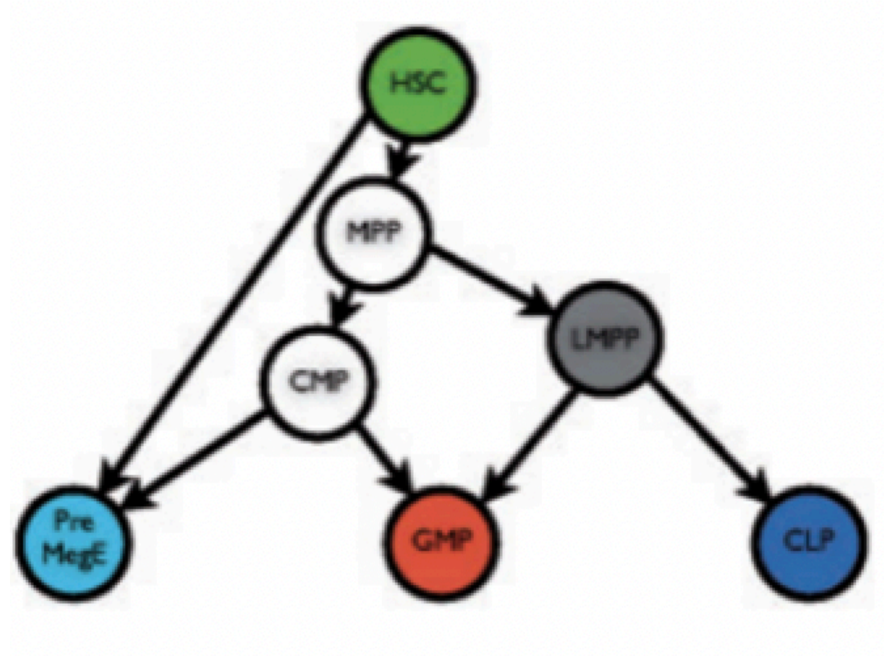
Aubry et al., *ICCV workshop, IEEE* (2011)

Measurements, including statistics of natural processes, constitute data

What is "data"?

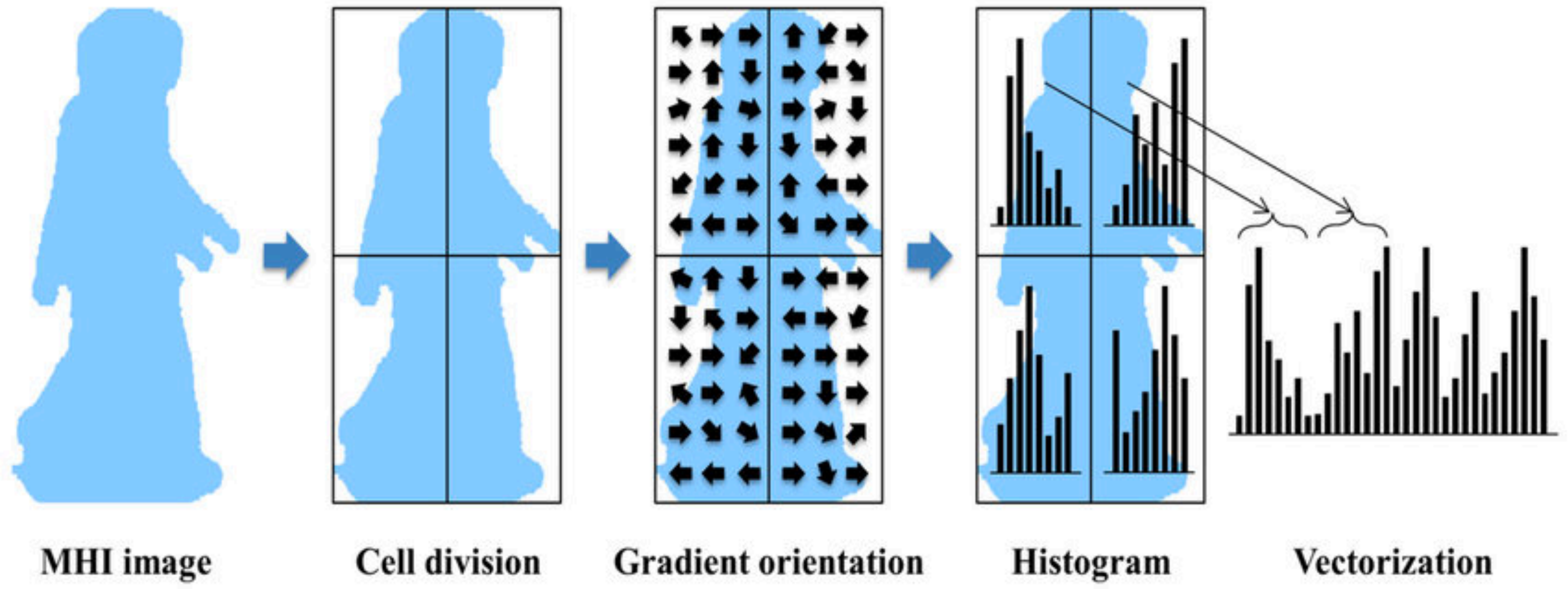
N vectors in $\mathbb{R}^D$



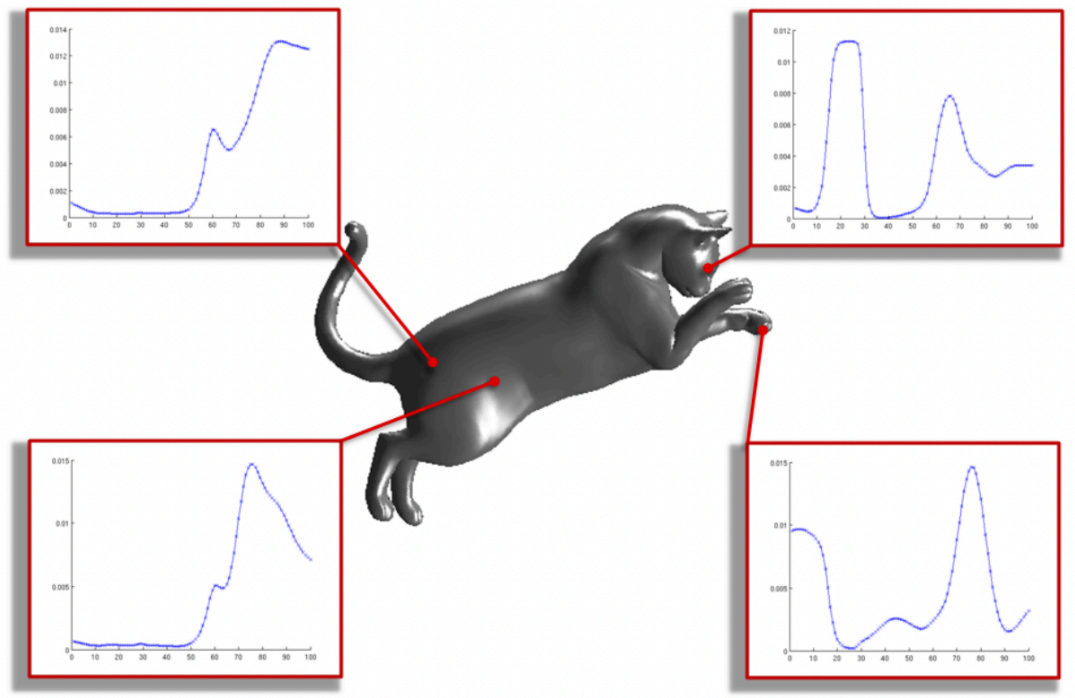


Hierarchy of cell types

Haghverdi et al., *Bioinformatics* (2015)



Eum et al., *Sensors* (2015)



Aubry et al., *ICCV workshop, IEEE* (2011)

Measurements, including statistics of natural processes, constitute data

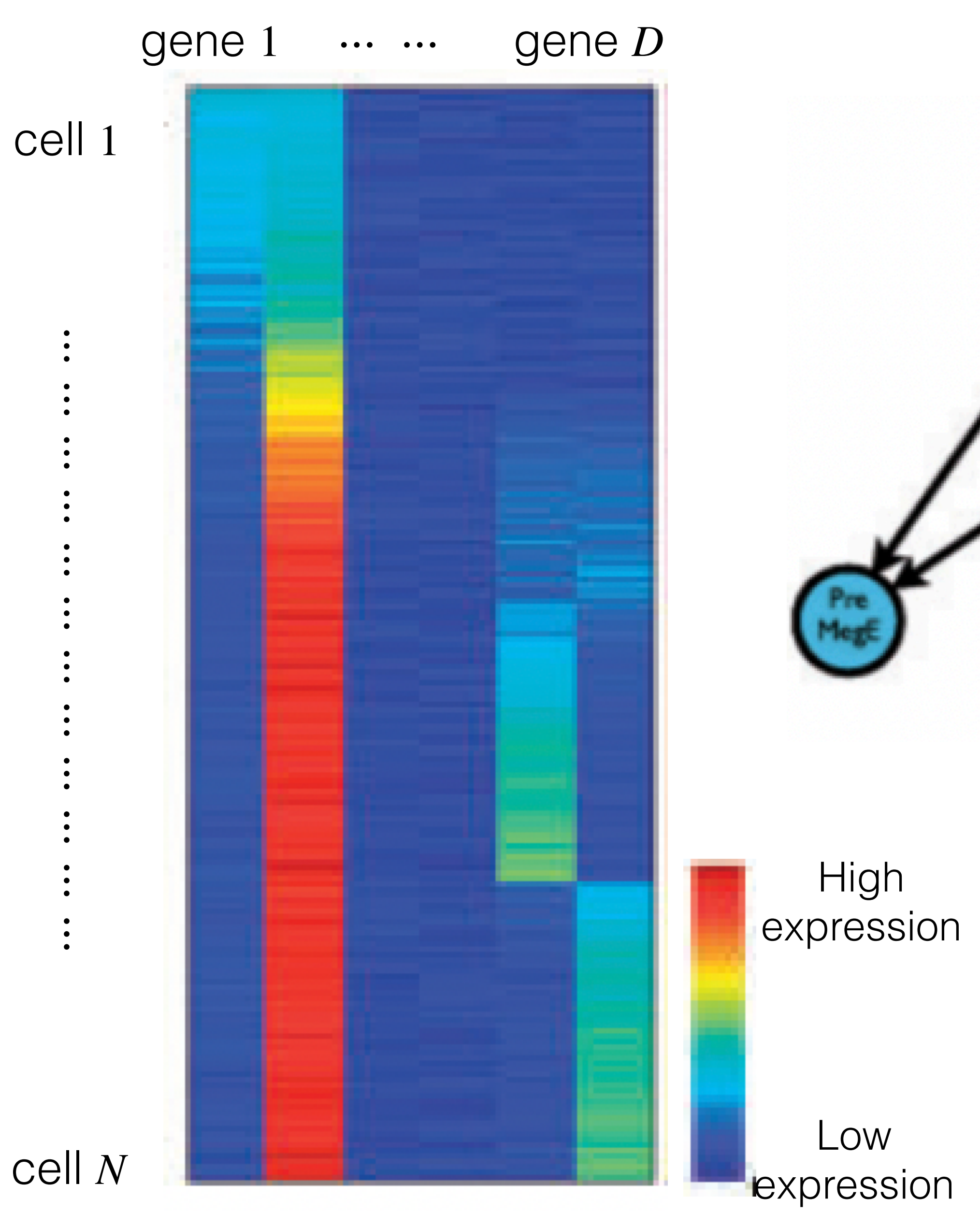
- measuring/sampling D variables in an experiment
- recording the output of D sensors

$$v := (\text{meas.}_1, \ldots, \text{meas.}_D) \in \mathbb{R}^D$$

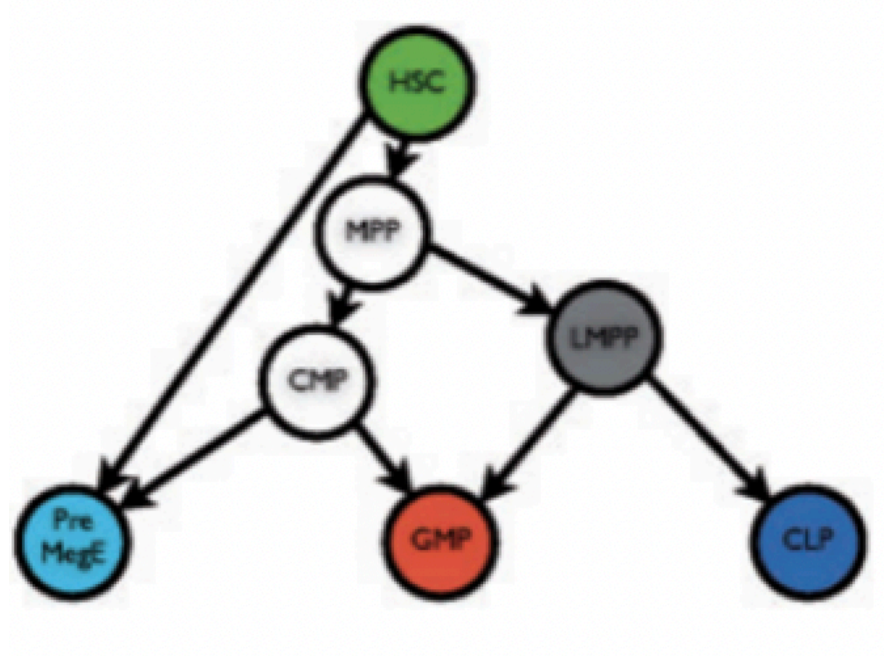
"feature vector"

What is “data”?

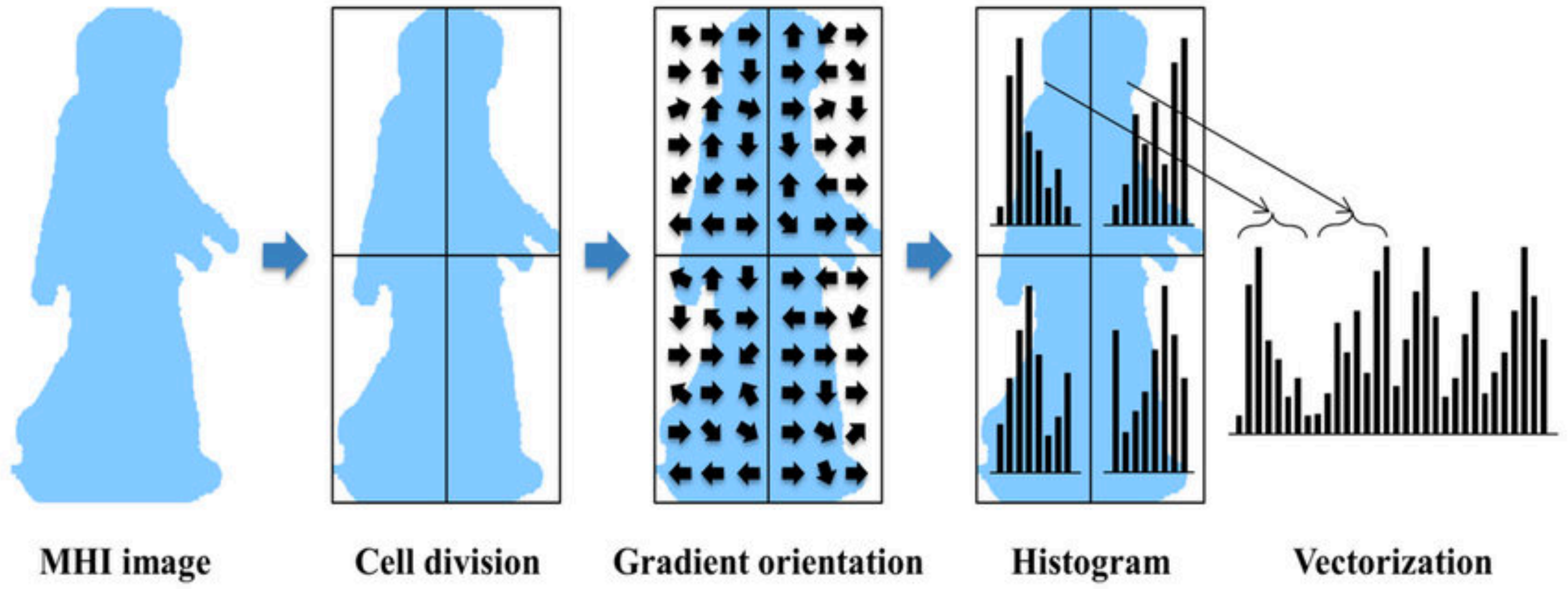
N vectors on low-dim. *submanifold*



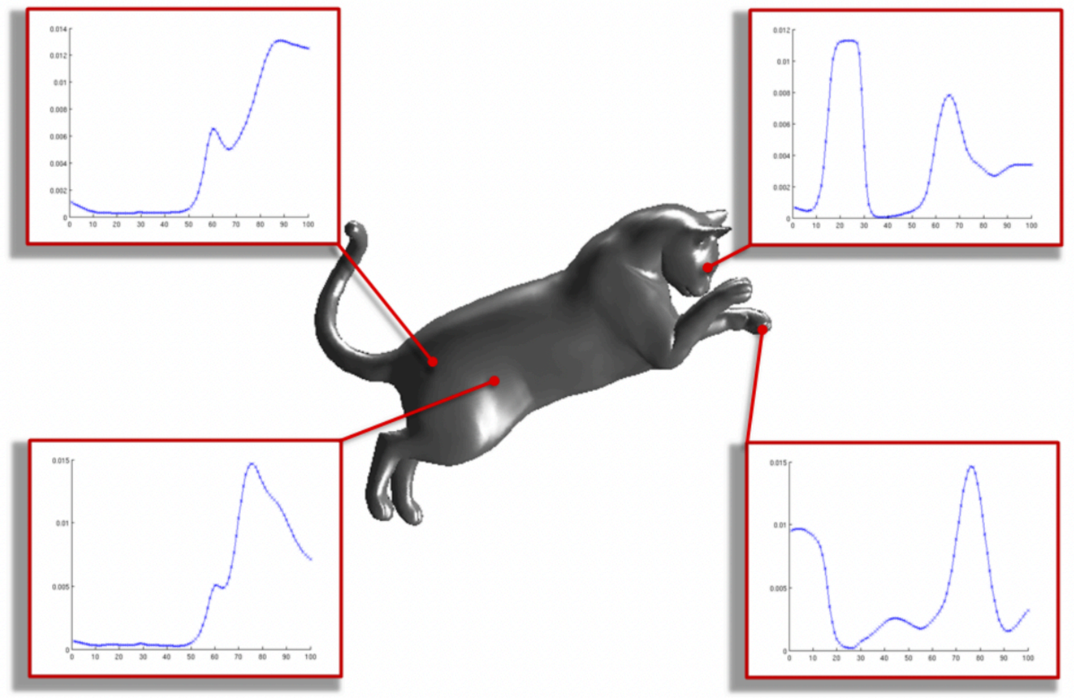
Haghverdi et al., *Bioinformatics* (2015)



Hierarchy of cell types



Eum et al., *Sensors* (2015)



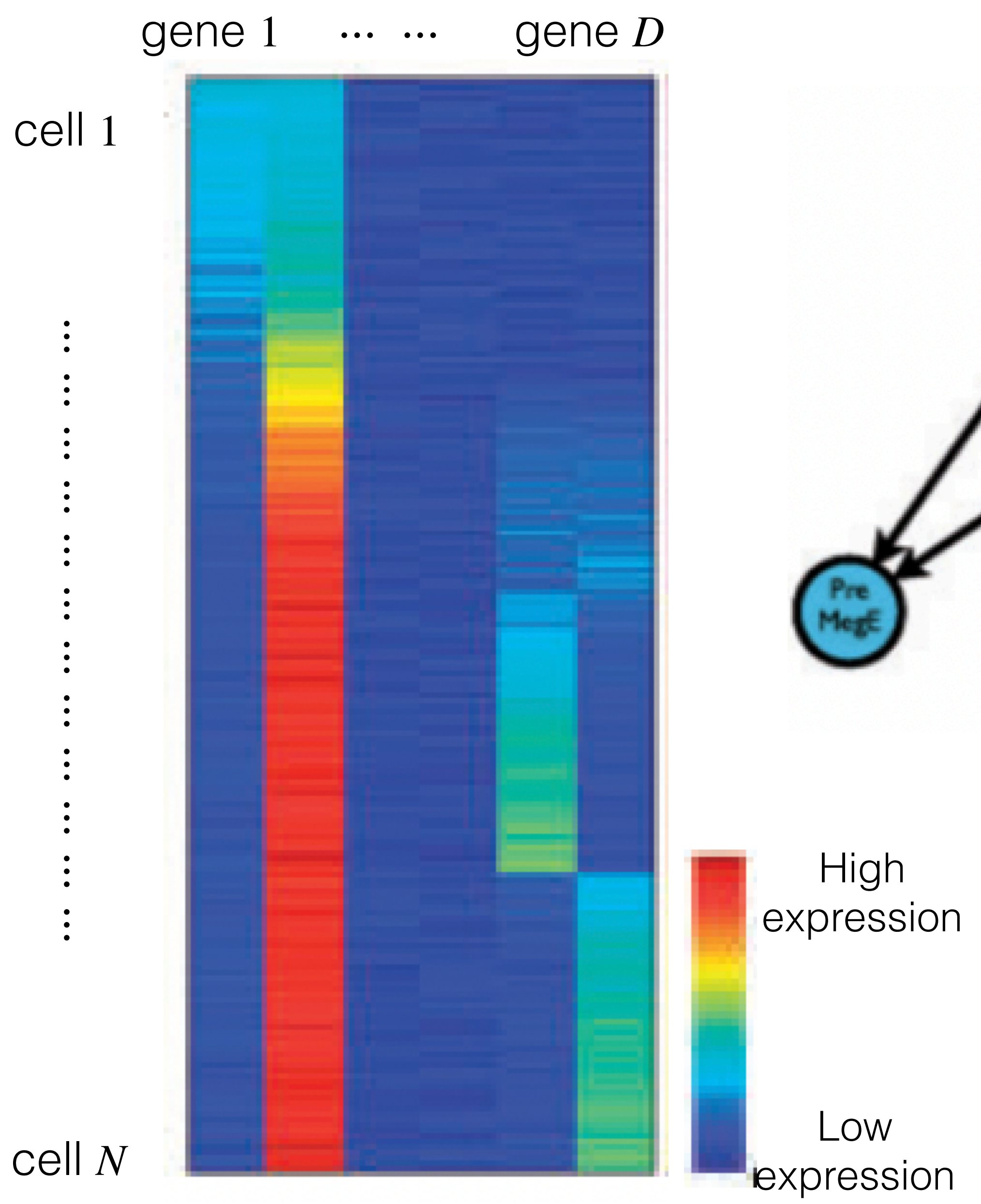
Aubry et al., *ICCV workshop, IEEE* (2011)

The manifold hypothesis:
“high dimensional data tend to lie in the vicinity of a low dimensional manifold”.

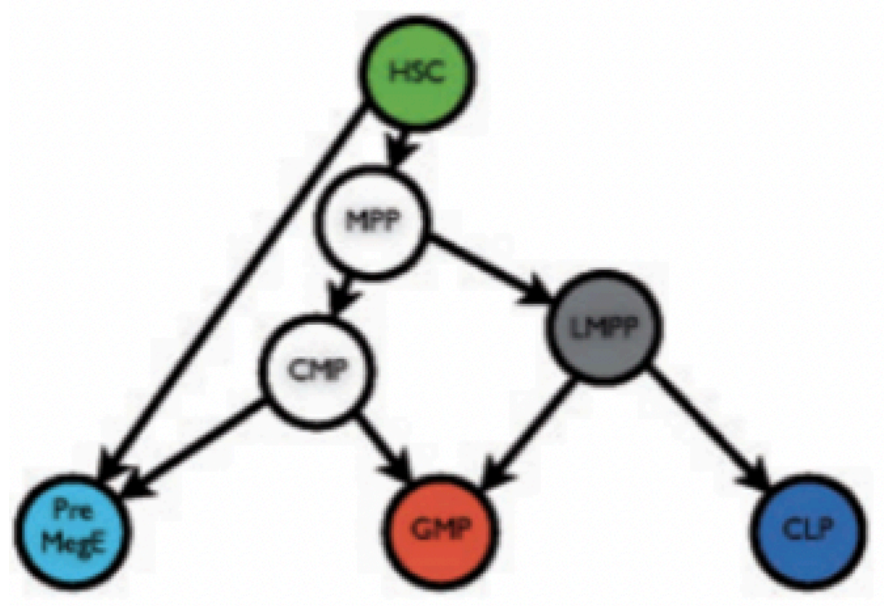
Fefferman, Mitter, Narayanan. J. Am. Math. Soc., 29, 983 (2016)

What is “data”?

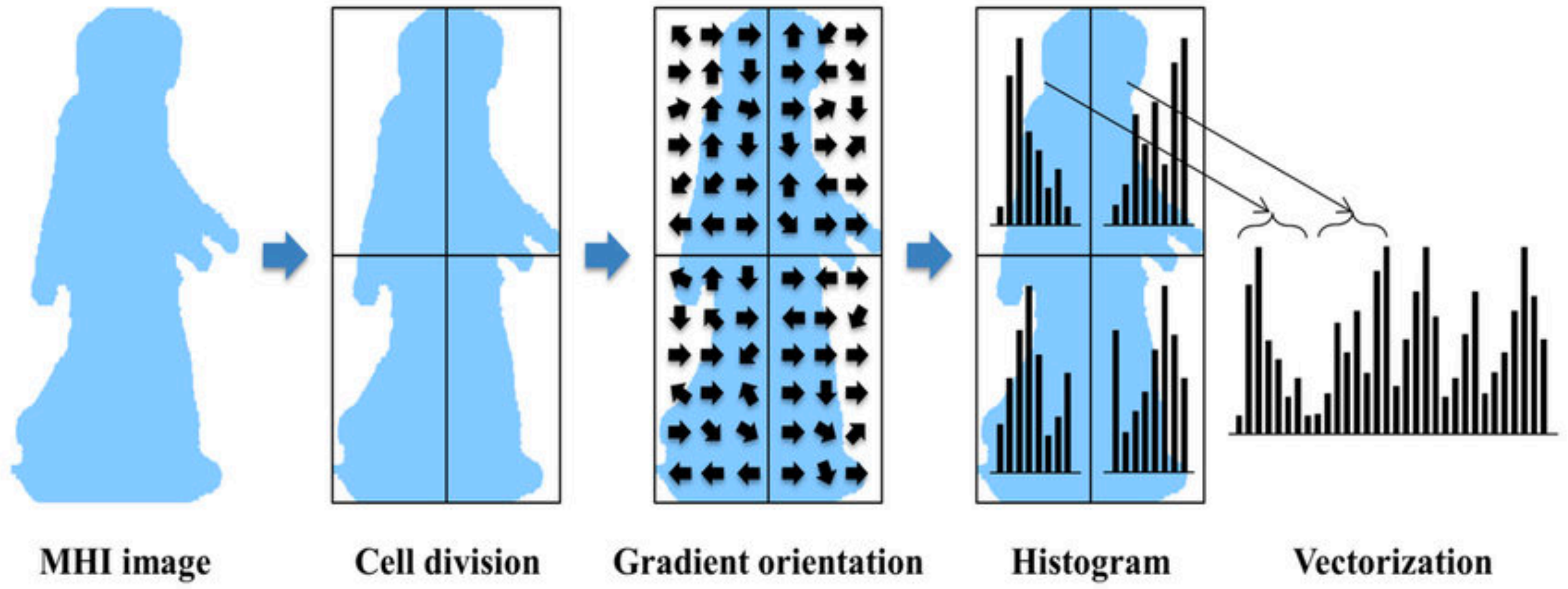
N vectors on low-dim. *submanifold*



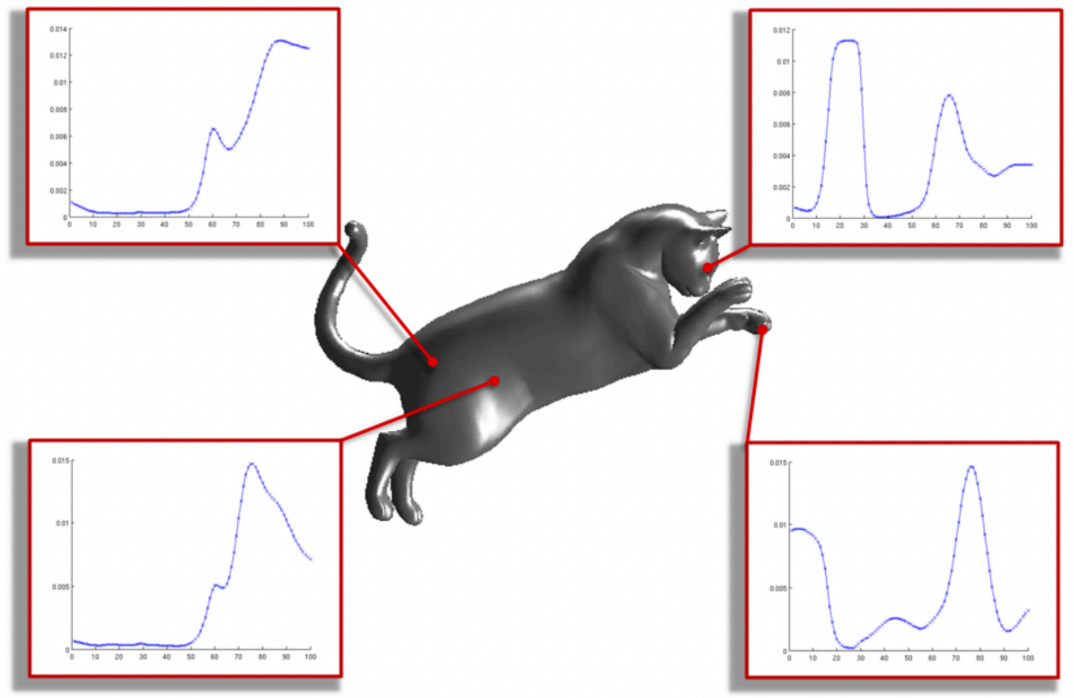
Haghverdi et al., *Bioinformatics* (2015)



Hierarchy of cell types



Eum et al., *Sensors* (2015)



Aubry et al., *ICCV workshop, IEEE* (2011)

The manifold hypothesis:
“high dimensional data tend to lie in the vicinity of a low dimensional manifold”.

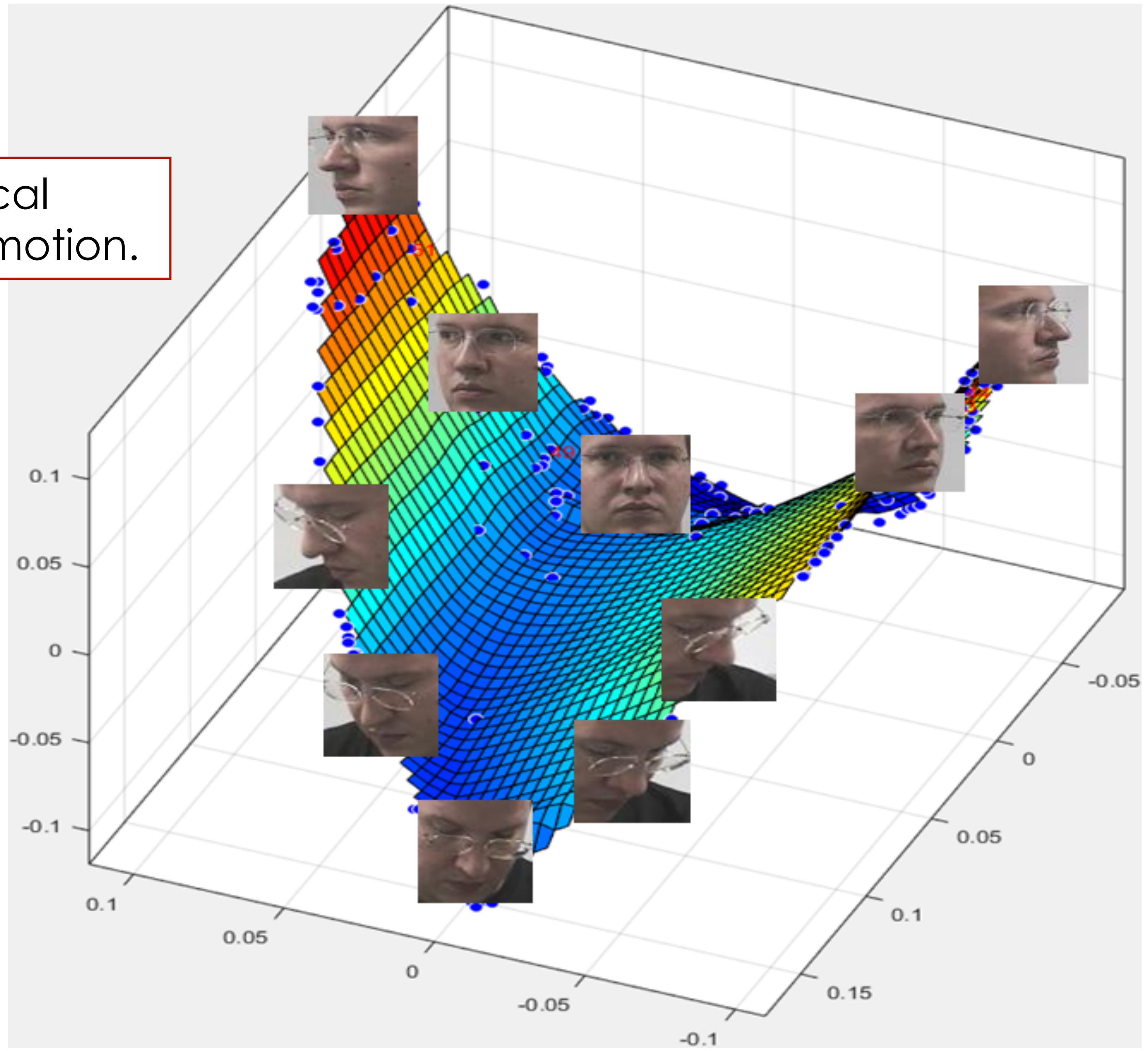
Fefferman, Mitter, Narayanan. J. Am. Math. Soc., 29, 983 (2016)

Real-world data has low-dimensional structure

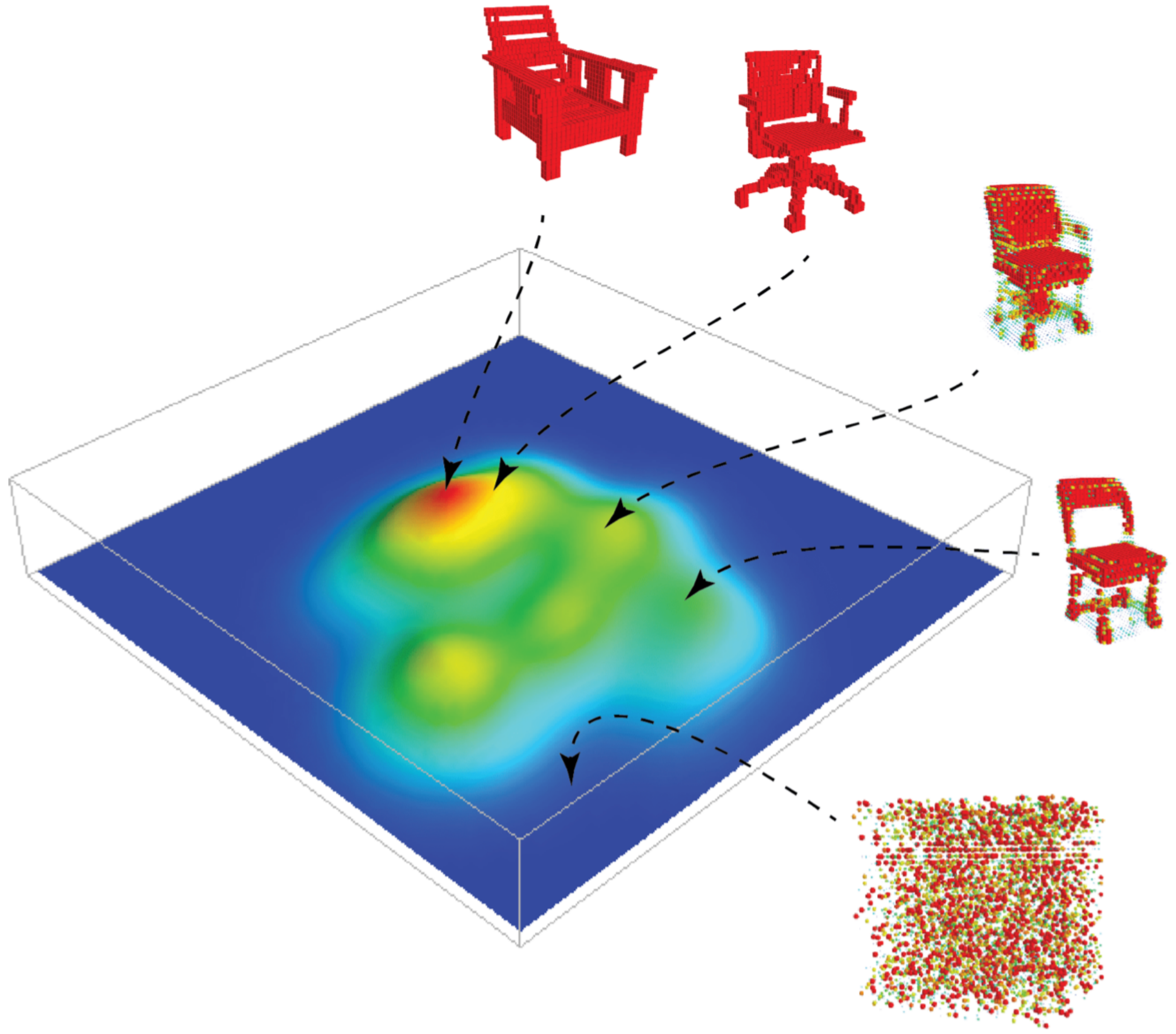
What is “data”?

N vectors on low-dim. *submanifold*

Data generated by dynamical system follows equations of motion.



https://www.intechopen.com/books/manifolds-current-research-areas/head-pose-estimation-via-manifold-learning



Randomly generated image of NxN pixels ≠ real world scene.

Gwak et al. *International Conference on 3D Vision (3DV)*, IEEE (2017).

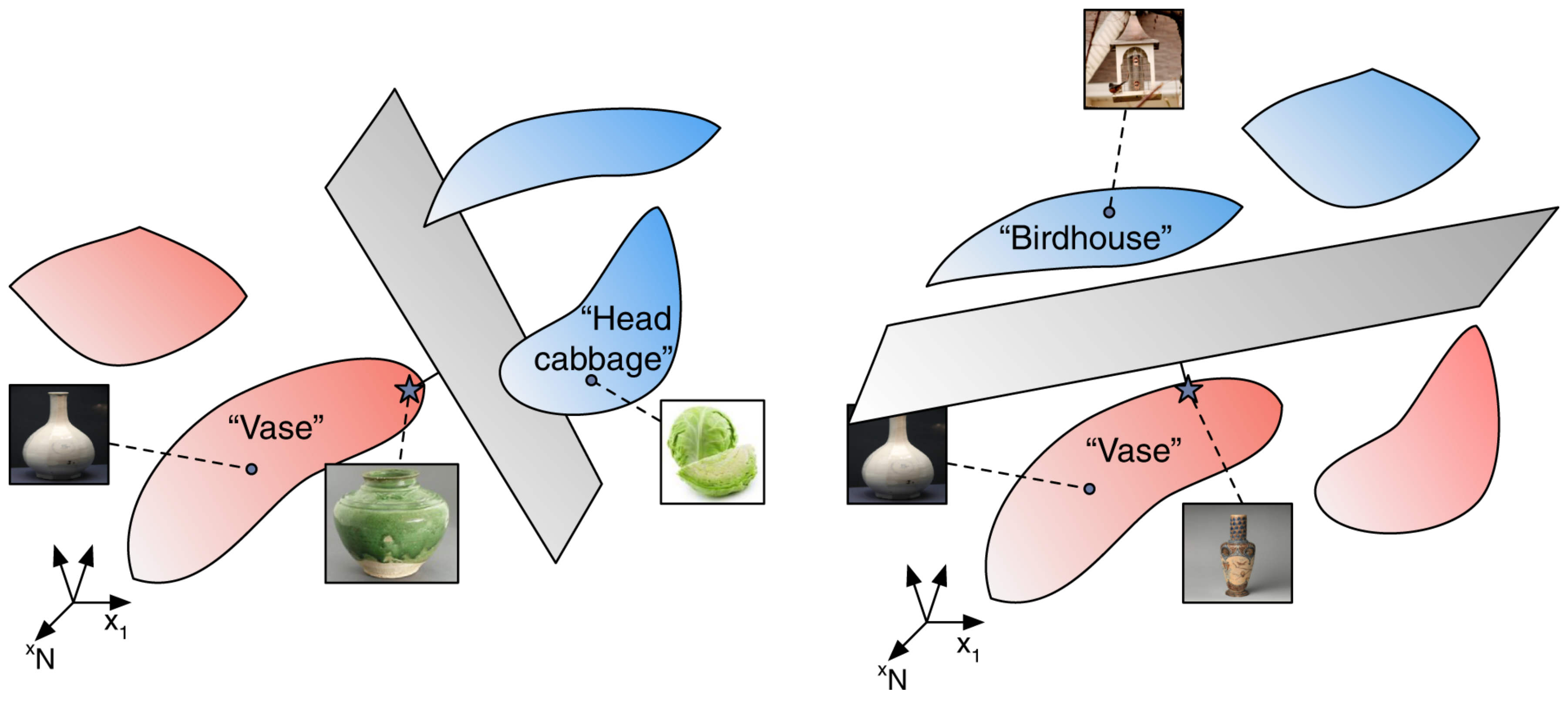
The manifold hypothesis:
“high dimensional data tend to lie in the vicinity of a low dimensional manifold”.

Fefferman, Mitter, Narayanan. J. Am. Math. Soc., 29, 983 (2016)

Real-world data has low-dimensional structure

Data analysis $\approx$ Recovering geometry of data

Classification within neural network's *concept space*:

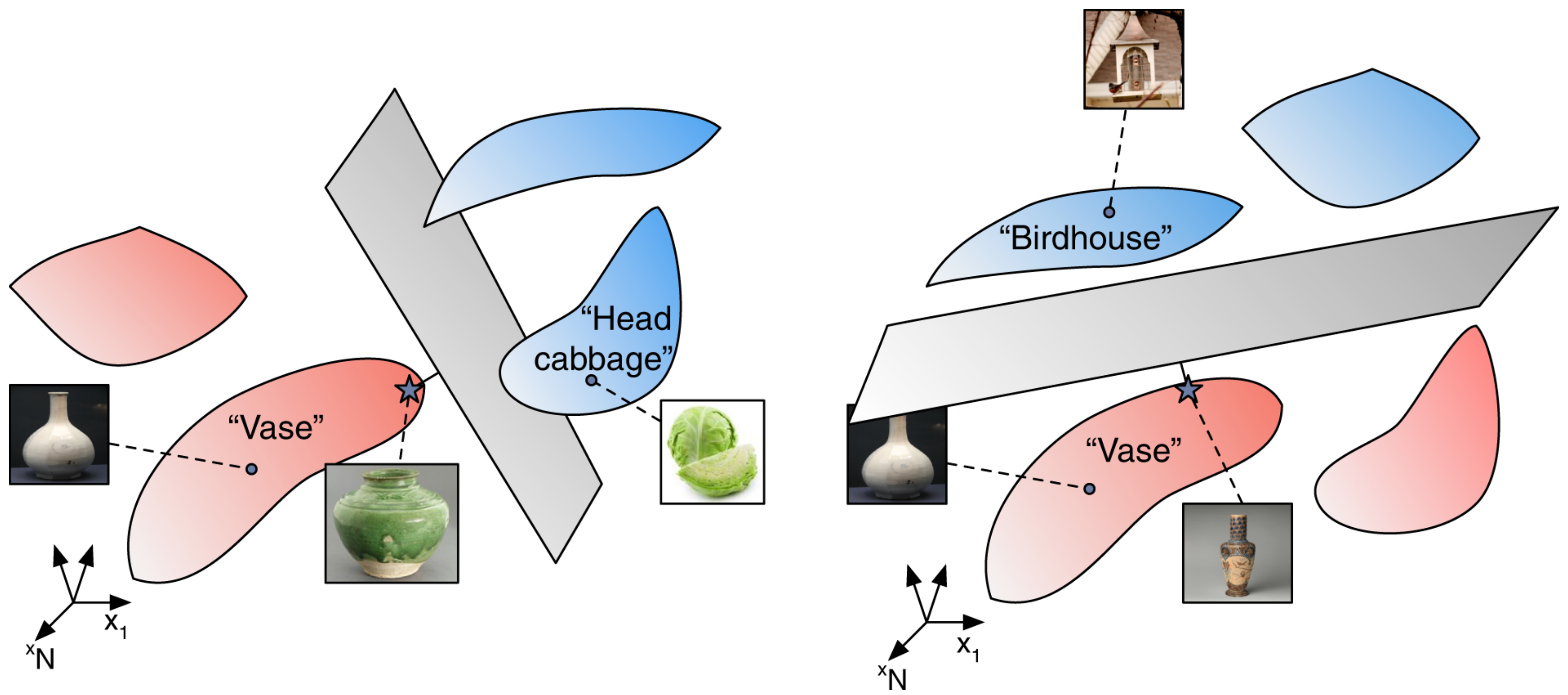


Cohen et al. *Nat Commun* 11, 746 (2020).

- Classification

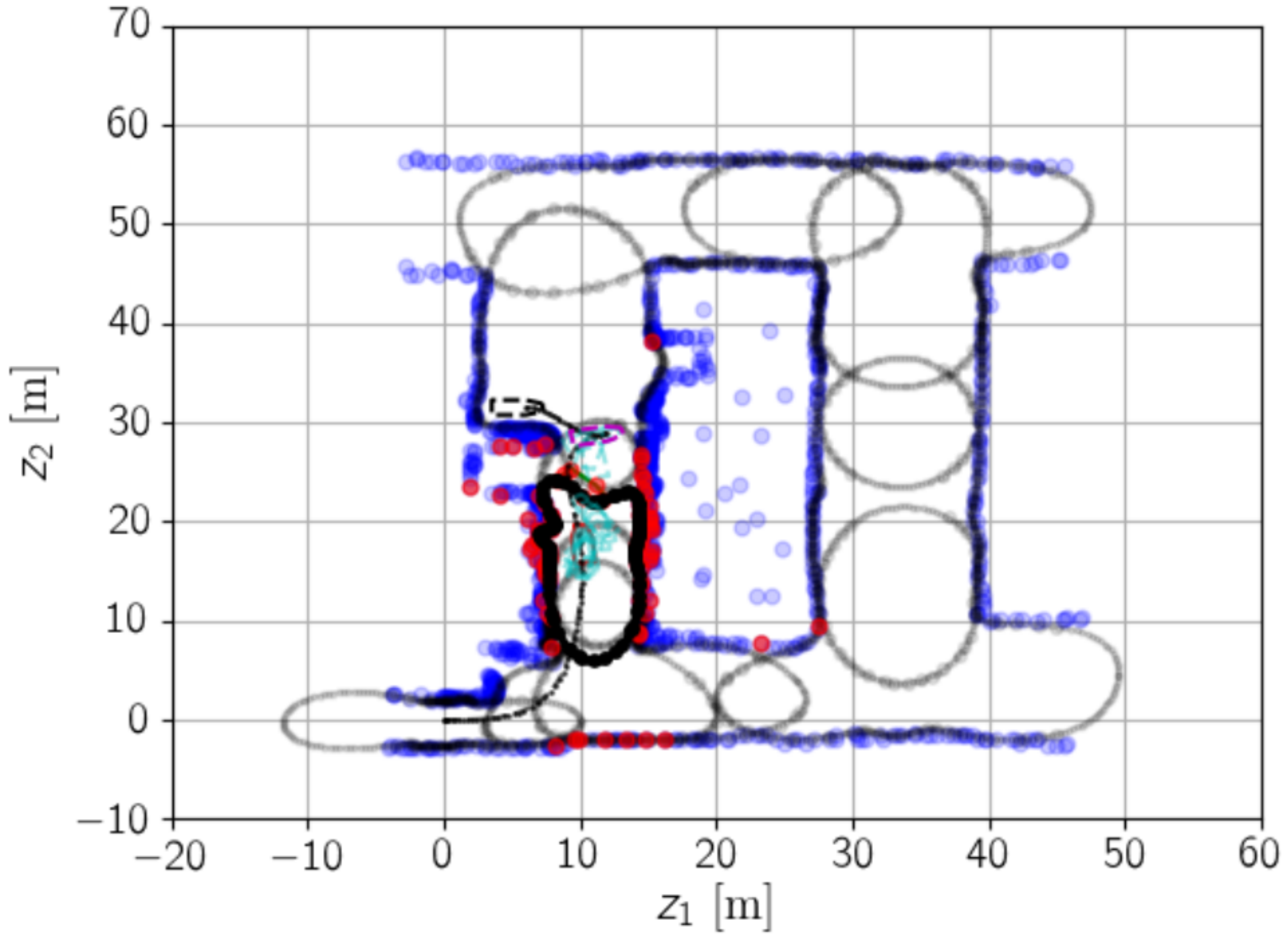
Data analysis ≈ Recovering geometry of data

Classification within neural network's *concept space*:



Cohen et al. *Nat Commun* 11, 746 (2020).

Obstacle avoidance in autonomous driving:

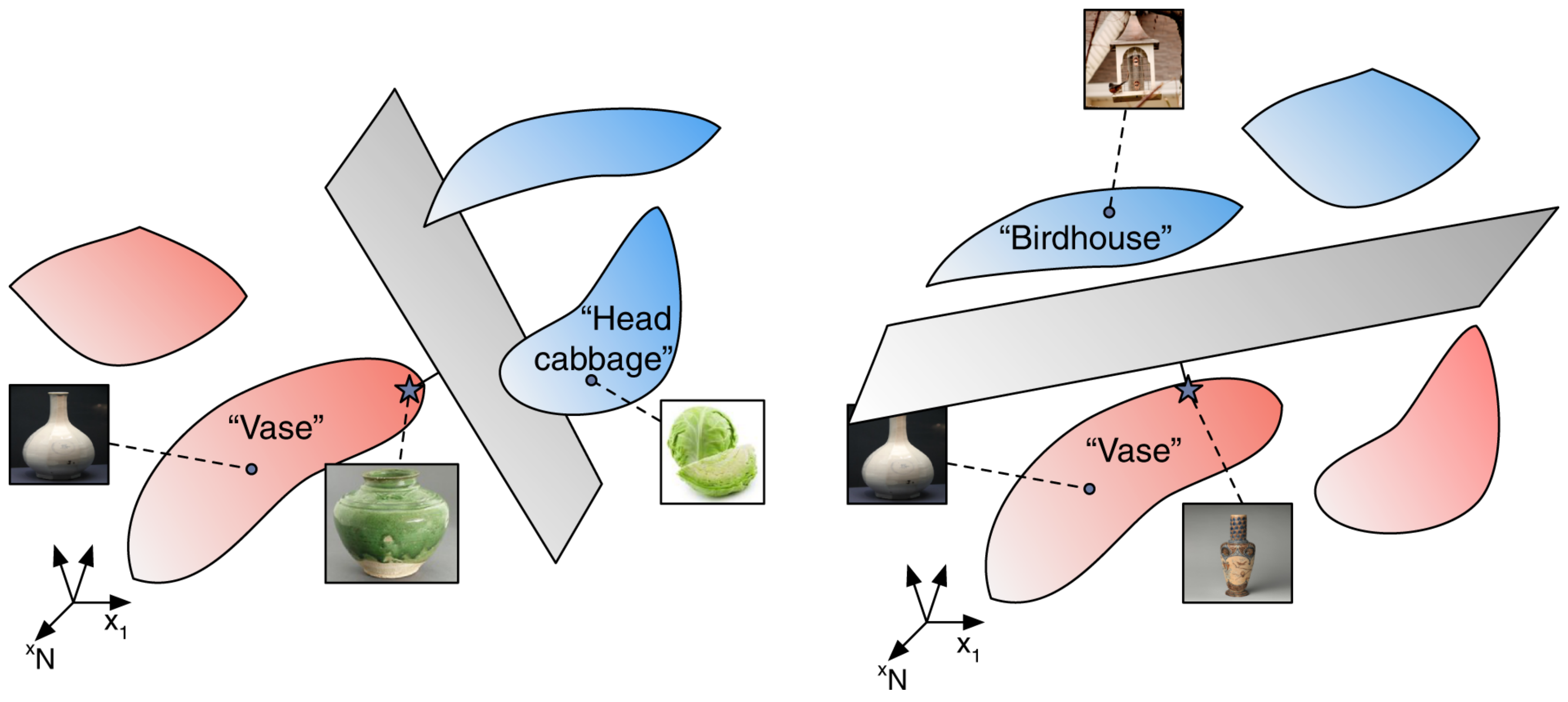


Diwale et al., https://infoscience.epfl.ch/record/265381

- Classification
- Route planning

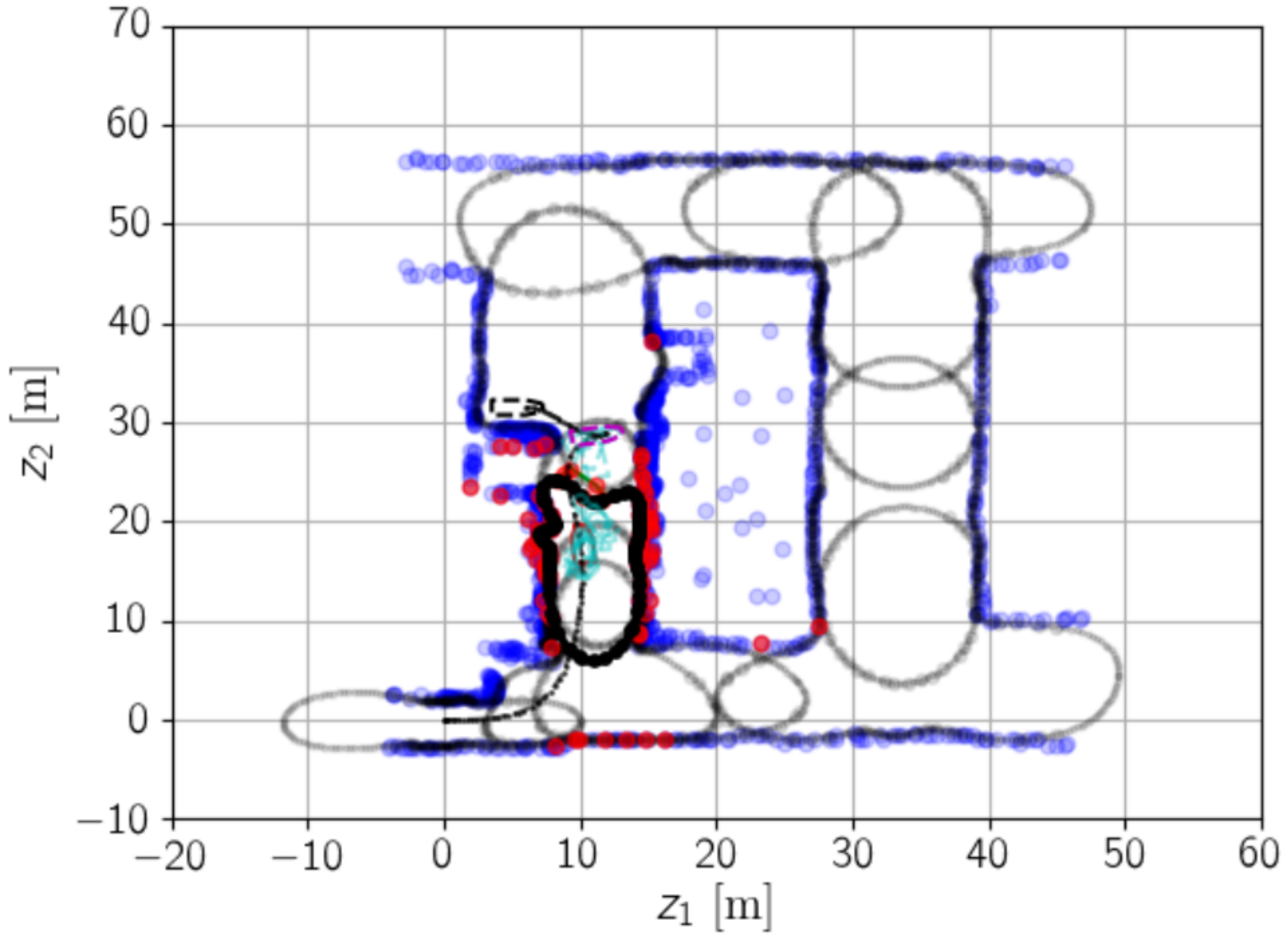
Data analysis ≈ Recovering geometry of data

Classification within neural network's *concept space*:



Cohen et al. *Nat Commun* 11, 746 (2020).

Obstacle avoidance in autonomous driving:

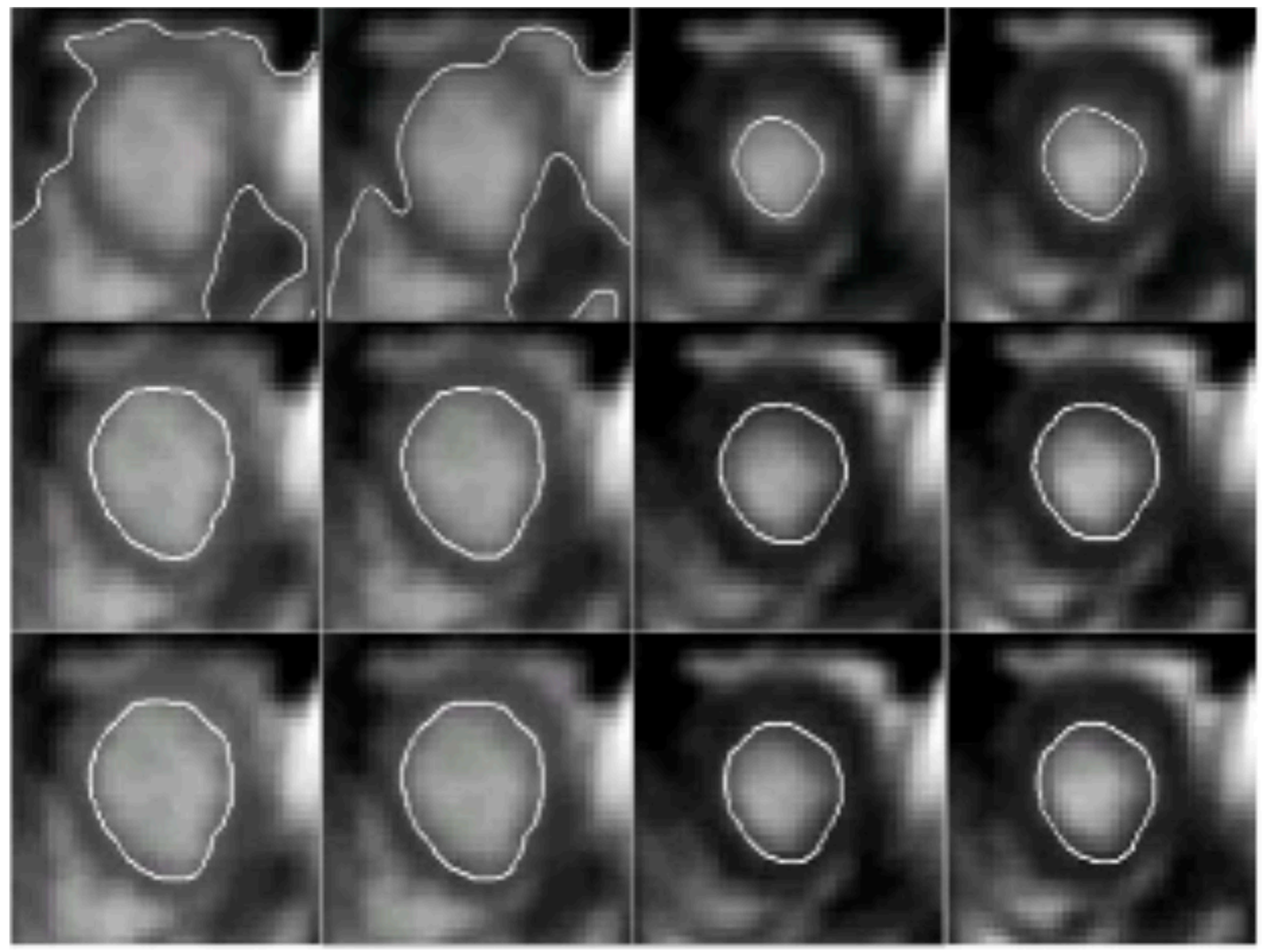


Diwale et al., https://infoscience.epfl.ch/record/265381

- Classification
- Route planning

Data analysis ≈ Recovering geometry of data

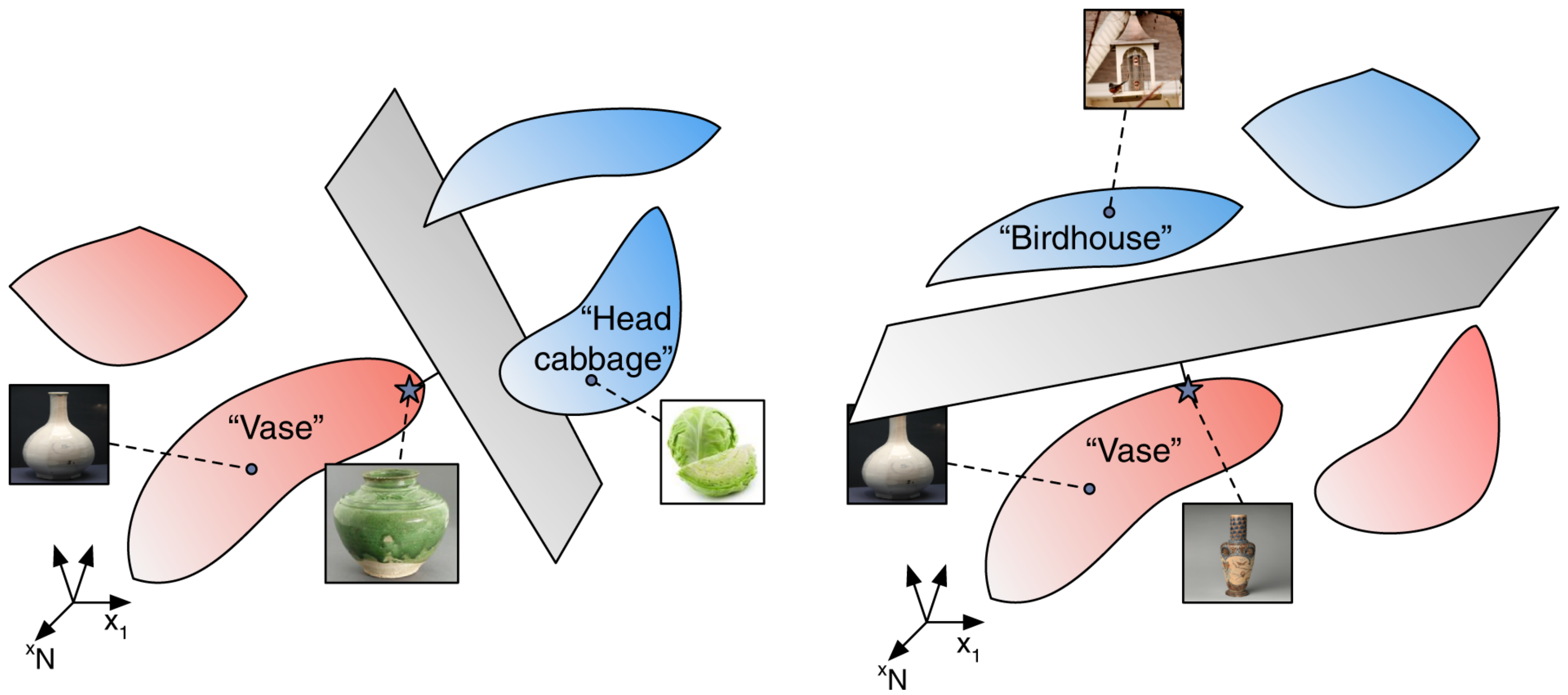
Image segmentation for medical imaging:



Zhang et al., IEEE CVPR, 1092 (2006).

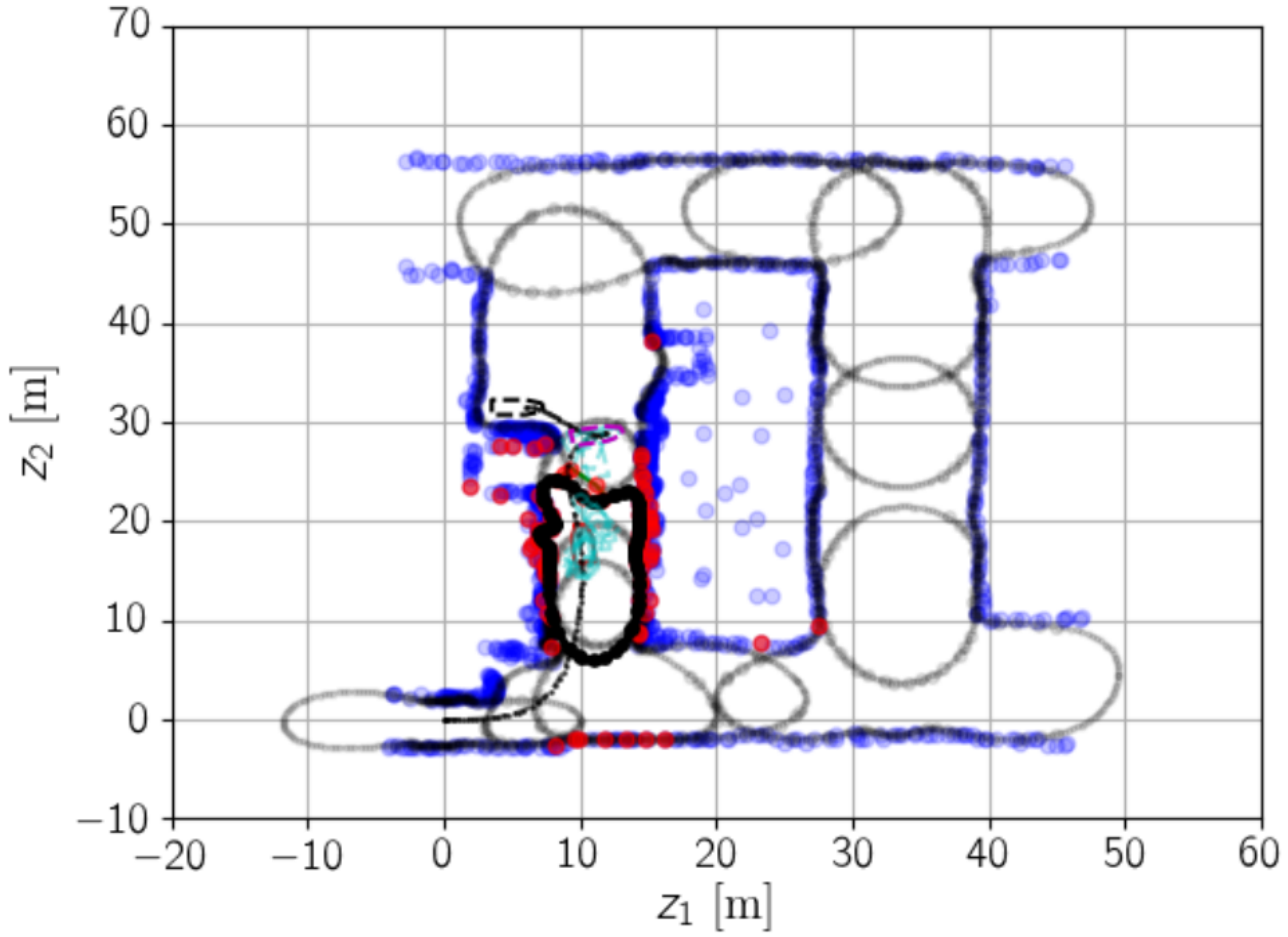
- Image segmentation

Classification within neural network's *concept space*:



Cohen et al. *Nat Commun* 11, 746 (2020).

Obstacle avoidance in autonomous driving:

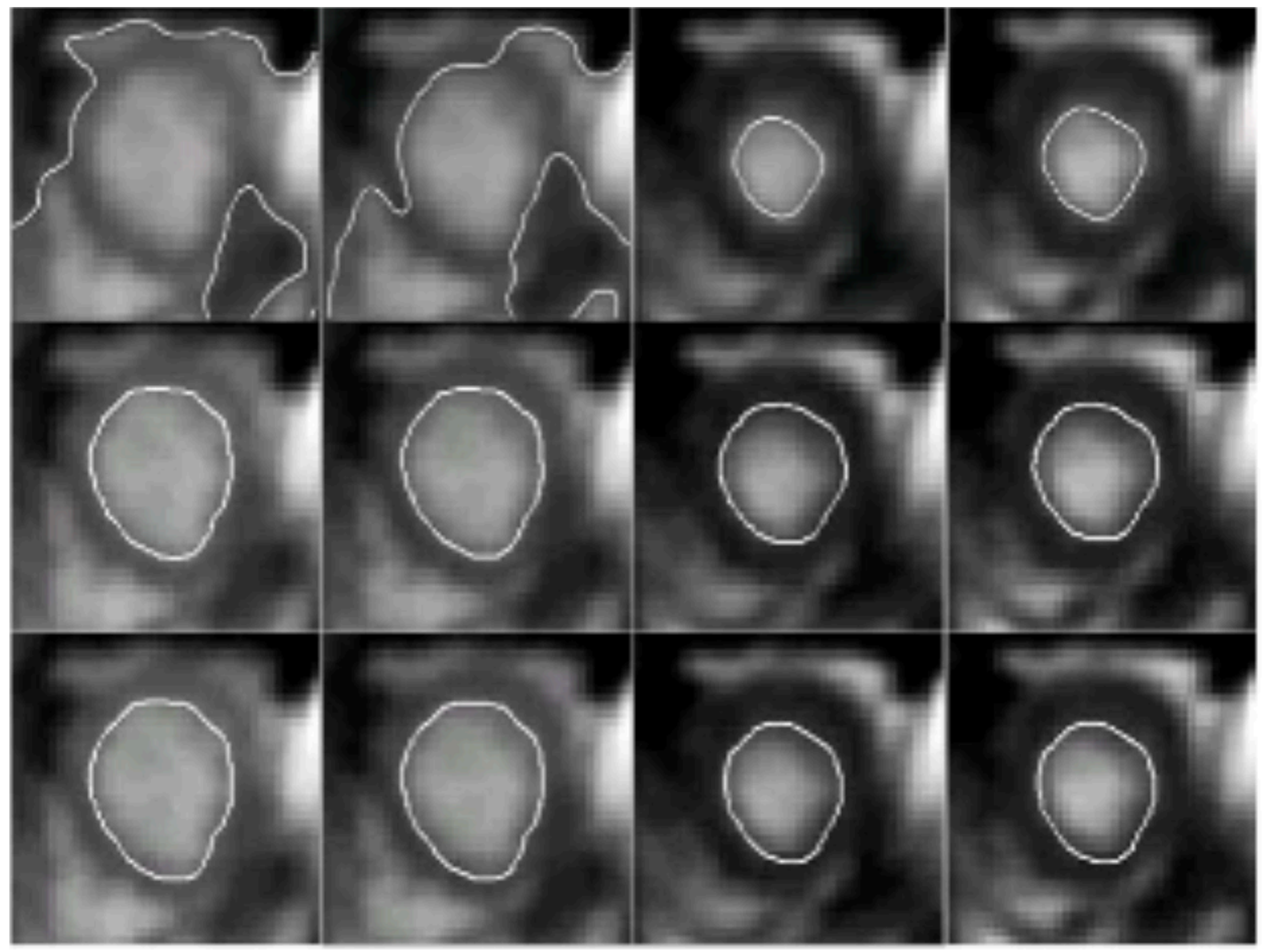


Diwale et al., https://infoscience.epfl.ch/record/265381

- Classification
- Route planning

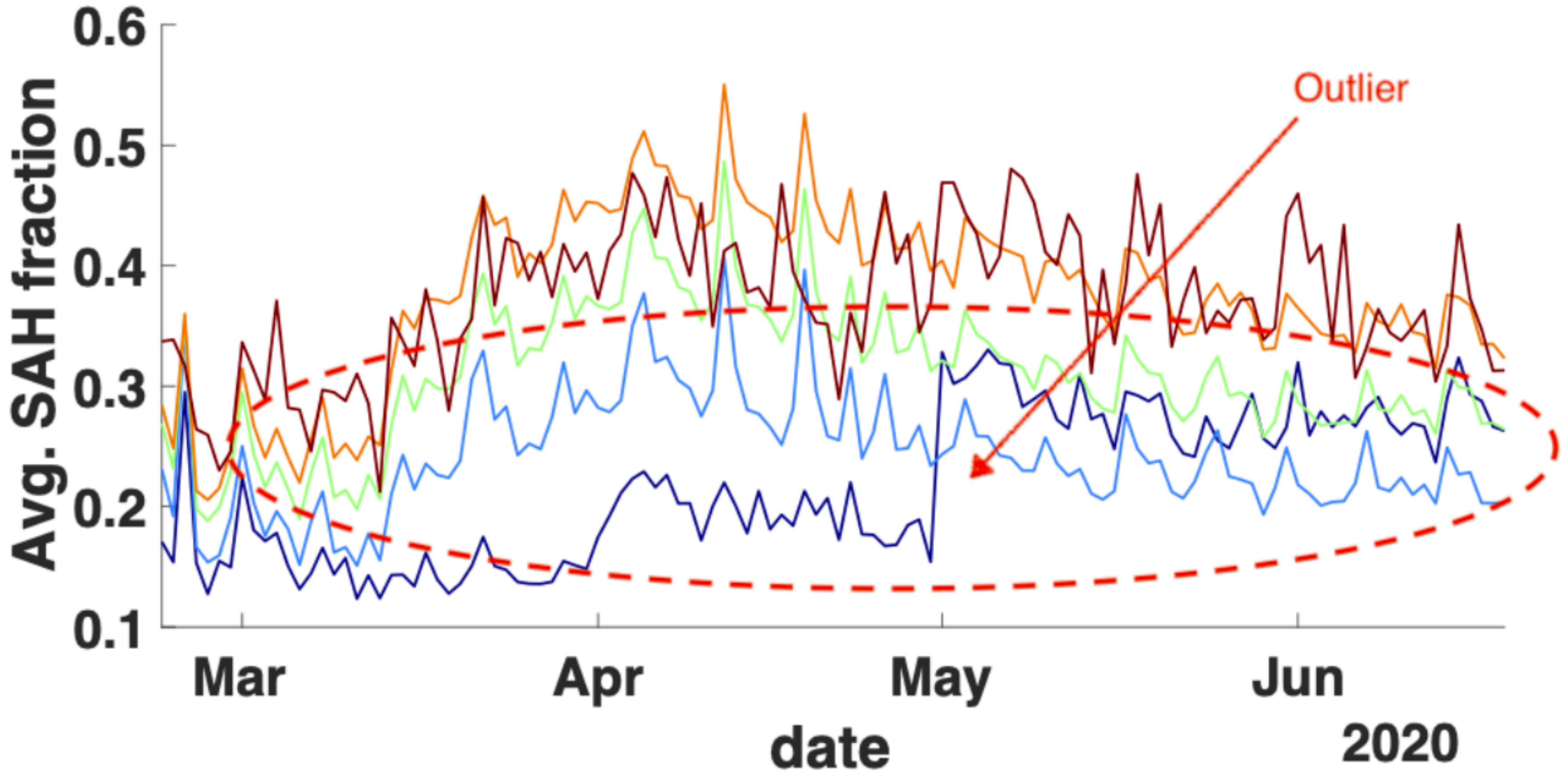
Data analysis ≈ Recovering geometry of data

Image segmentation for medical imaging:



Zhang et al., IEEE CVPR, 1092 (2006).

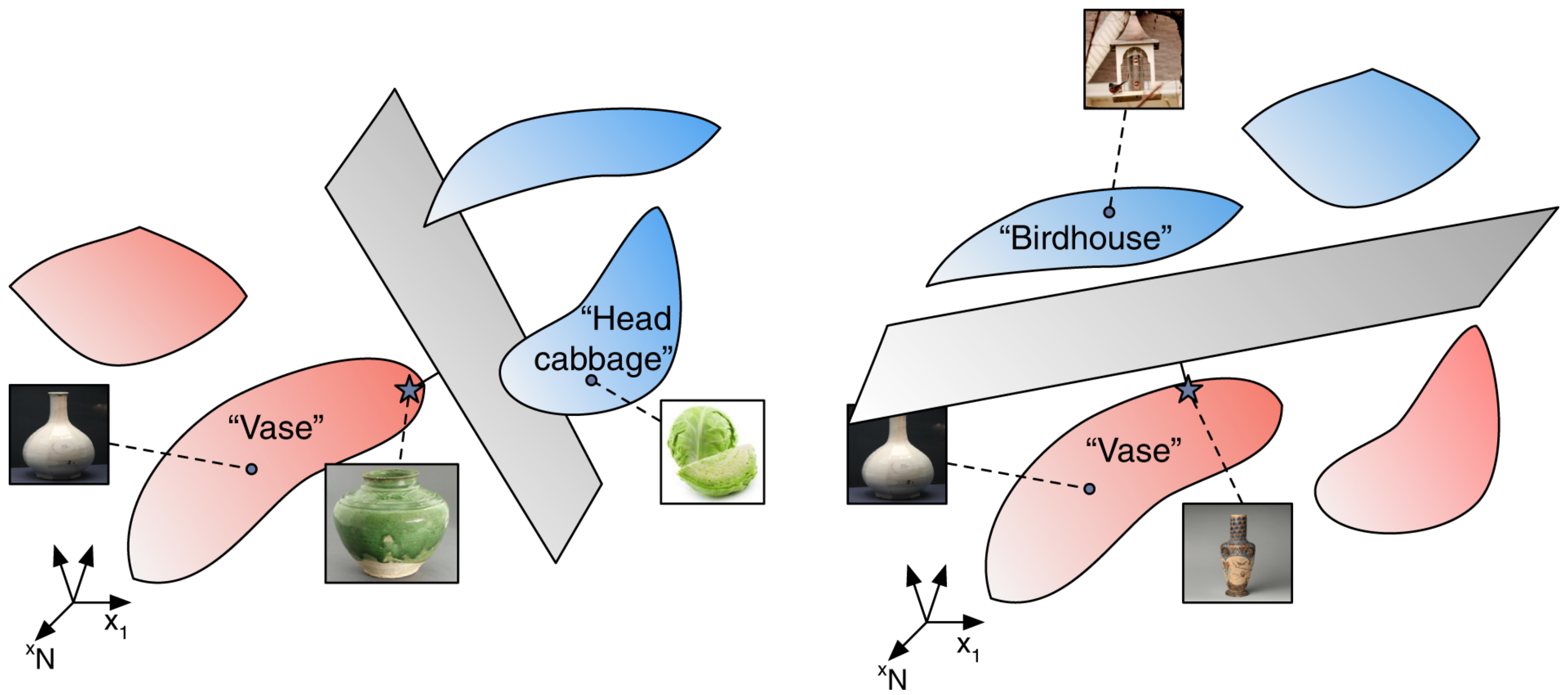
Anomaly detection in statistical dataset:



Kumar & Sarovar, arXiv (2022).

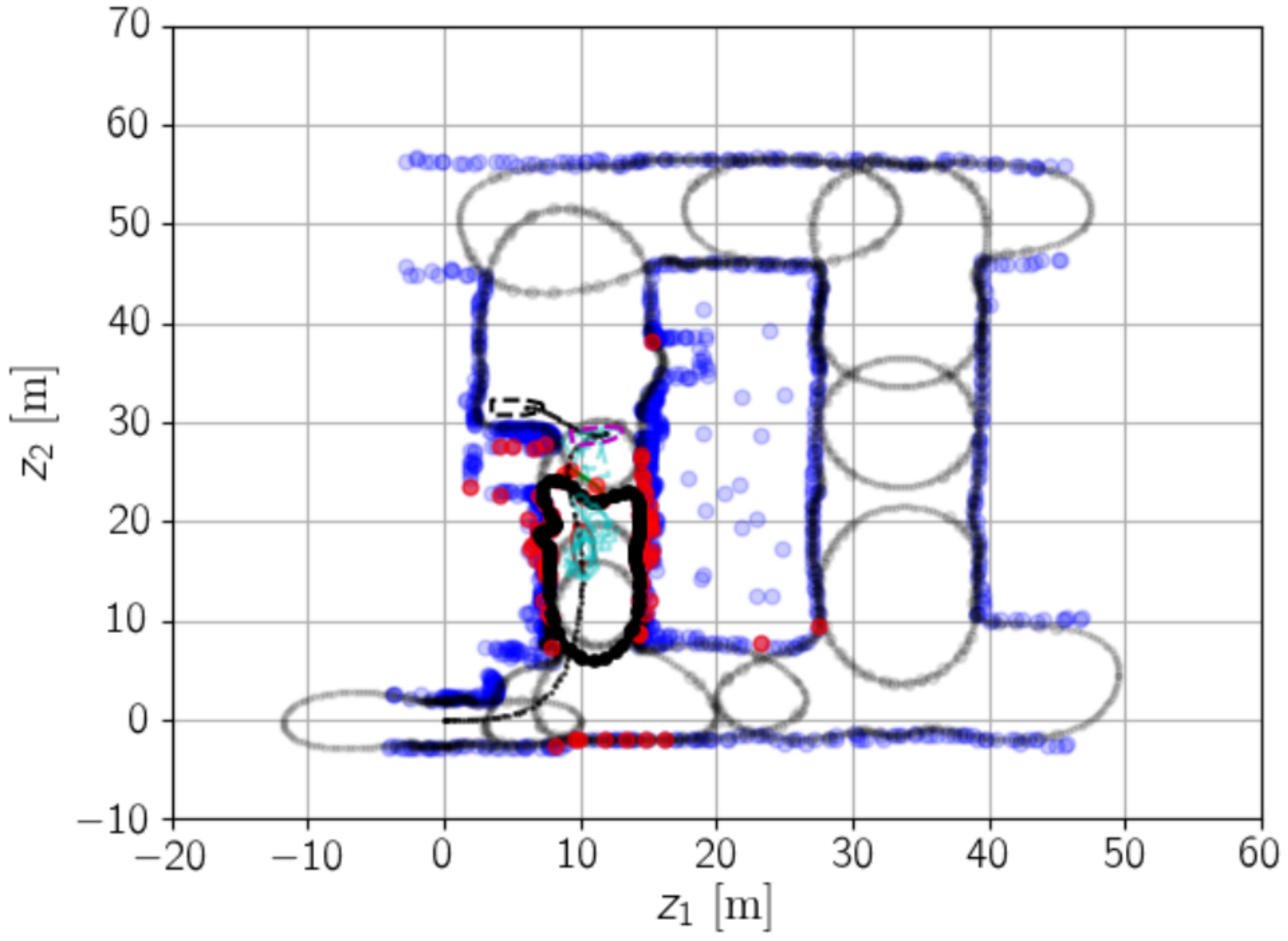
- Image segmentation
- Outlier detection

Classification within neural network's *concept space*:



Cohen et al. *Nat Commun* 11, 746 (2020).

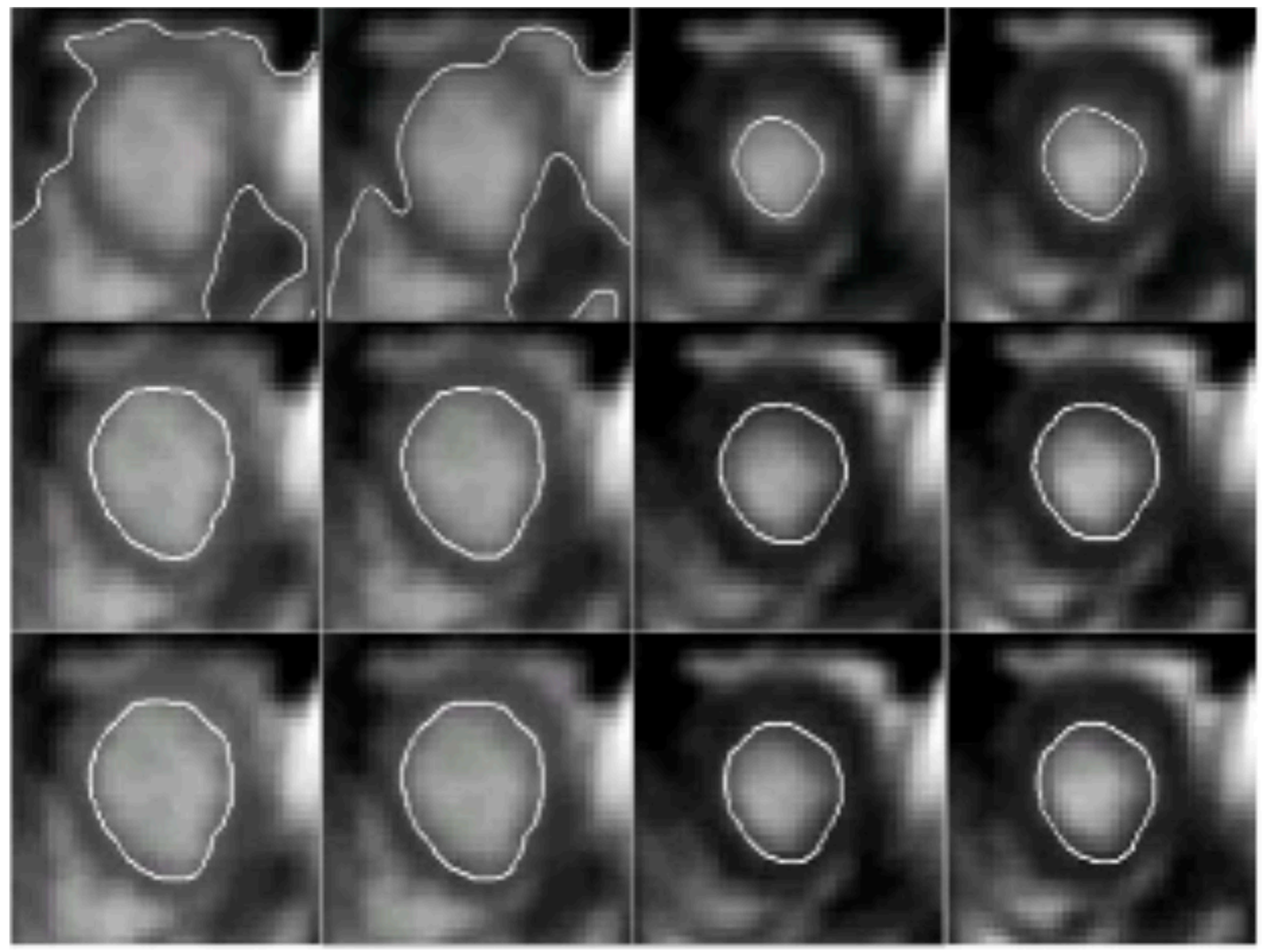
Obstacle avoidance in autonomous driving:



Diwale et al., https://infoscience.epfl.ch/record/265381

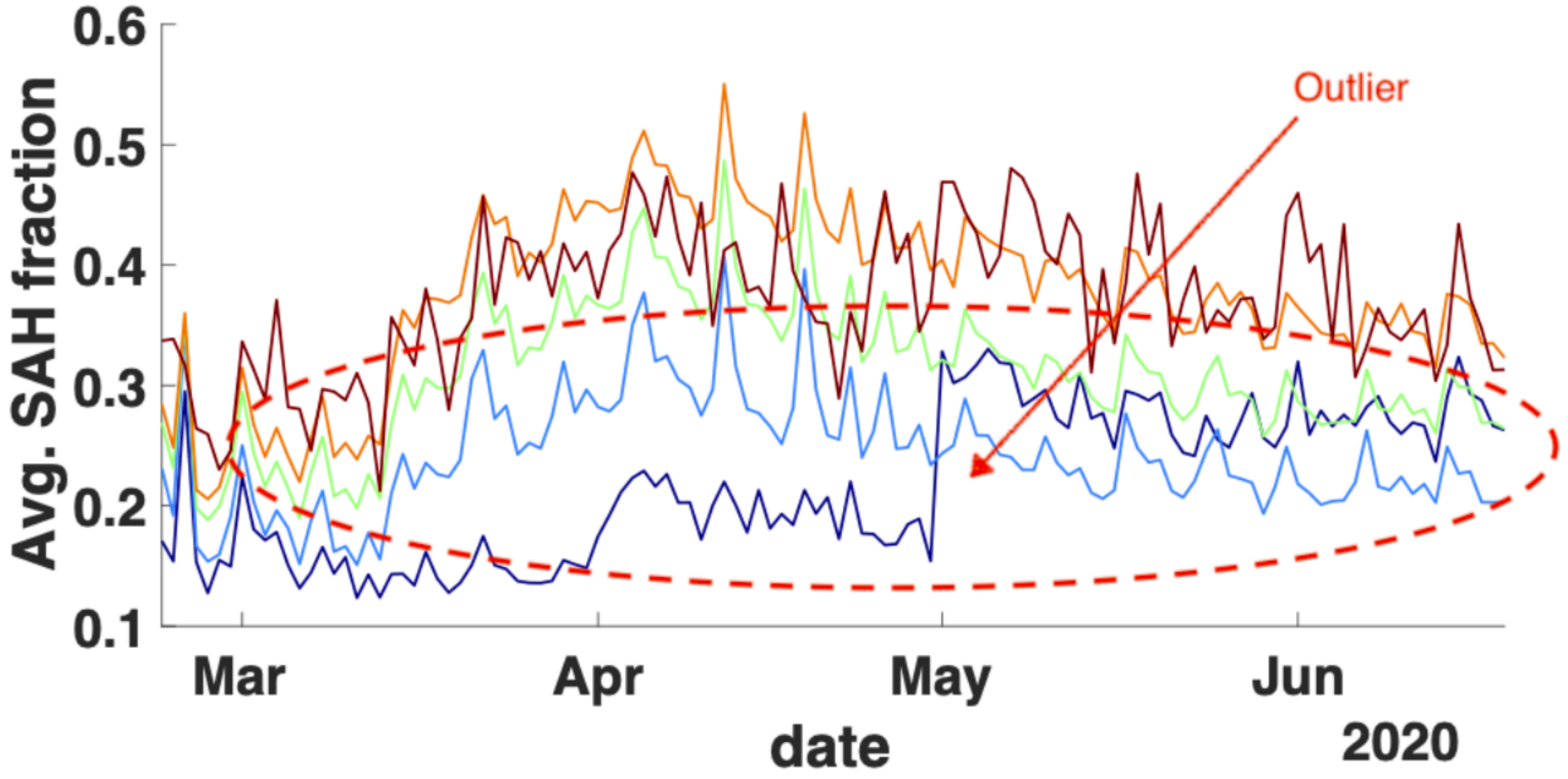
Data analysis ≈ Recovering geometry of data

Image segmentation for medical imaging:



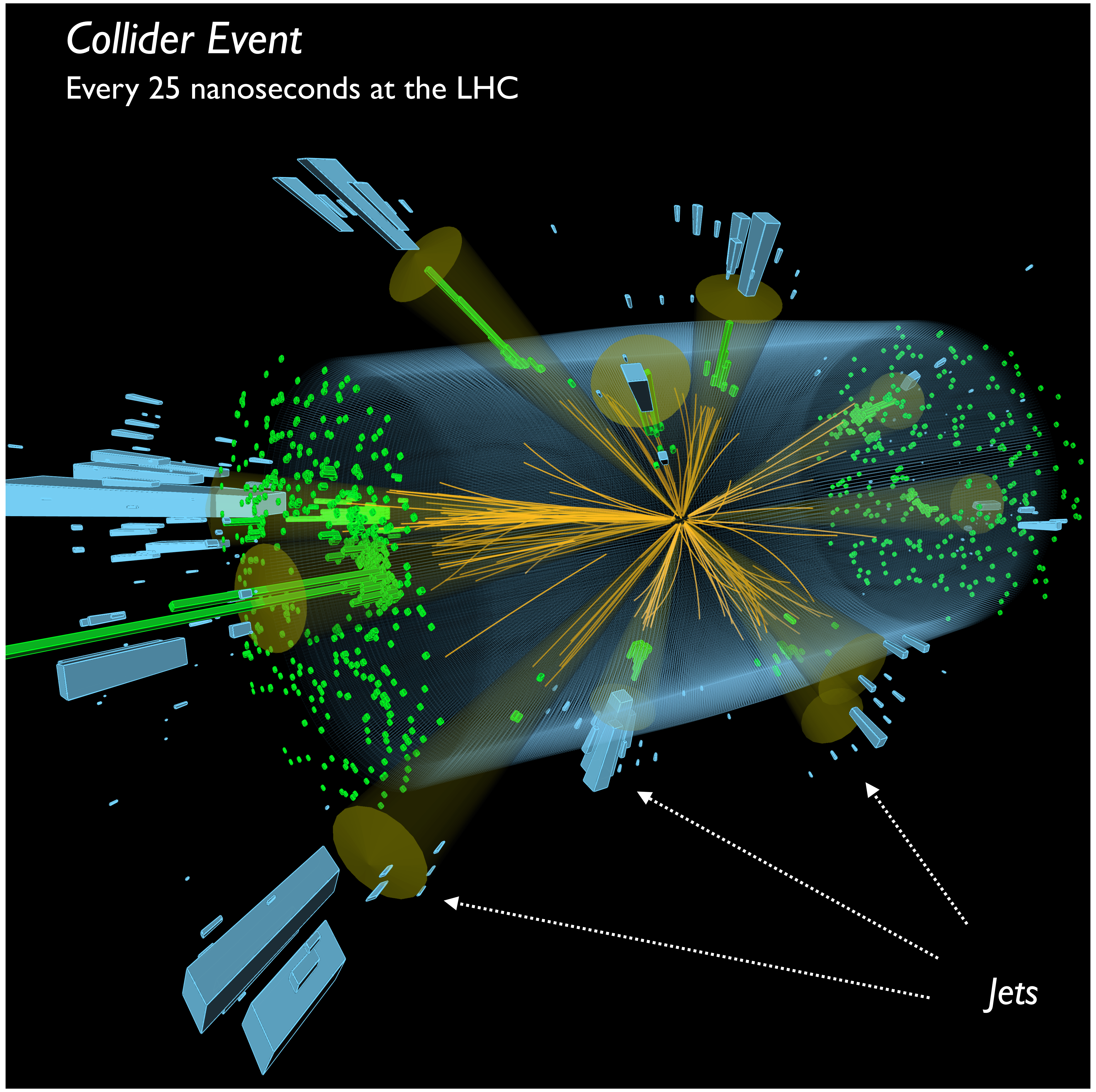
Zhang et al., IEEE CVPR, 1092 (2006).

Anomaly detection in statistical dataset:



Kumar & Sarovar, arXiv (2022).

- Classification
- Route planning
- Dimensionality reduction, clustering, augmented reality, regularization of learning methods, visualization, …
- Image segmentation
- Outlier detection

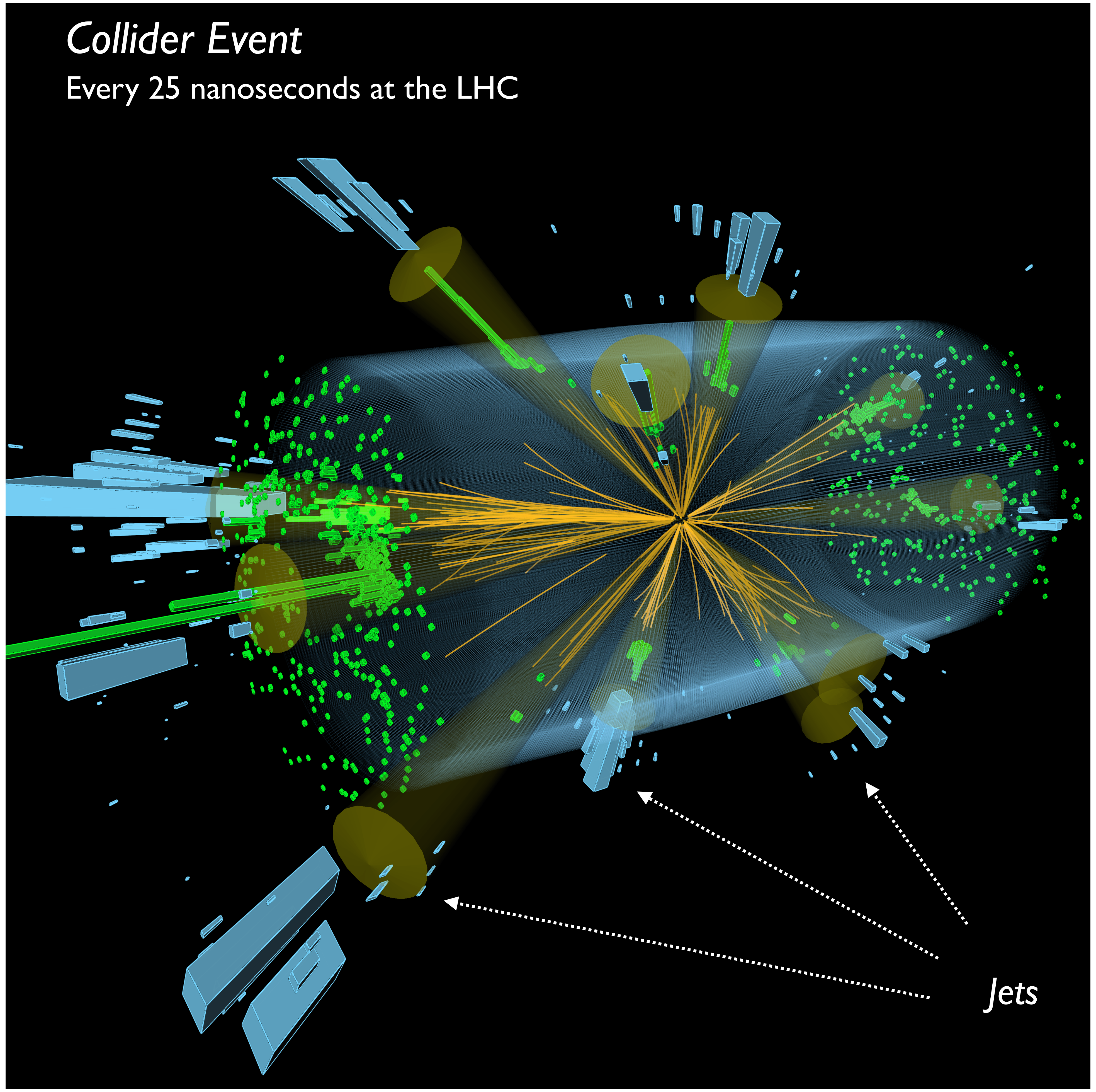


Collider data has geometric structure

Komiske, Metodiev, Thaler, **The hidden geometry of particle collisions**, J. HEP, 6 (2020)

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure

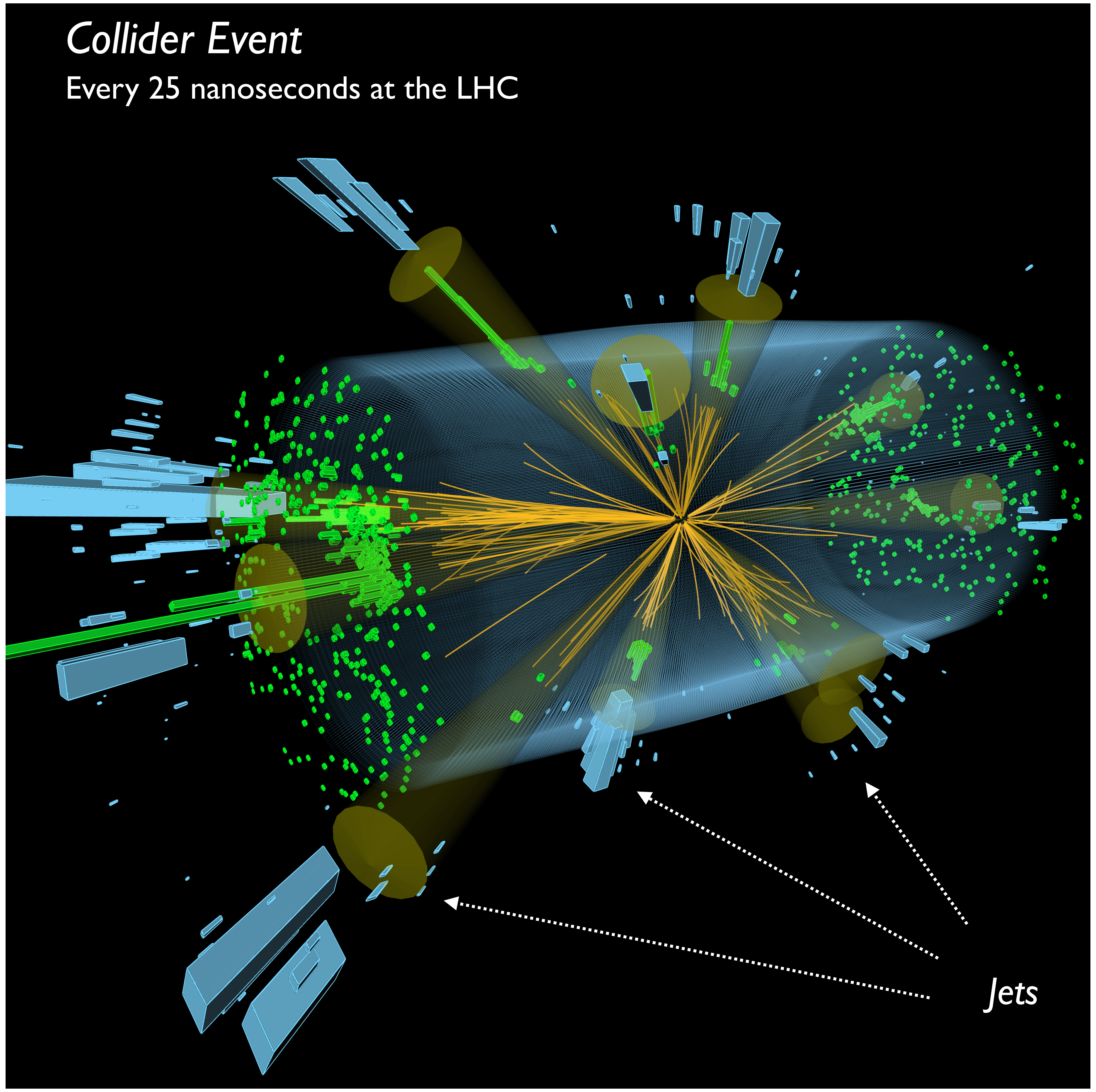


Energy flow (or Event):

$$\mathcal{E}(\hat{n}) = \sum_{j=1}^{N} E_j \delta(\hat{n} - \hat{n}_j)$$

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure



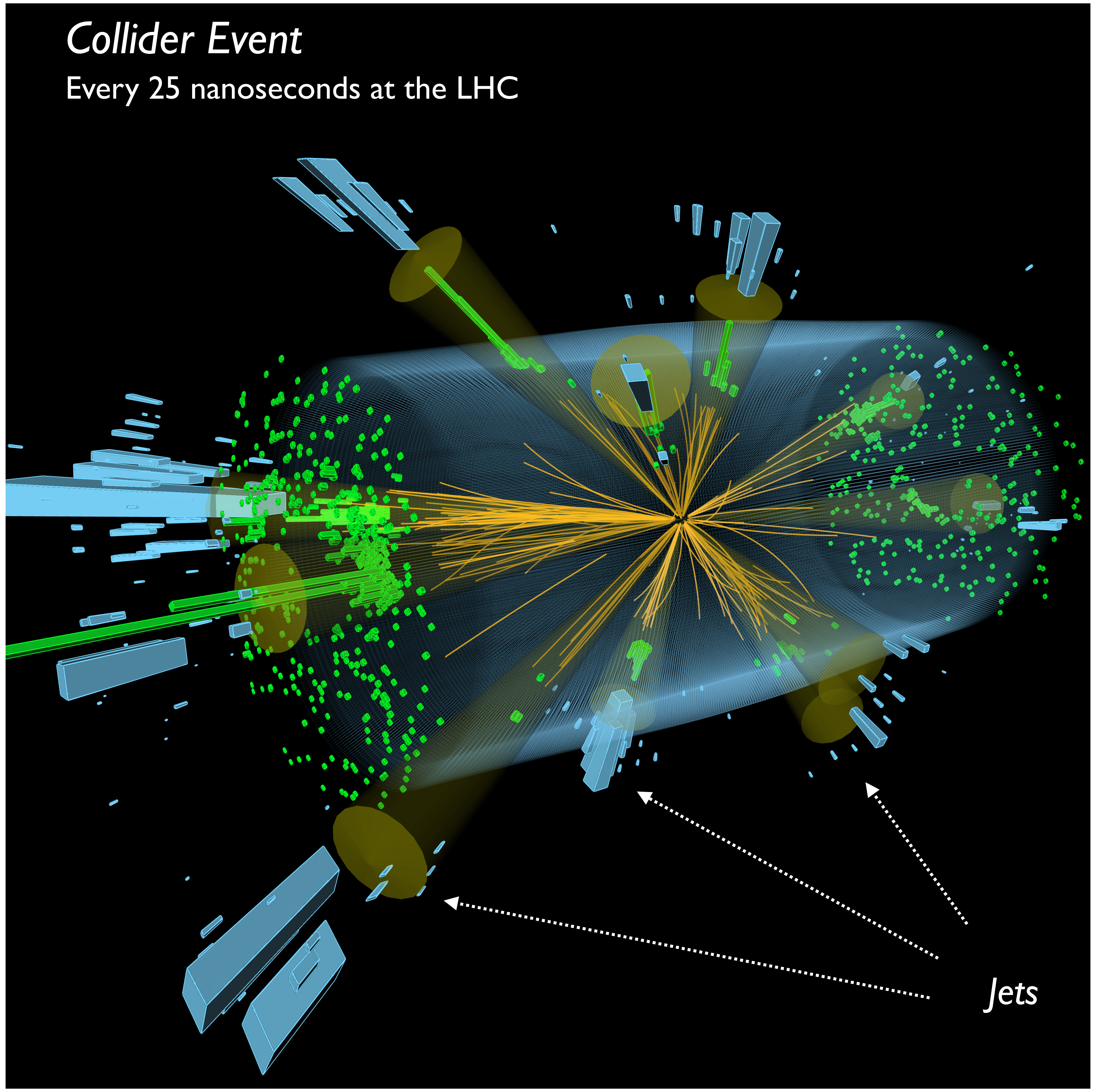
#particles

Energy flow (or Event):

$$\mathcal{E}(\hat{n}) = \sum_{j=1}^{N} E_j \delta(\hat{n} - \hat{n}_j)$$

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure



#particles

energies*

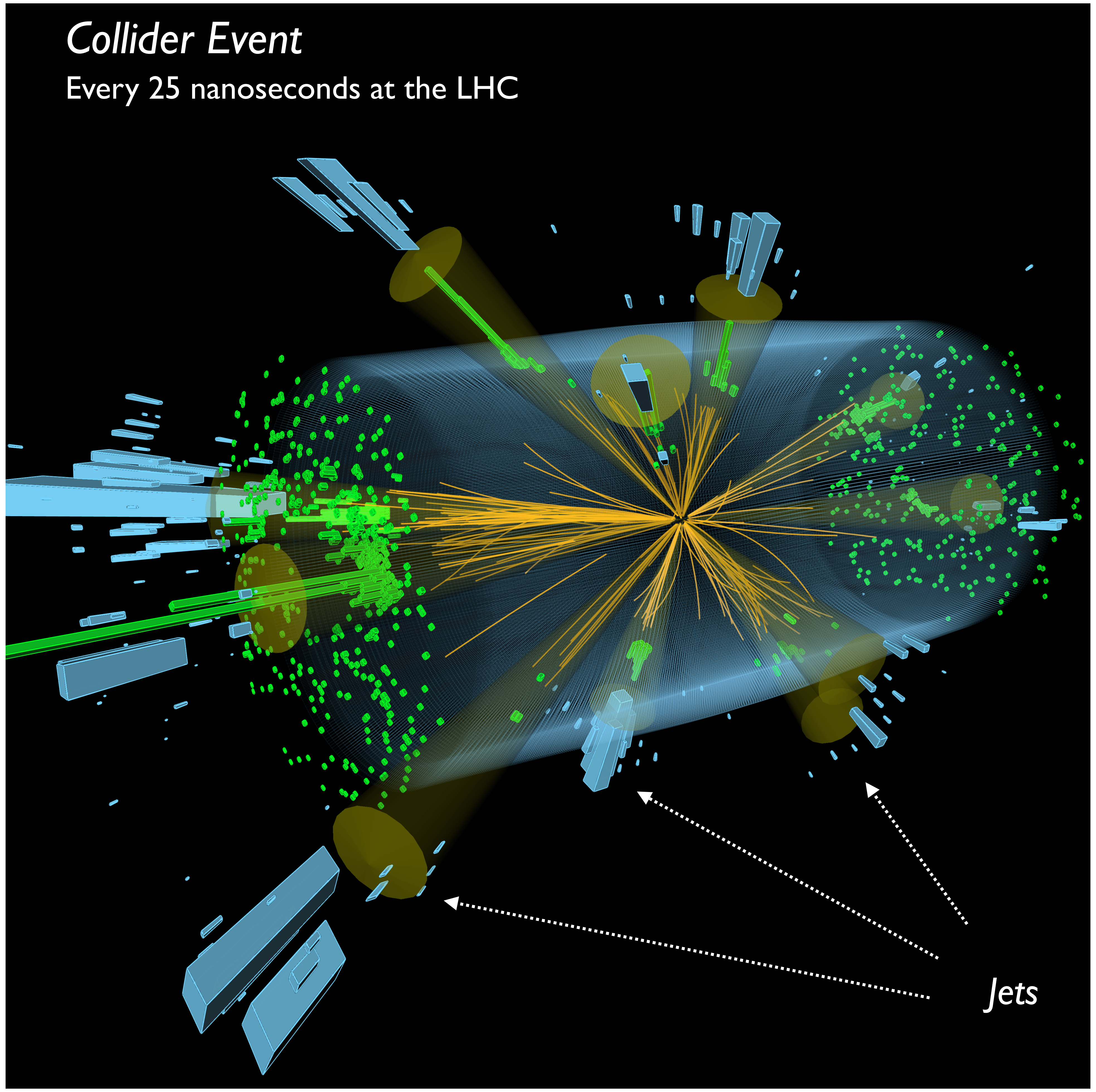
Energy flow (or Event):

$$\mathcal{E}(\hat{n}) = \sum_{j=1}^{N} E_j \delta(\hat{n} - \hat{n}_j)$$

* or particle transverse momenta

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure



Energy flow (or Event):

$$\mathcal{E}(\hat{n}) = \sum_{j=1}^{N} E_j \delta(\hat{n} - \hat{n}_j)$$

#particles

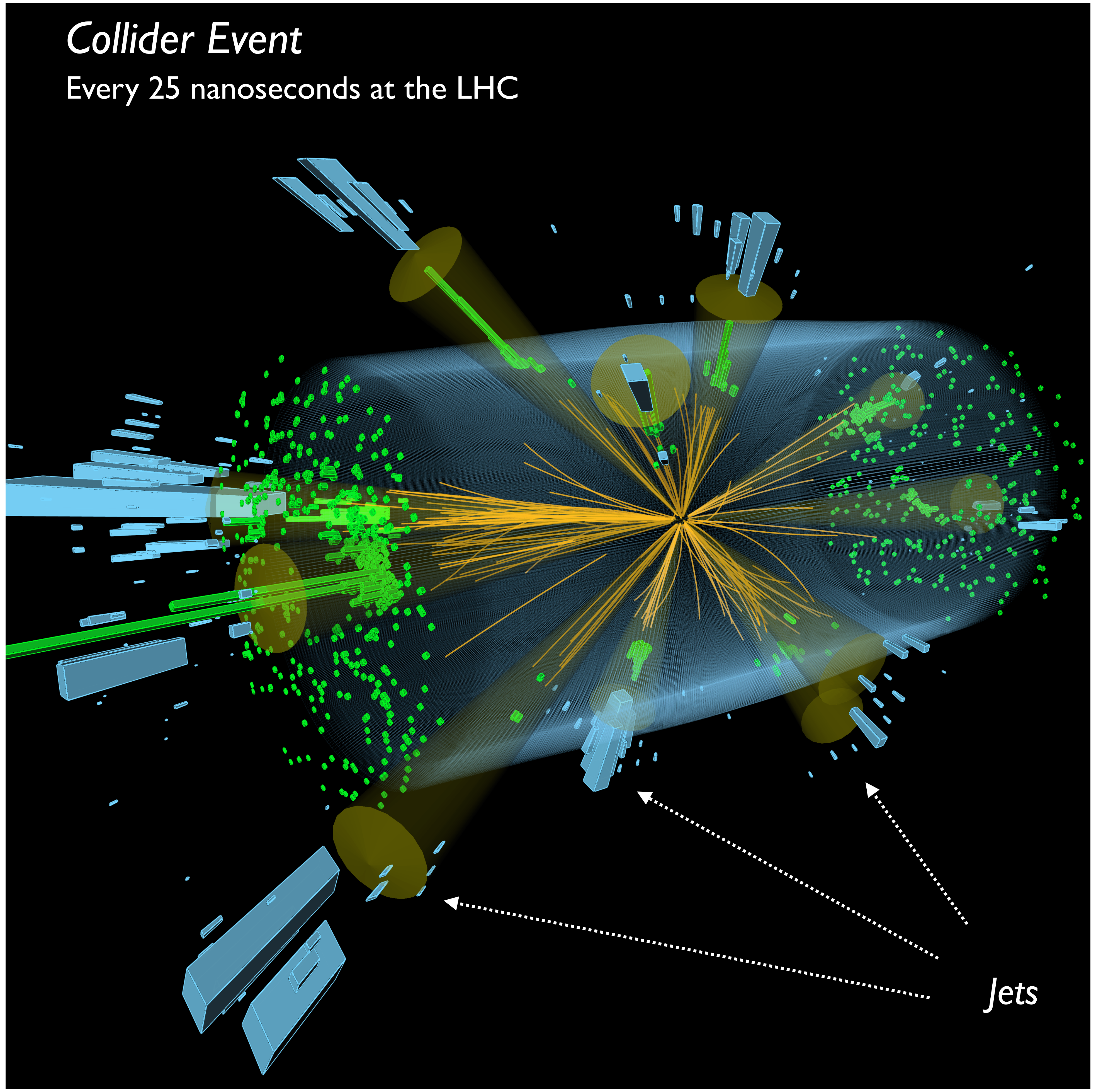
energies*

∠directions

* or particle transverse momenta

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science:
https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure



#particles

energies*

Energy flow (or Event):

$$\mathcal{E}(\hat{n}) = \sum_{j=1}^{N} E_j \delta(\hat{n} - \hat{n}_j)$$

∠directions

$$\wr\!\wr$$

Energy-weighted point cloud:

$$(E_j, \hat{n}_{x,j}, \hat{n}_{y,j}, \hat{n}_{z,j})_{j=1}^{N}$$

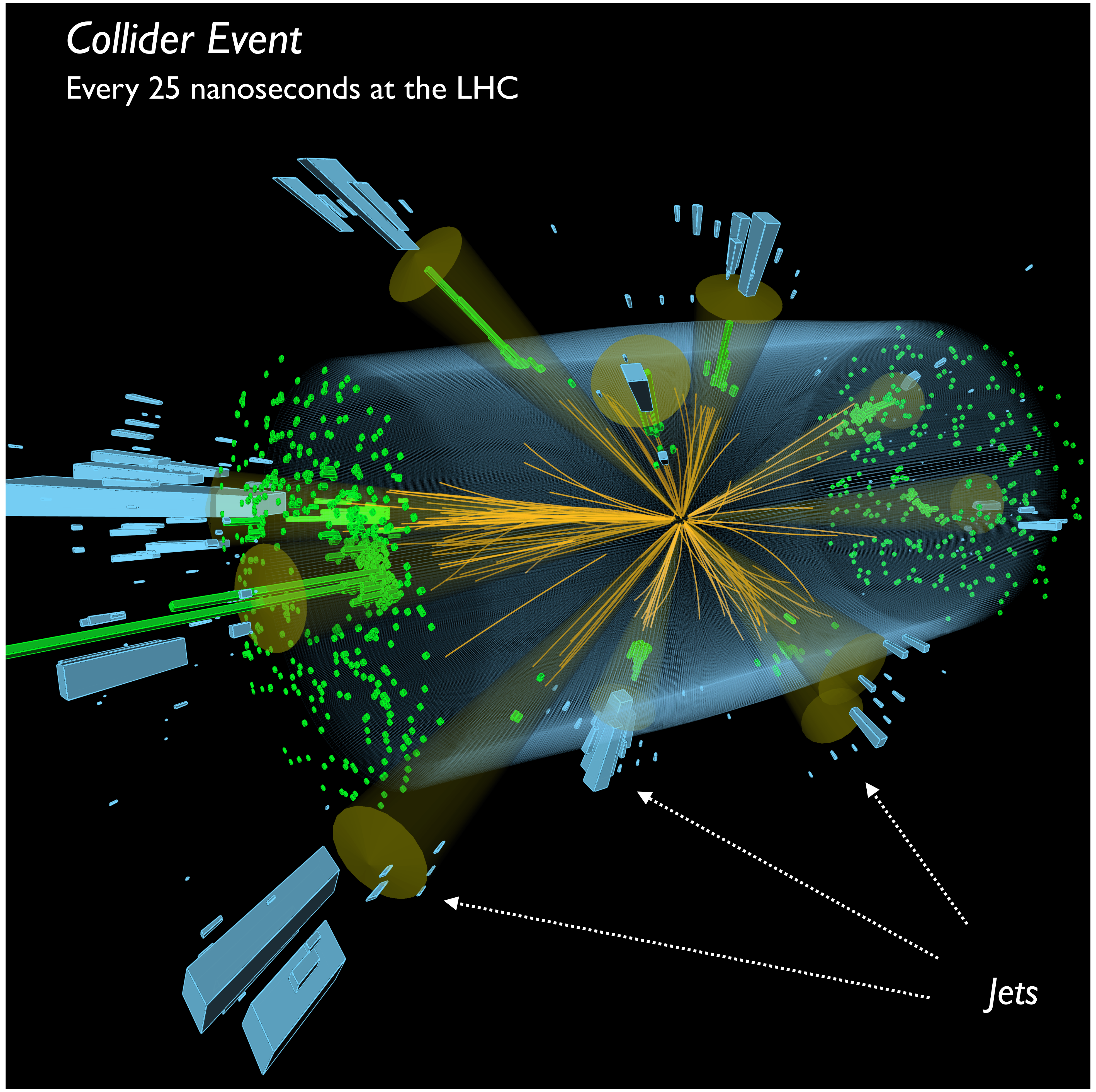
$$\cap$$

$$\mathbb{R}_+ \times \mathbb{S}^2$$

* or particle transverse momenta

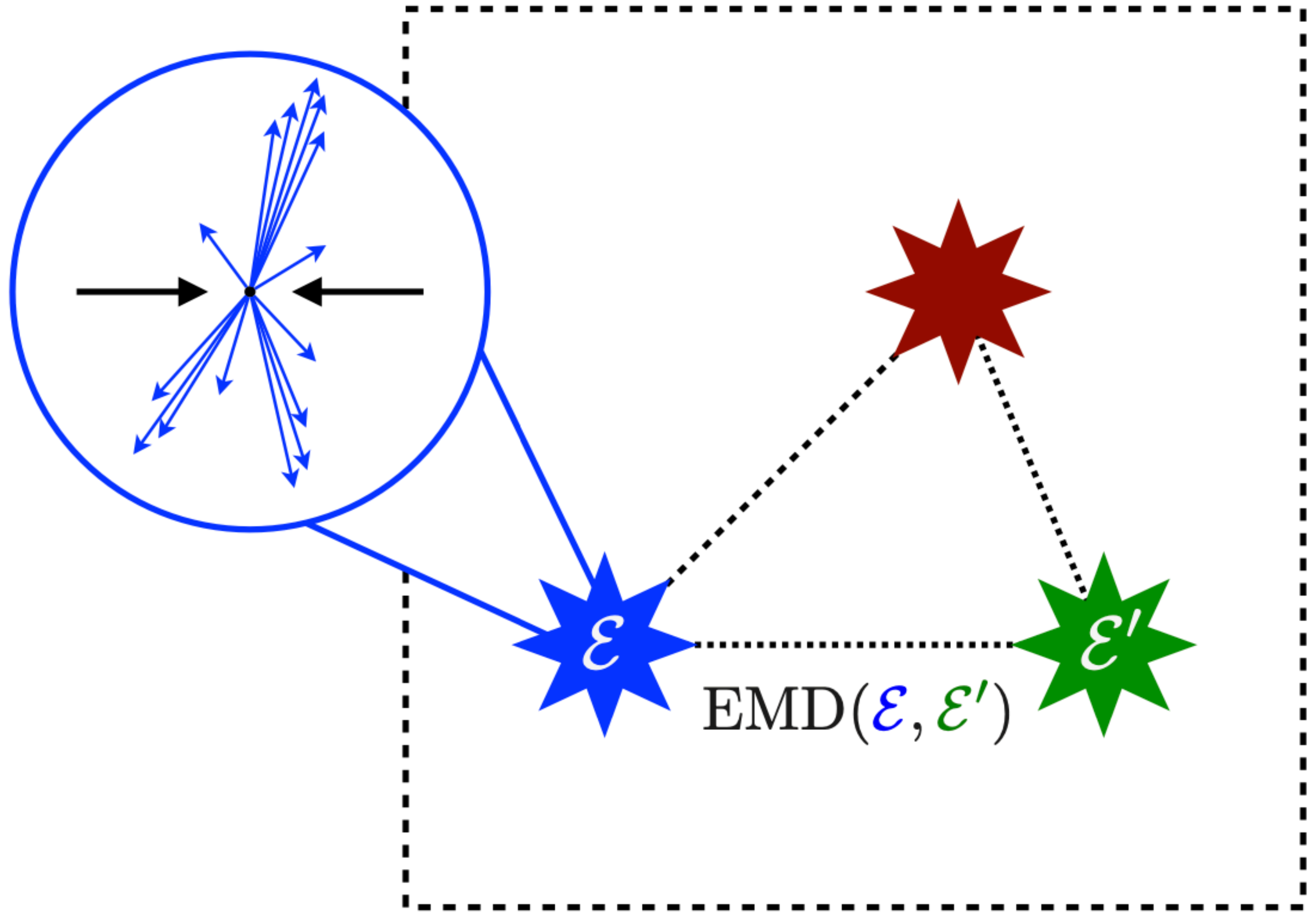
Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science:
https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure

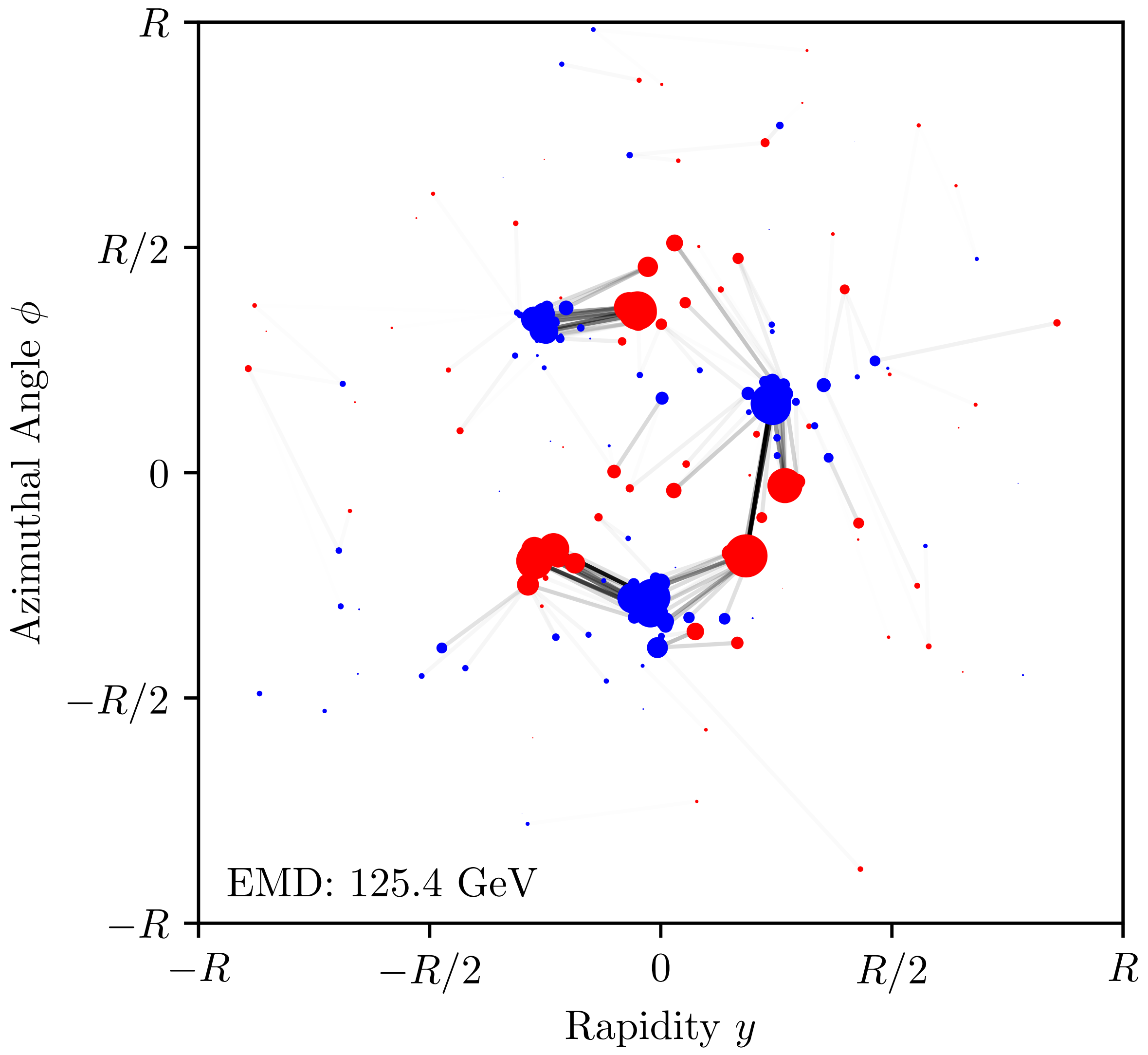


$$\text{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\frac{\text{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i - \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events

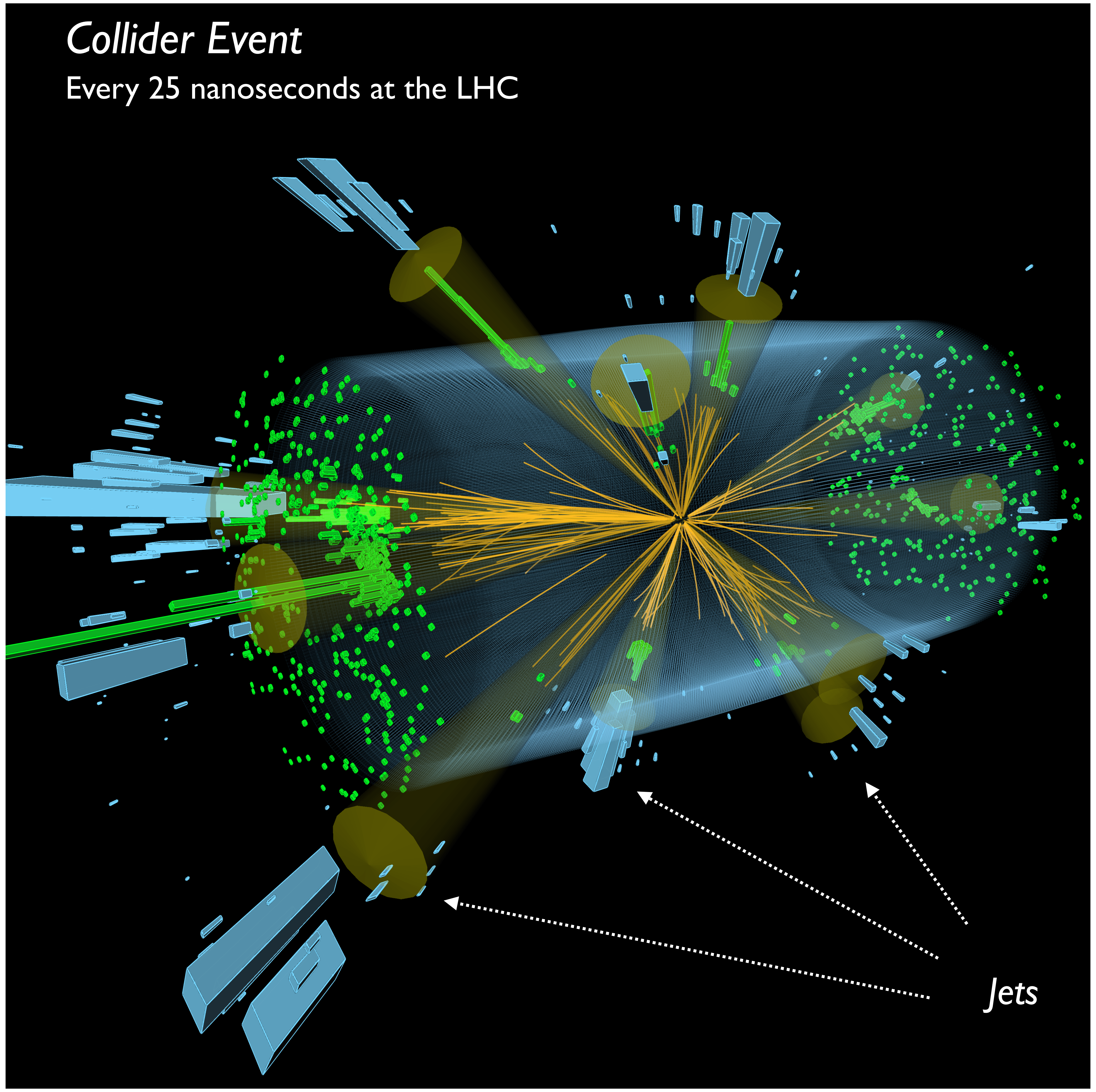


$\mathcal{E} \longrightarrow \mathcal{E}'$



Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science:
https://www.youtube.com/watch?v=BMBSAWUxBn4

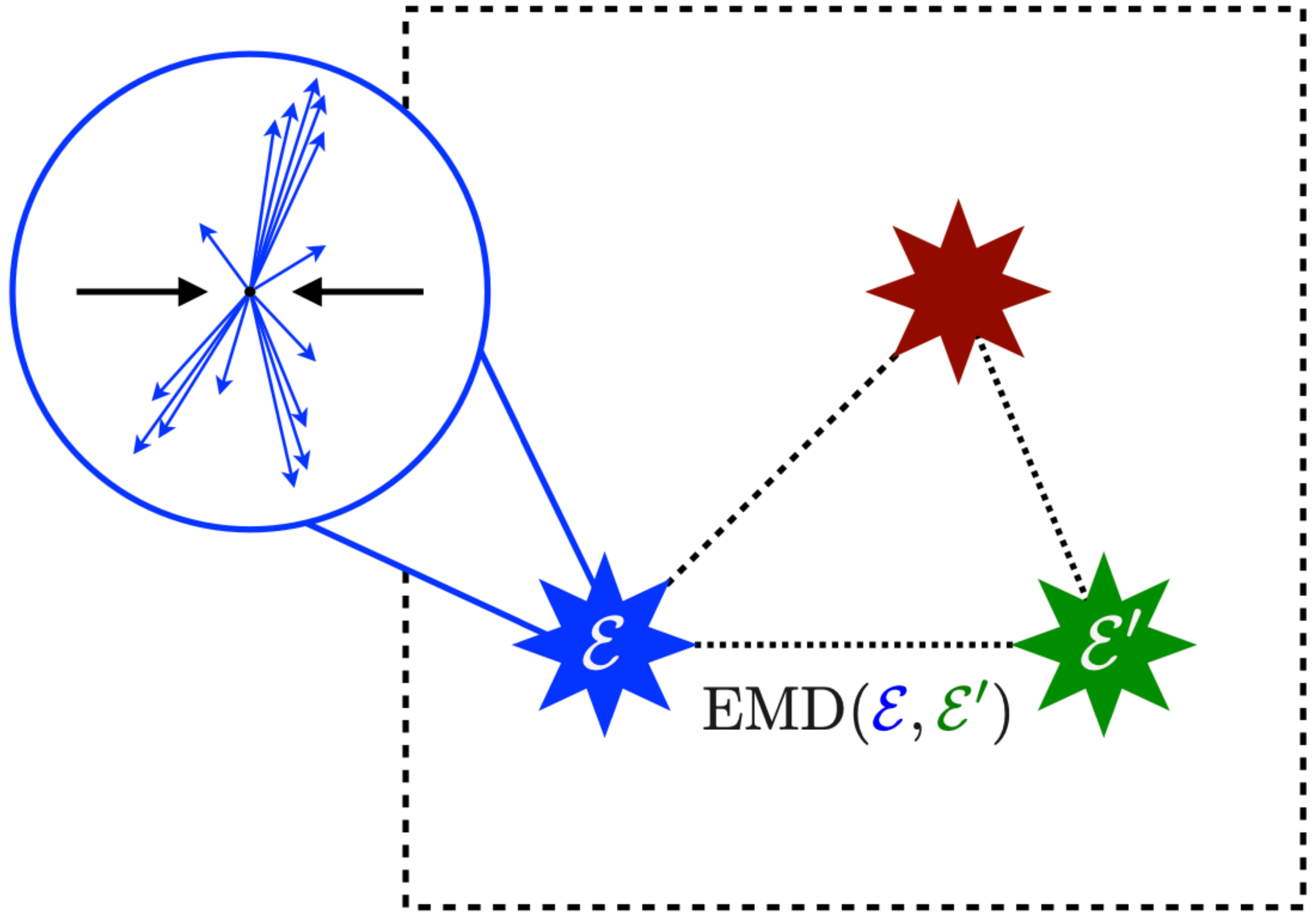
Collider data has geometric structure

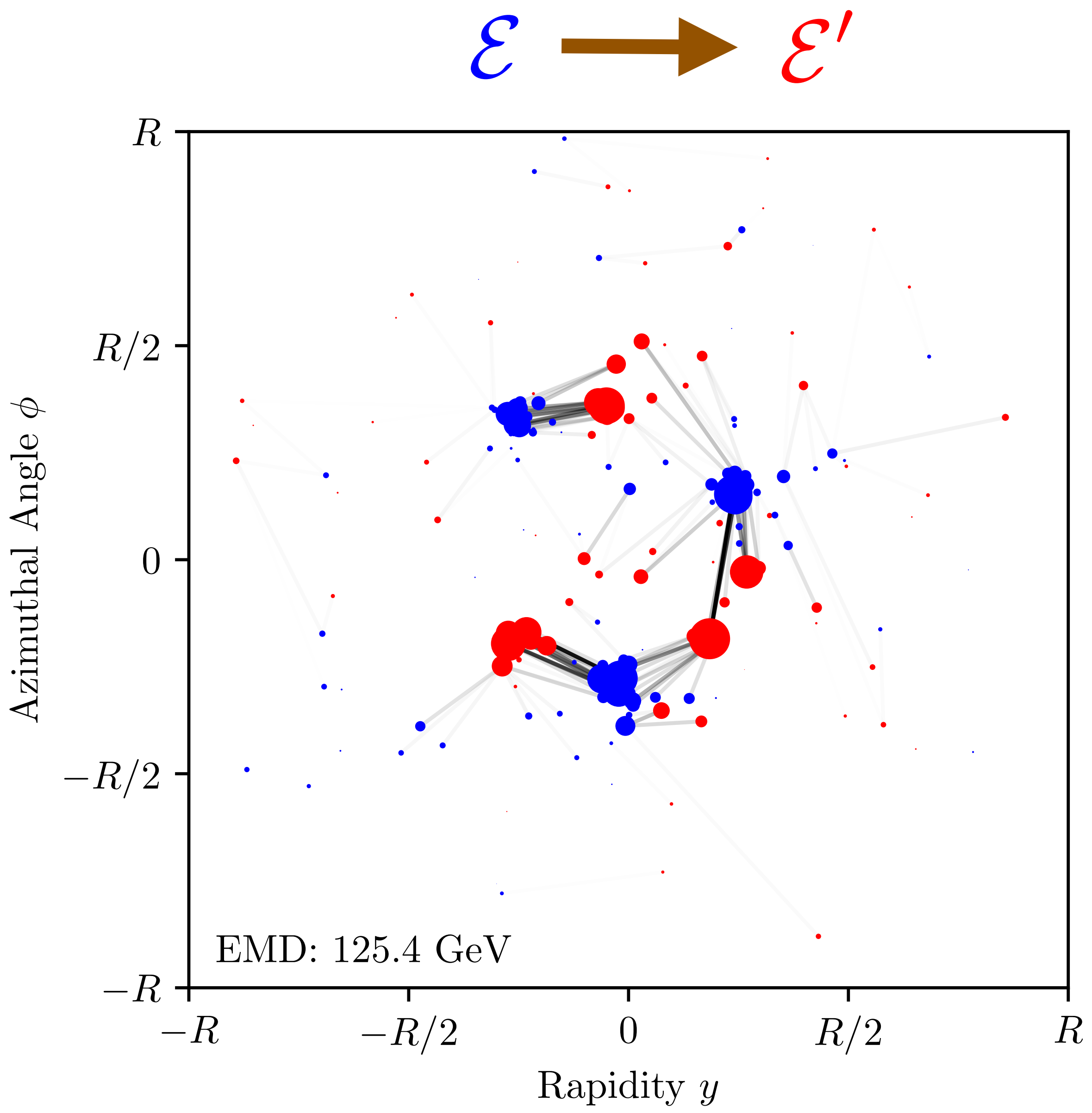


energy (GeV)

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i - \sum_{j=1}^{N'} E_j\right|$$

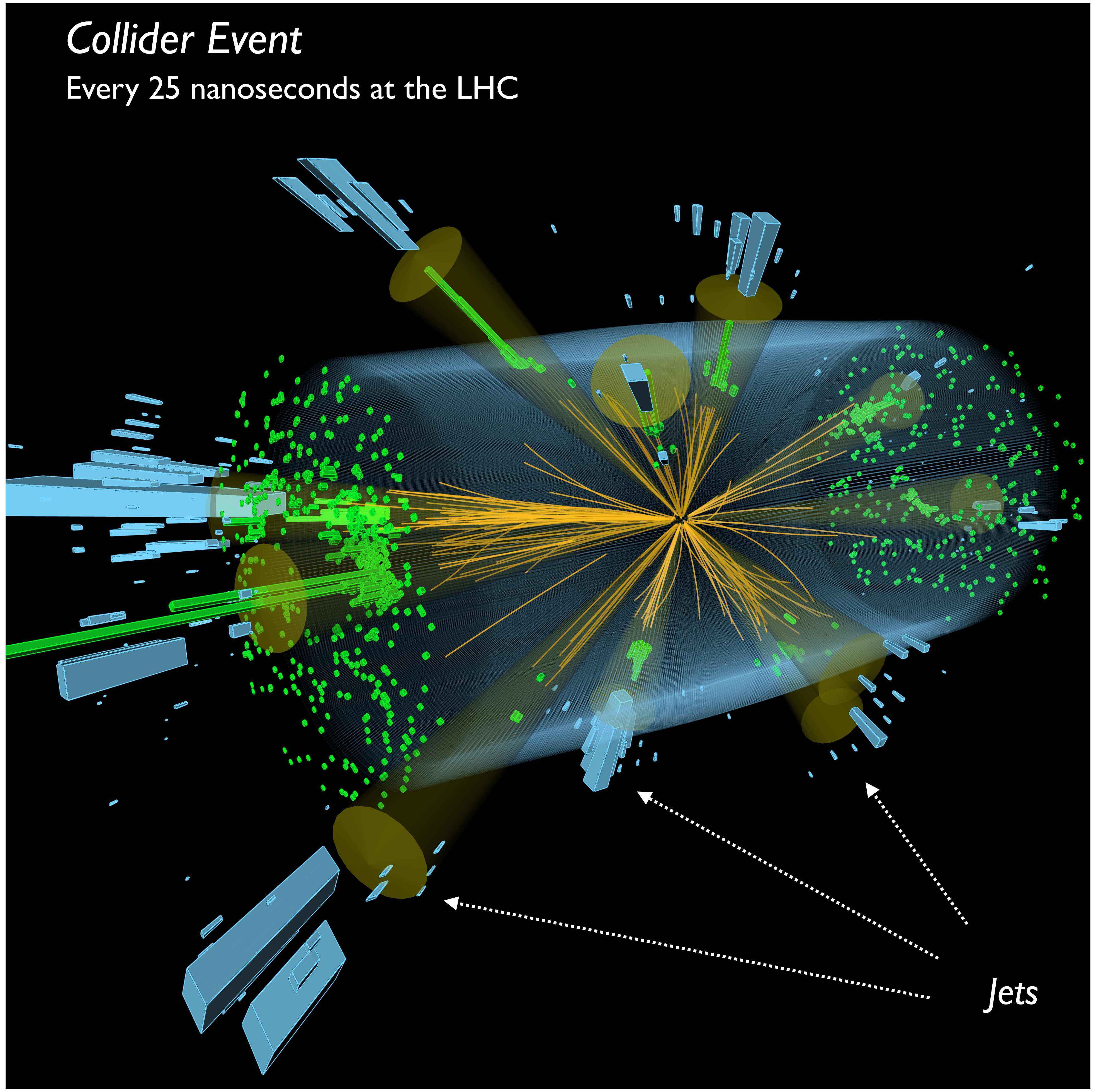
Metric* space of possible events





Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science:
https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure

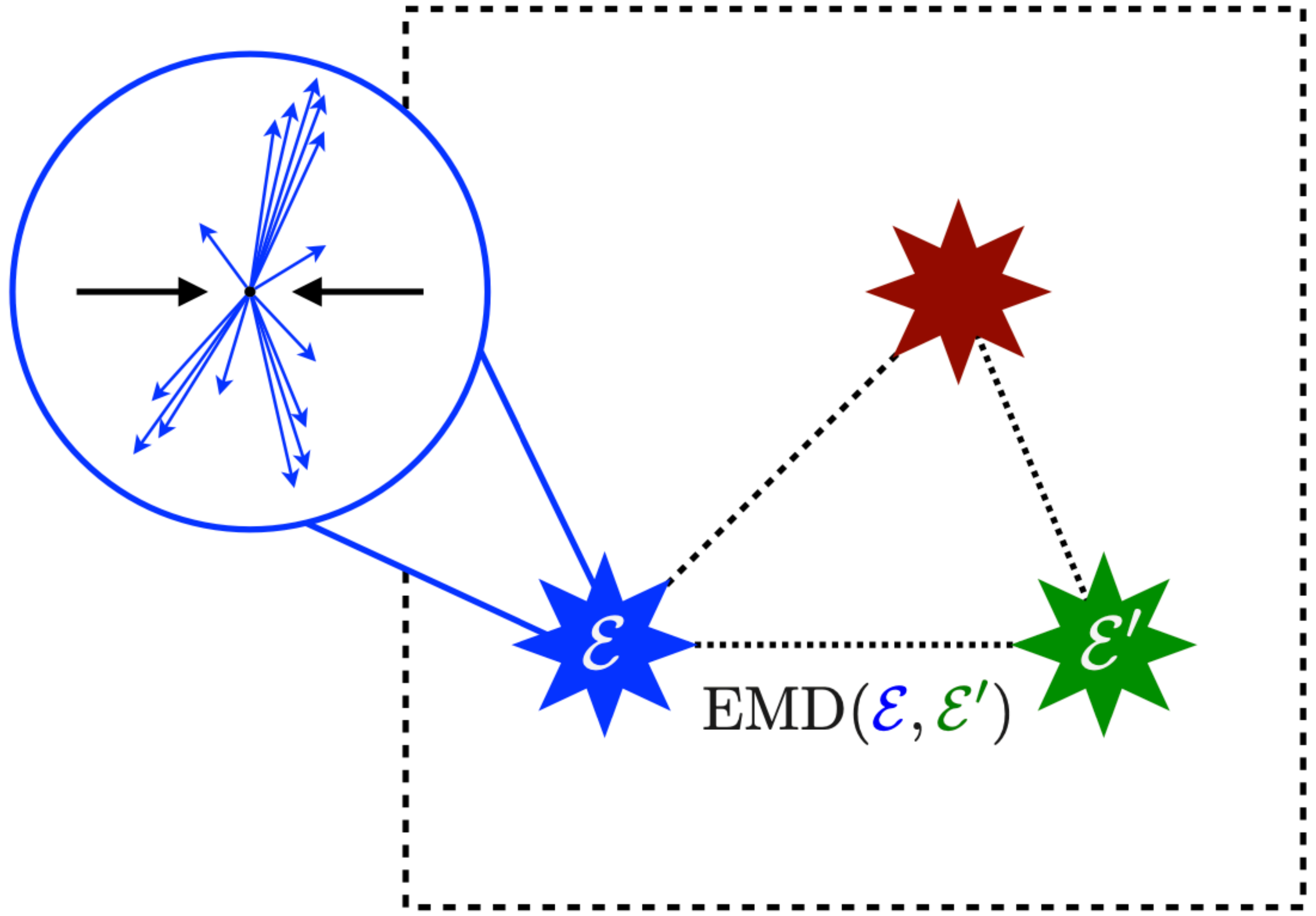


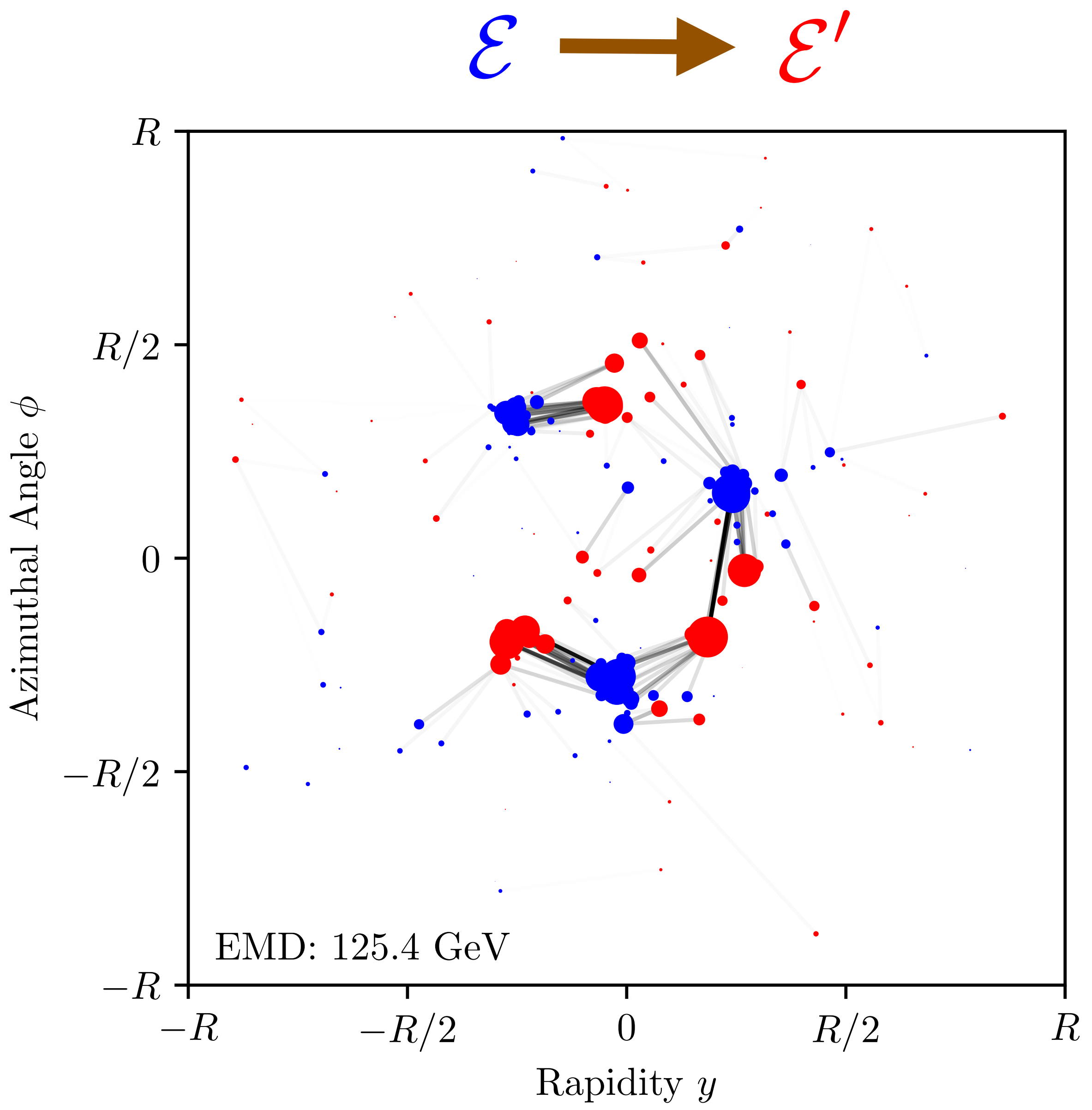
energy (GeV)

cost to move energy

$$\text{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\frac{\text{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i - \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events

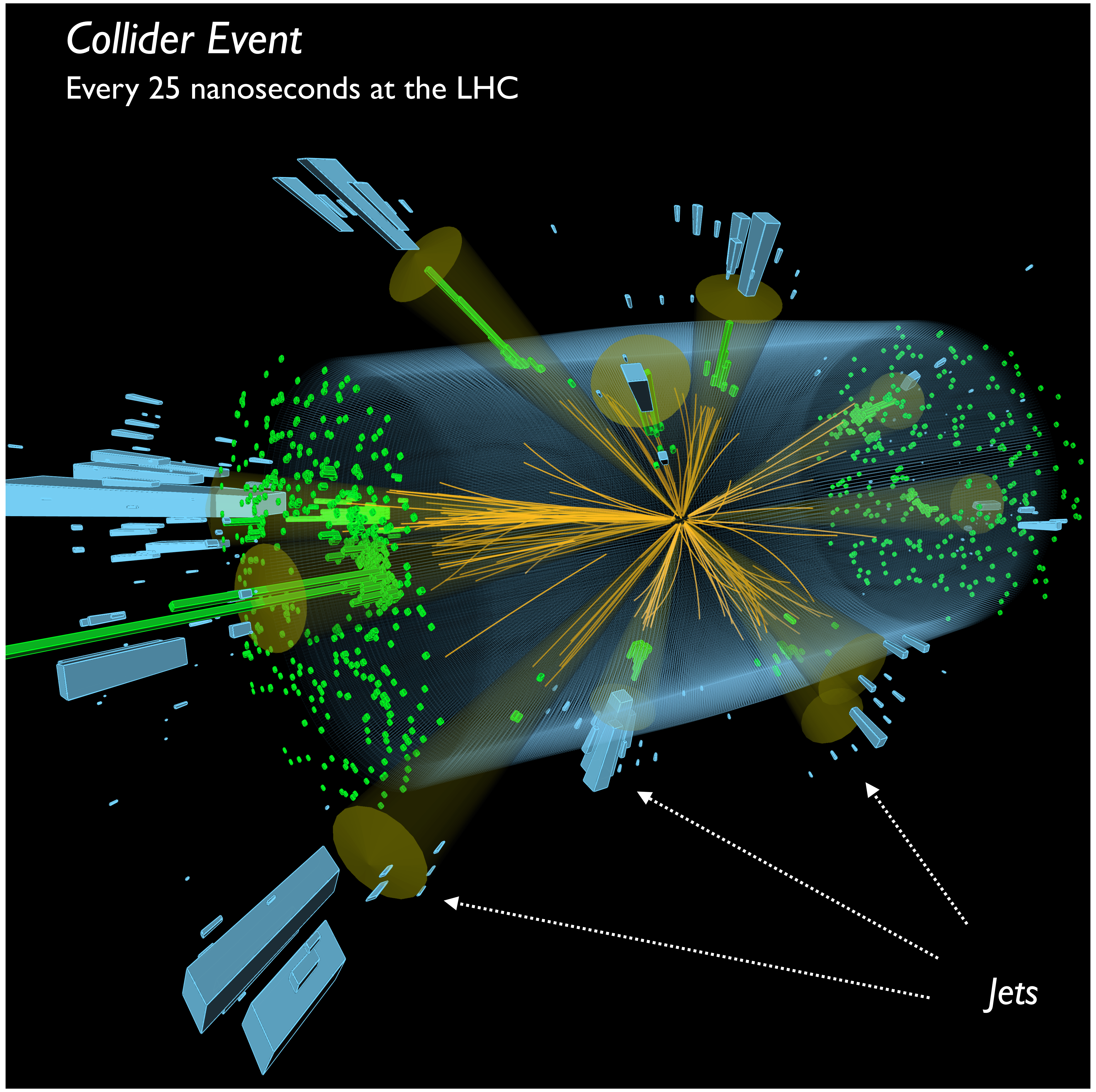




* for R sufficiently large
i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure



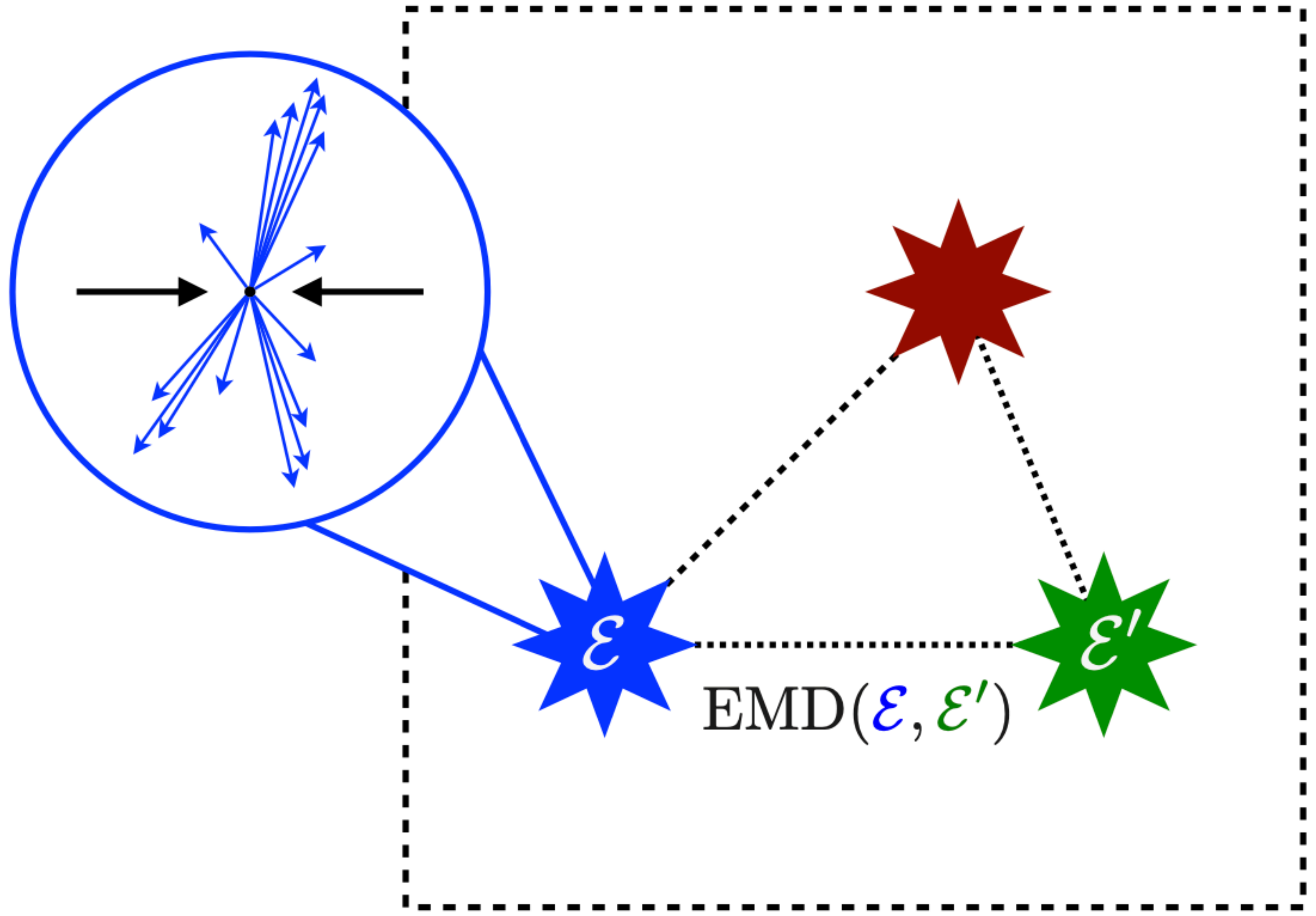
energy (GeV)

cost to move energy

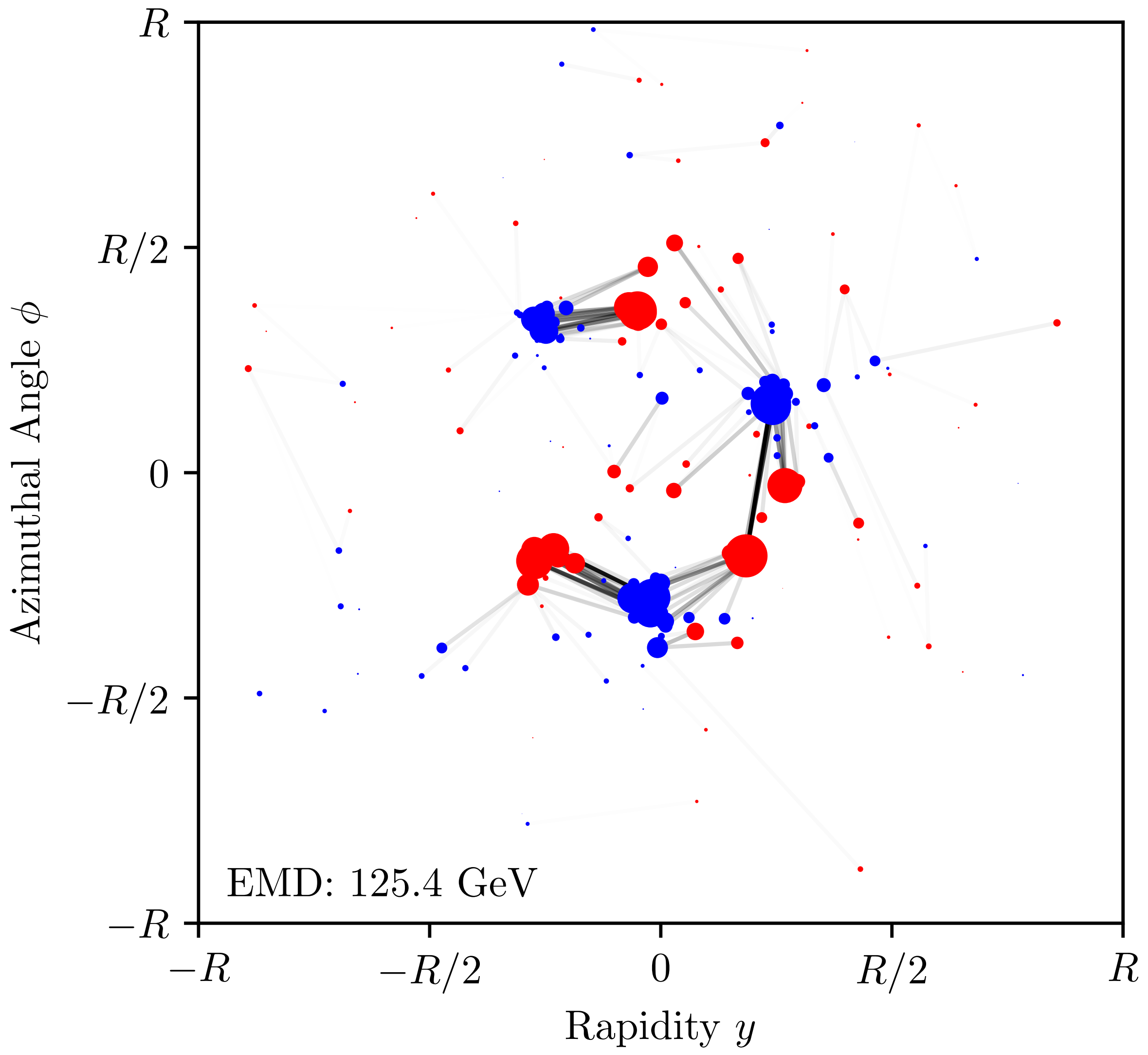
cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i - \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events



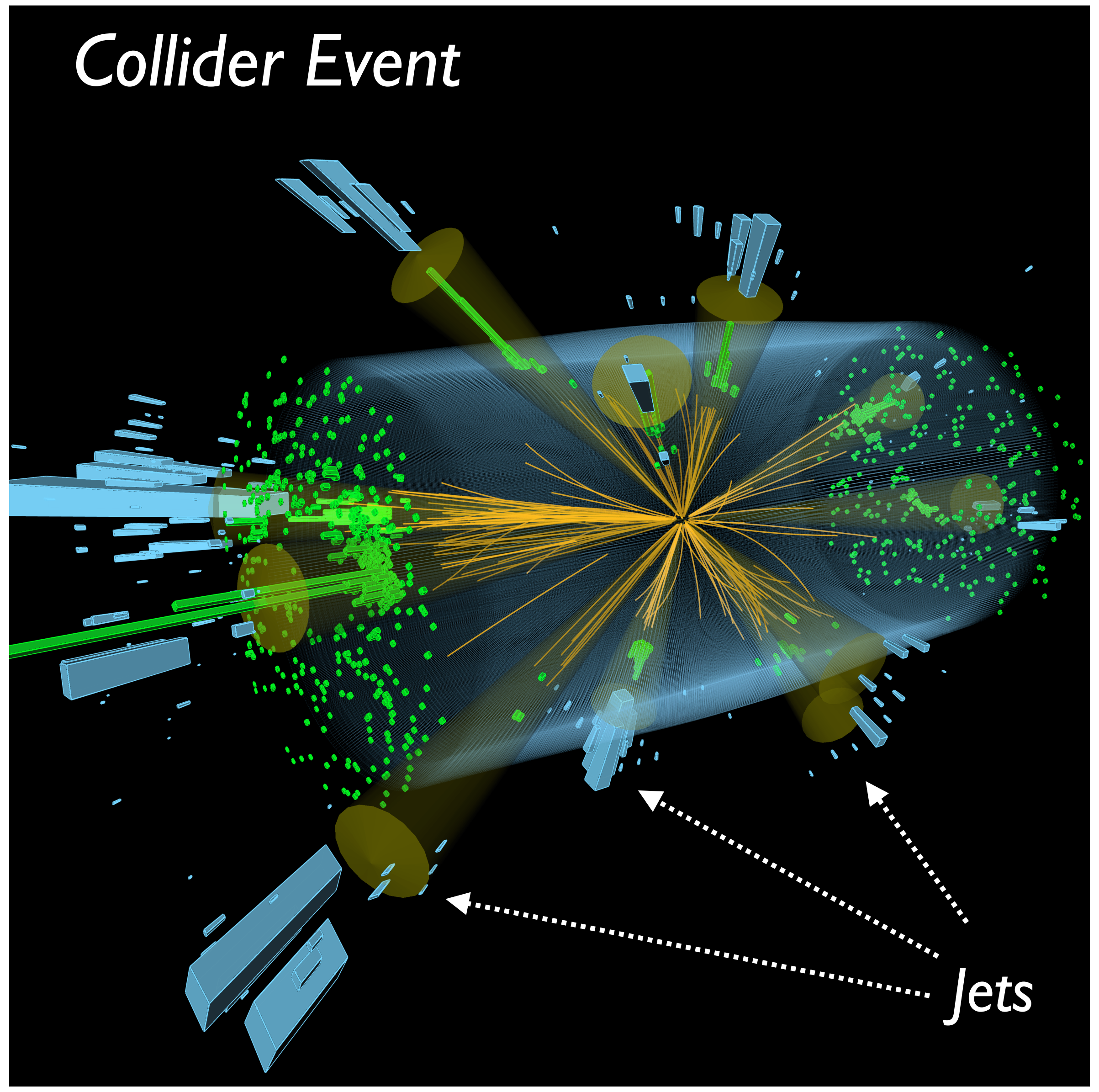
$\mathcal{E} \longrightarrow \mathcal{E}'$



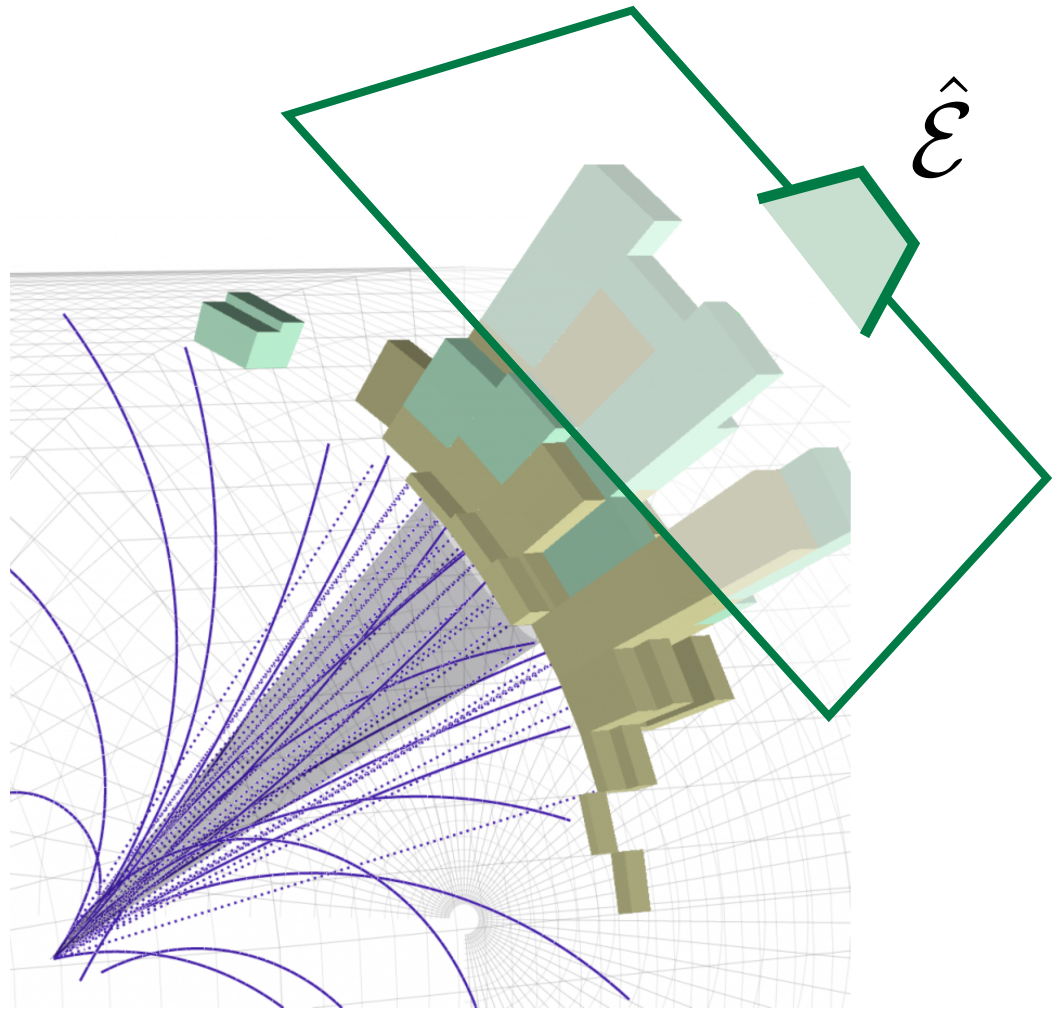
* for R sufficiently large
i.e., $R \geq$ jet radius for conical jets

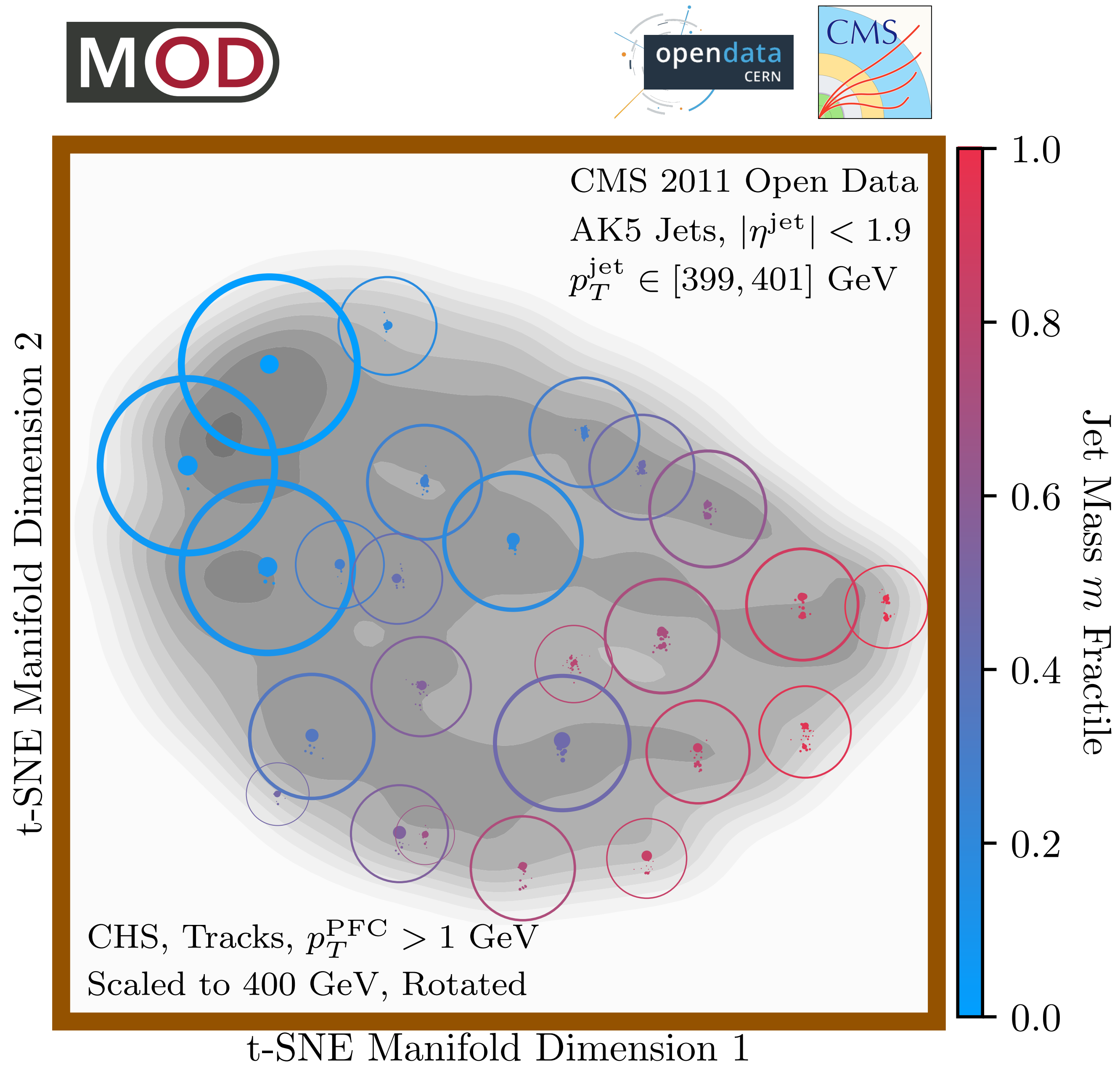
Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Collider data has geometric structure



Events summarized by
"energy flow"

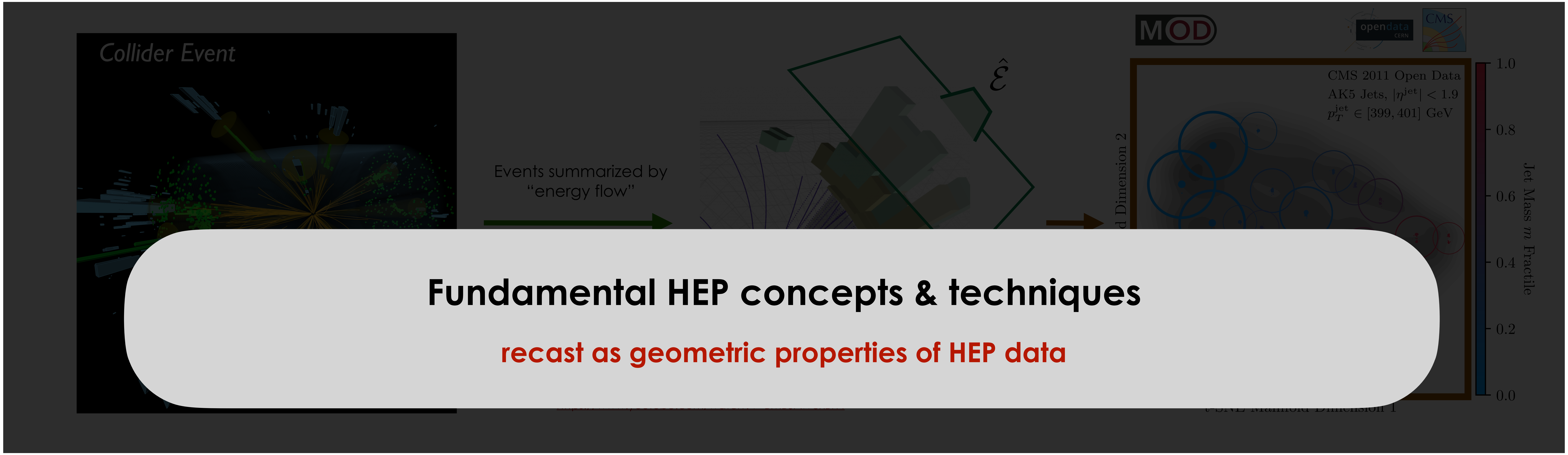




Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science:
https://www.youtube.com/watch?v=BMBSAWUxBn4

Komiske, Metodiev, Thaler, The hidden geometry of particle collisions, J. HEP, 6 (2020)

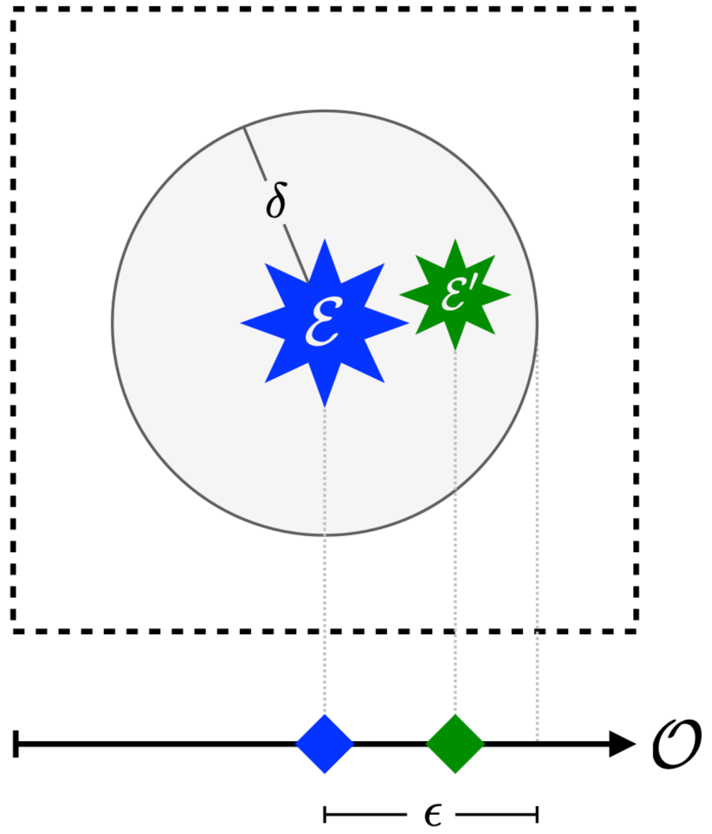
Collider data has geometric structure



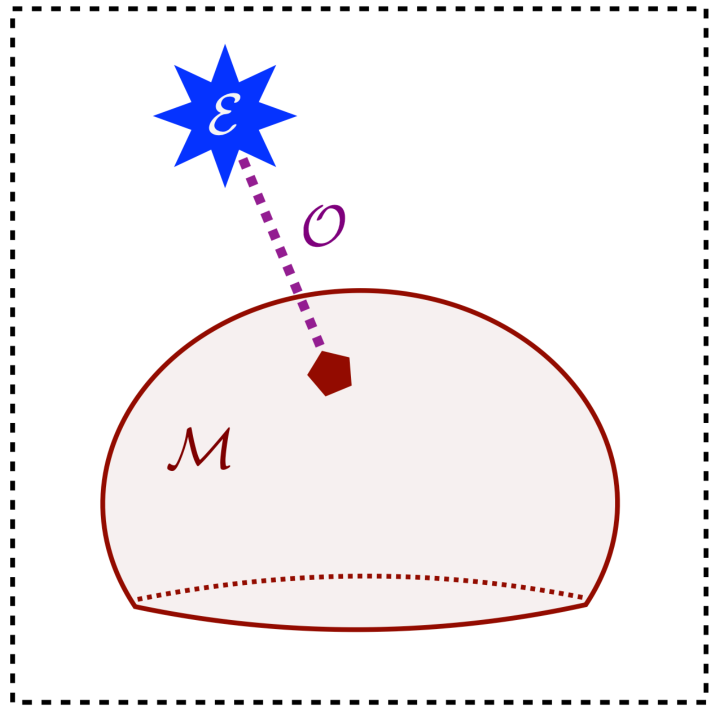
Fundamental HEP concepts & techniques

recast as geometric properties of HEP data

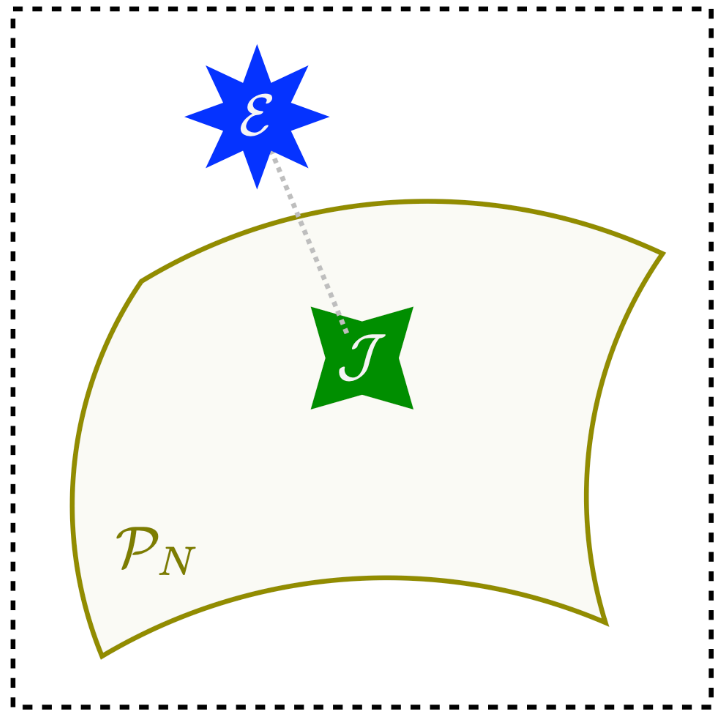
Infrared & Collinear Safety



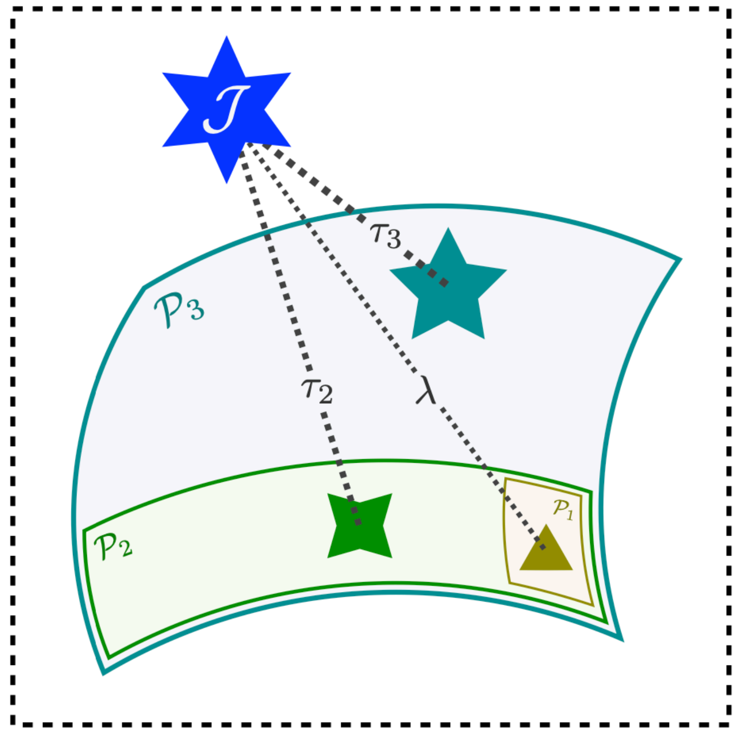
Observables



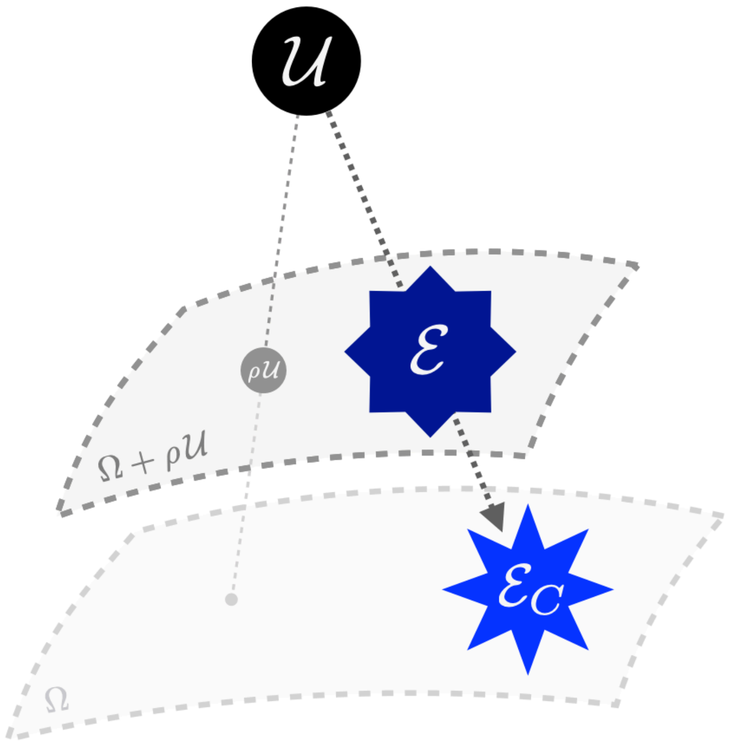
Jets



Jet substructure observables



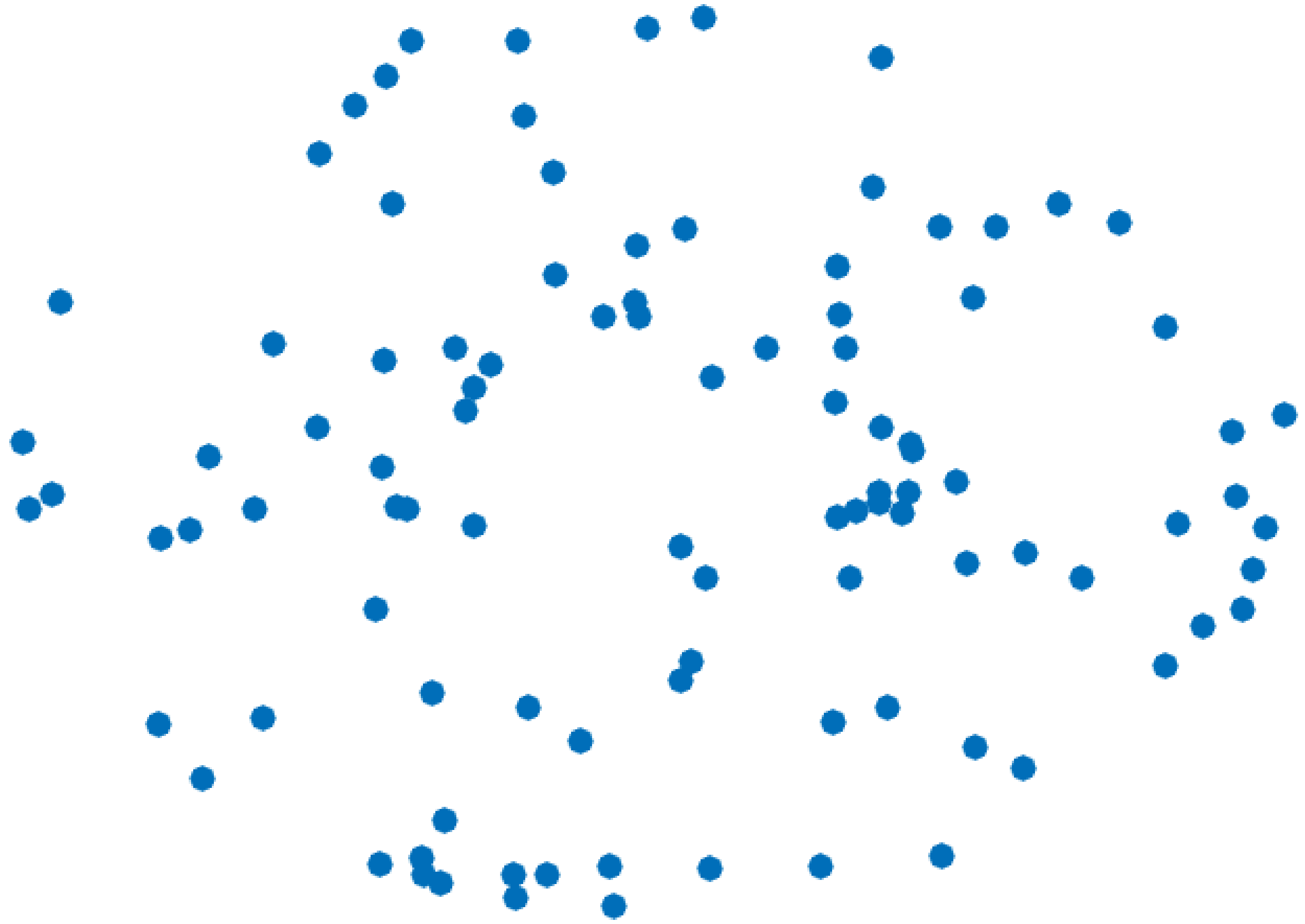
Pileup subtraction



Komiske, Metodiev, Thaler, The hidden geometry of particle collisions, J. HEP, 6 (2020)

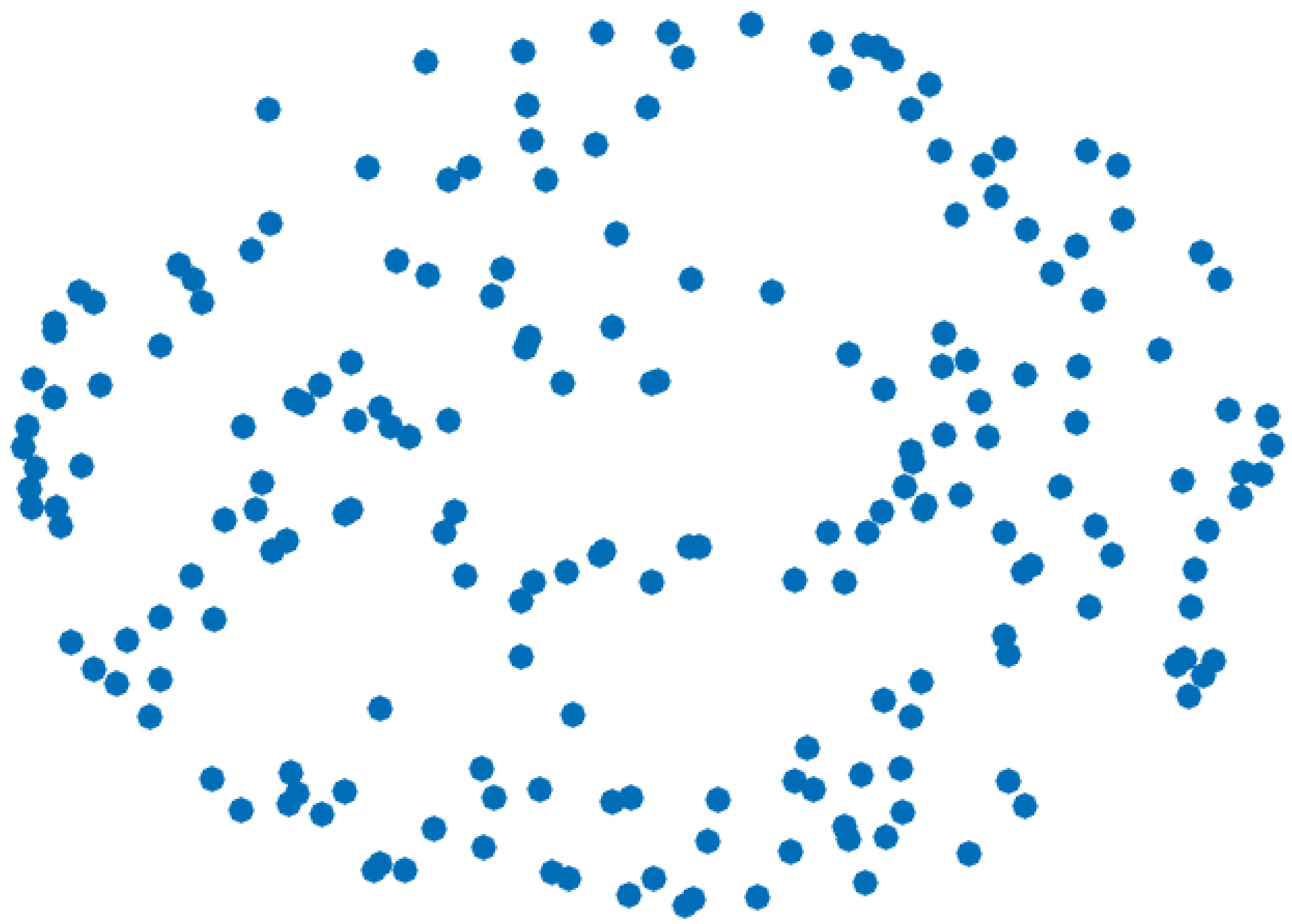
Recovering geometry of data:
problem of connecting data in the *right* way

What do you see?



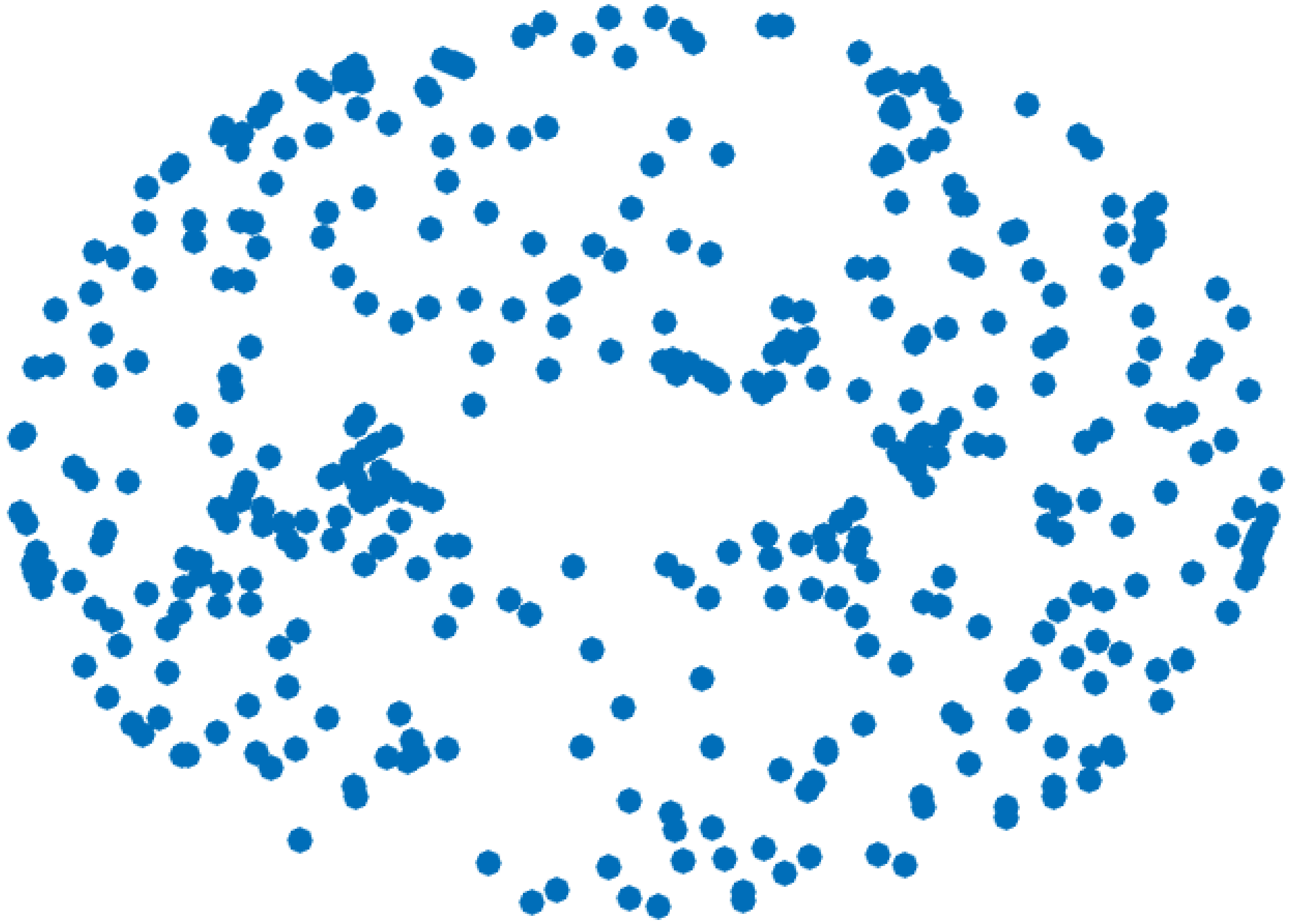
$N = 100$

What do you see?



$N = 200$

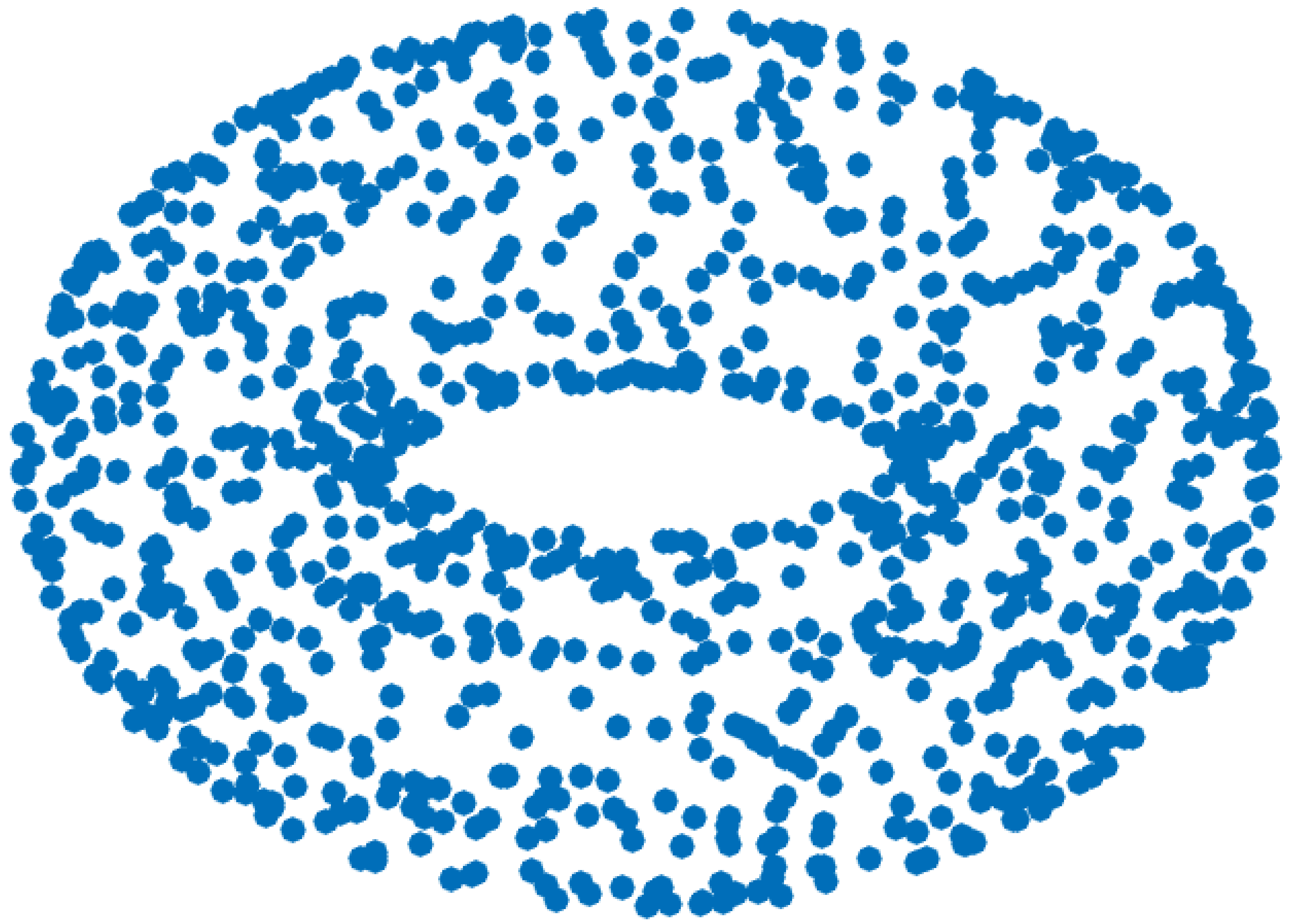
What do you see?



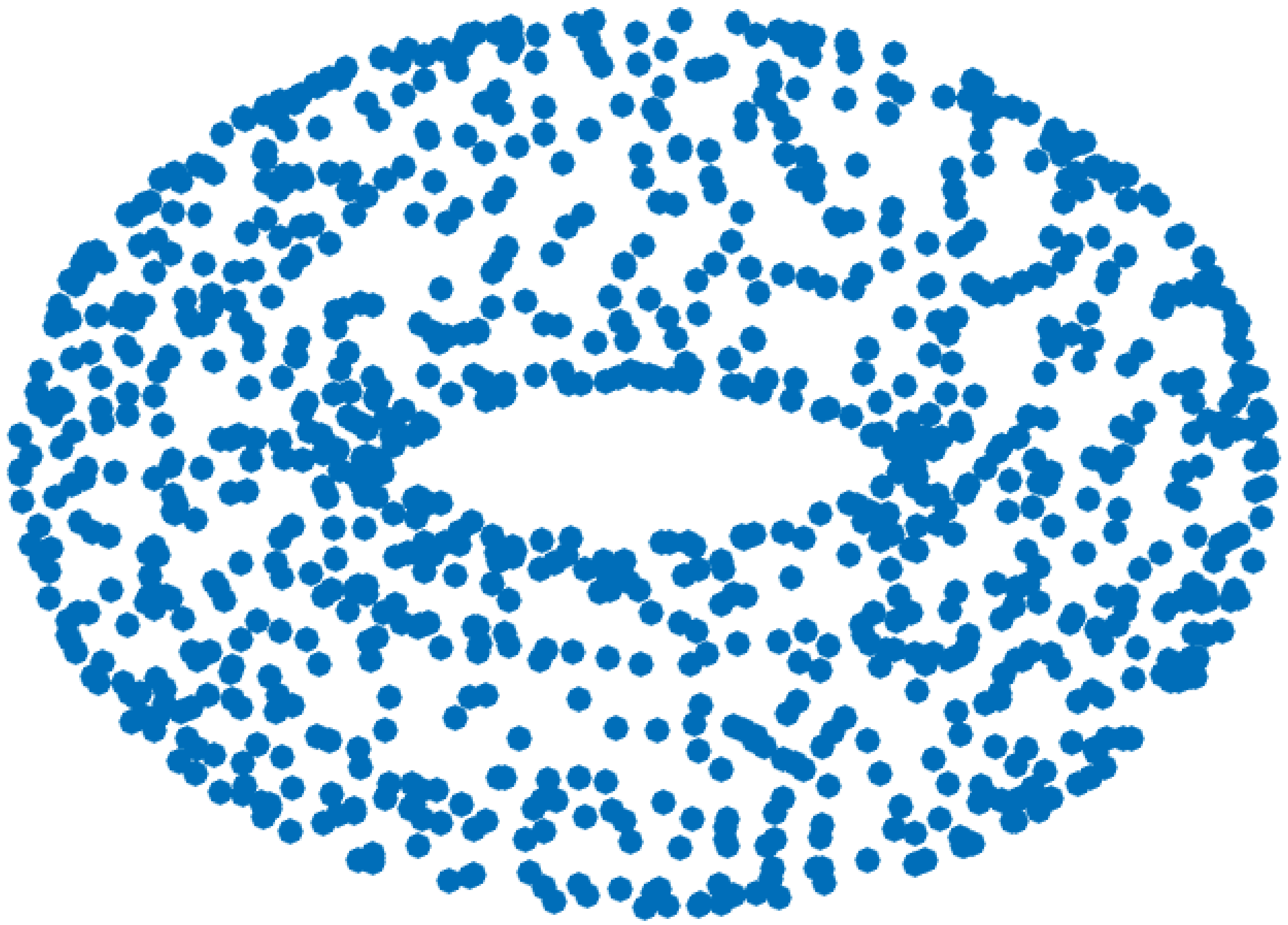
$N = 400$

What do you see?

$N = 1000$



What do you see?

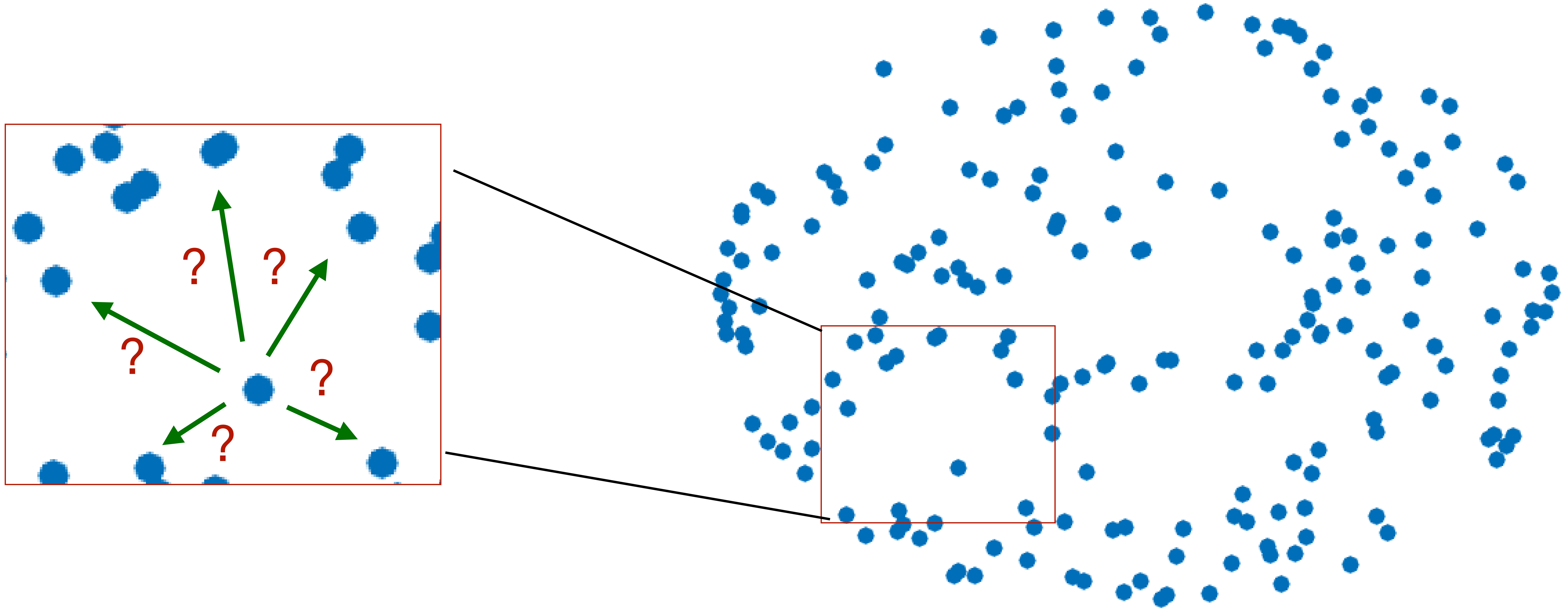


$N = 1000$

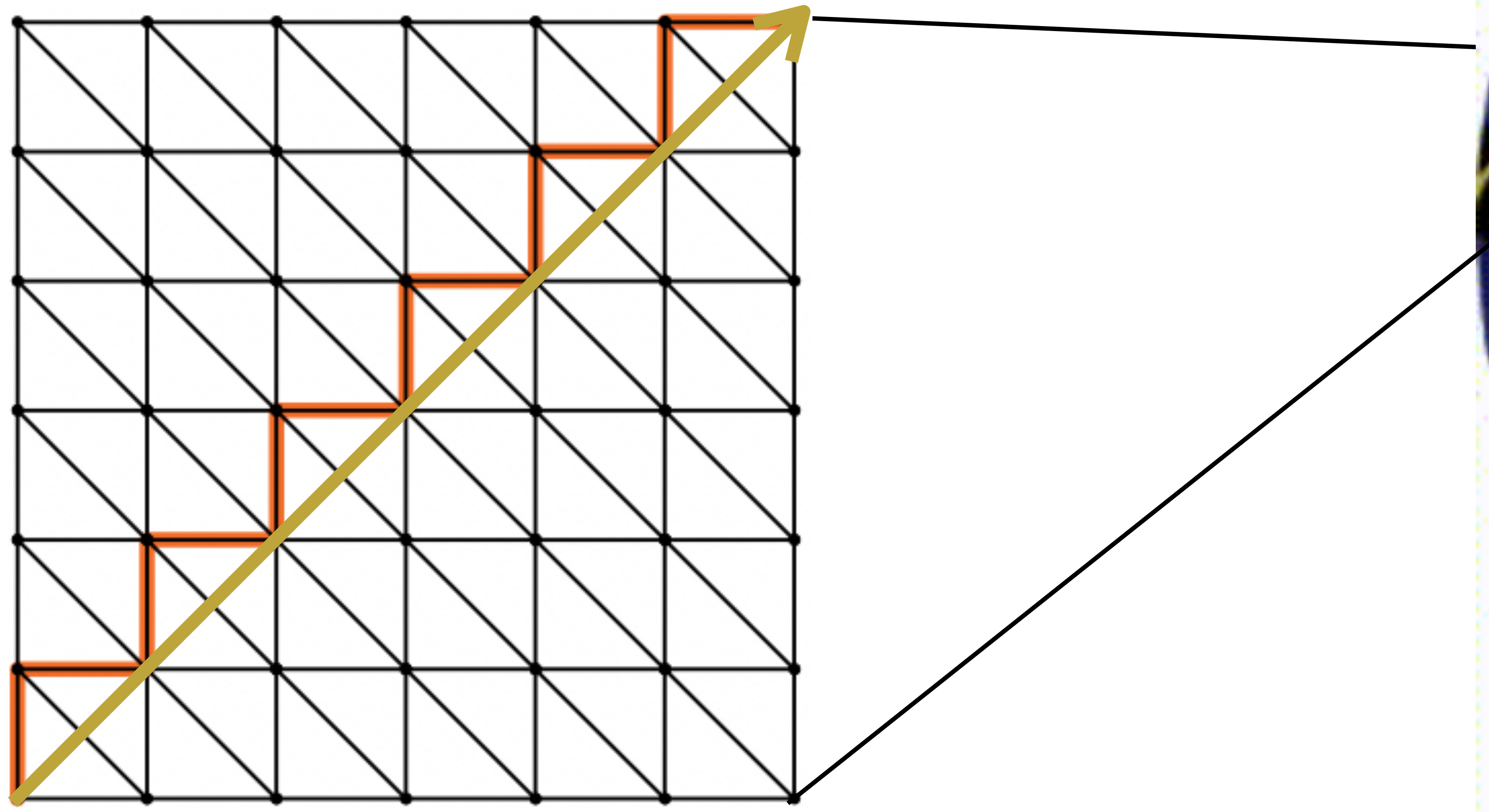
N data points in $\mathbb{R}^3$ sampled from a distribution constrained to a *submanifold* $T^2 \subset \mathbb{R}^3$

How is data **interconnected**?

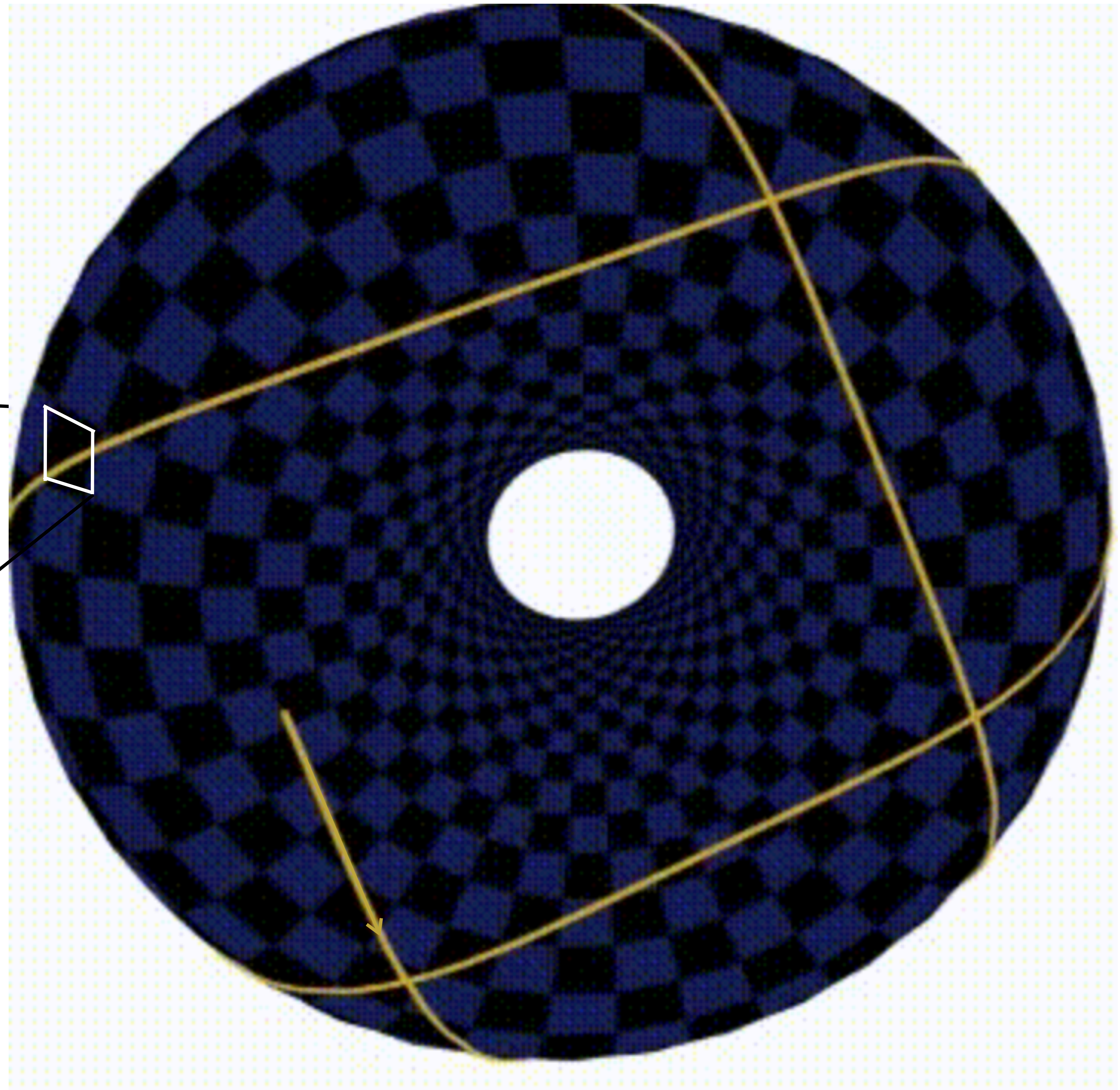
$N = 200$



How is data **interconnected**?



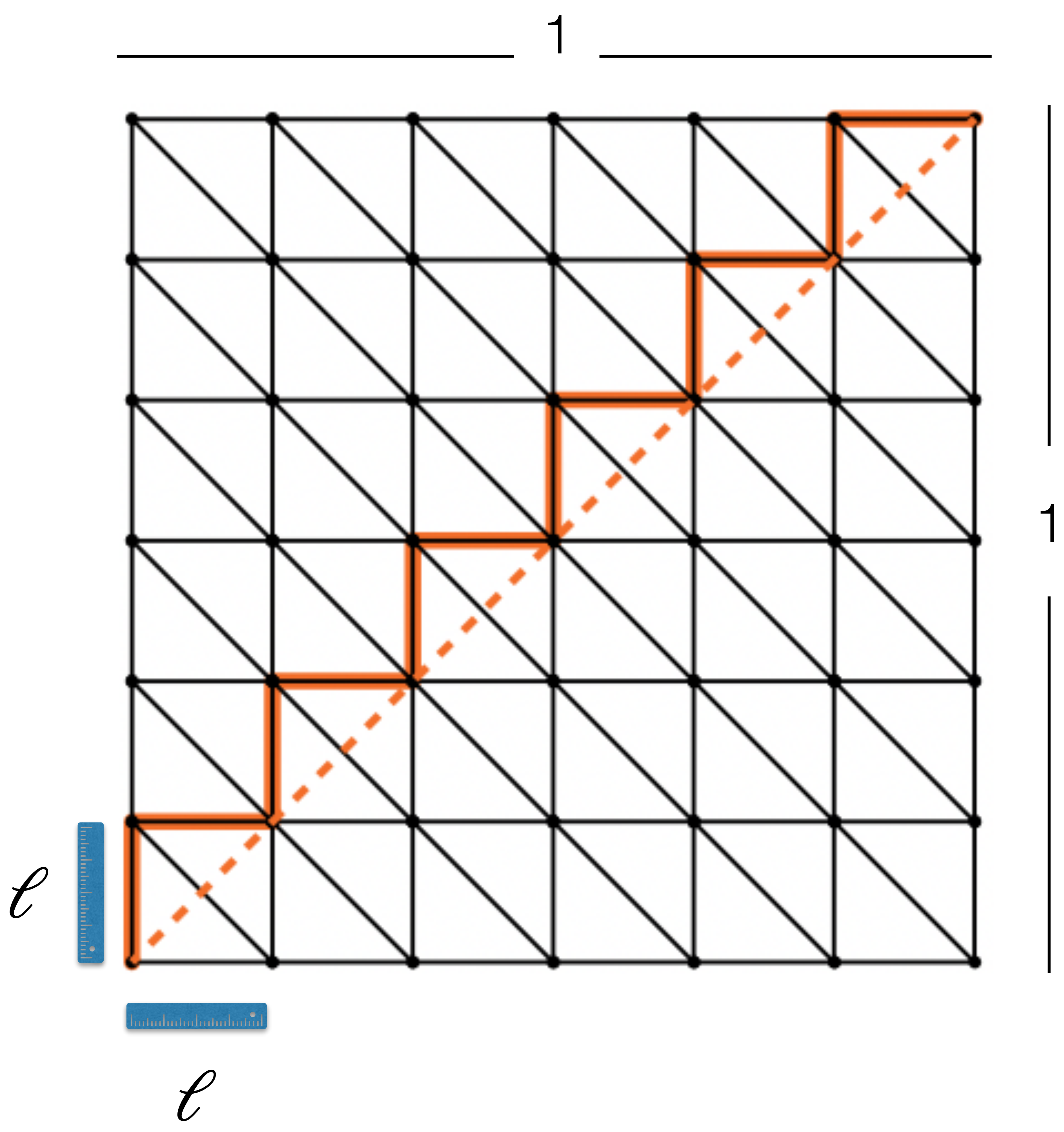
Crane et al., *Comm. ACM* (2017)



Wikipedia: *Geodesics*

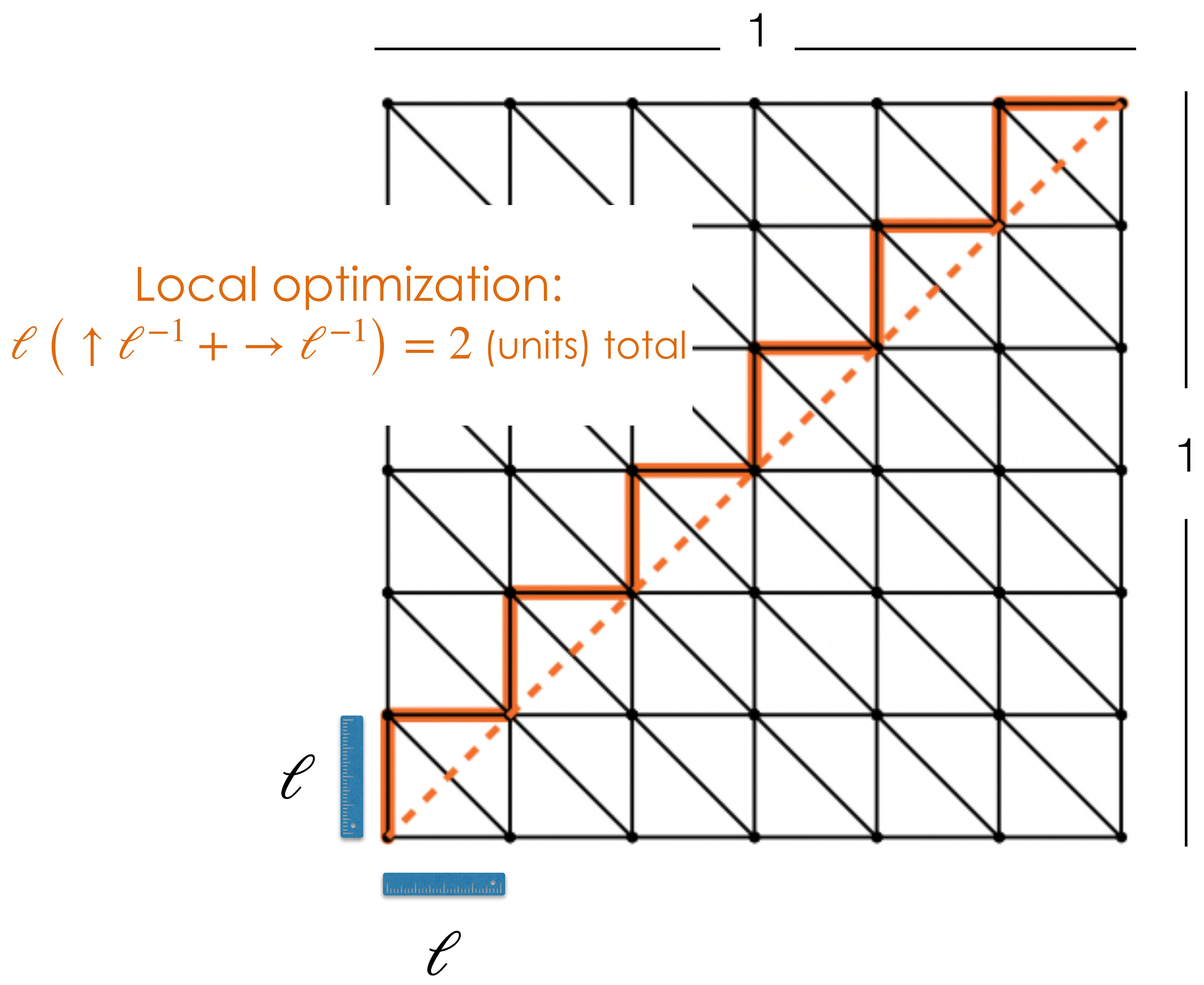
intrinsic local and **non-local** relationships between data points

How is data **interconnected**?



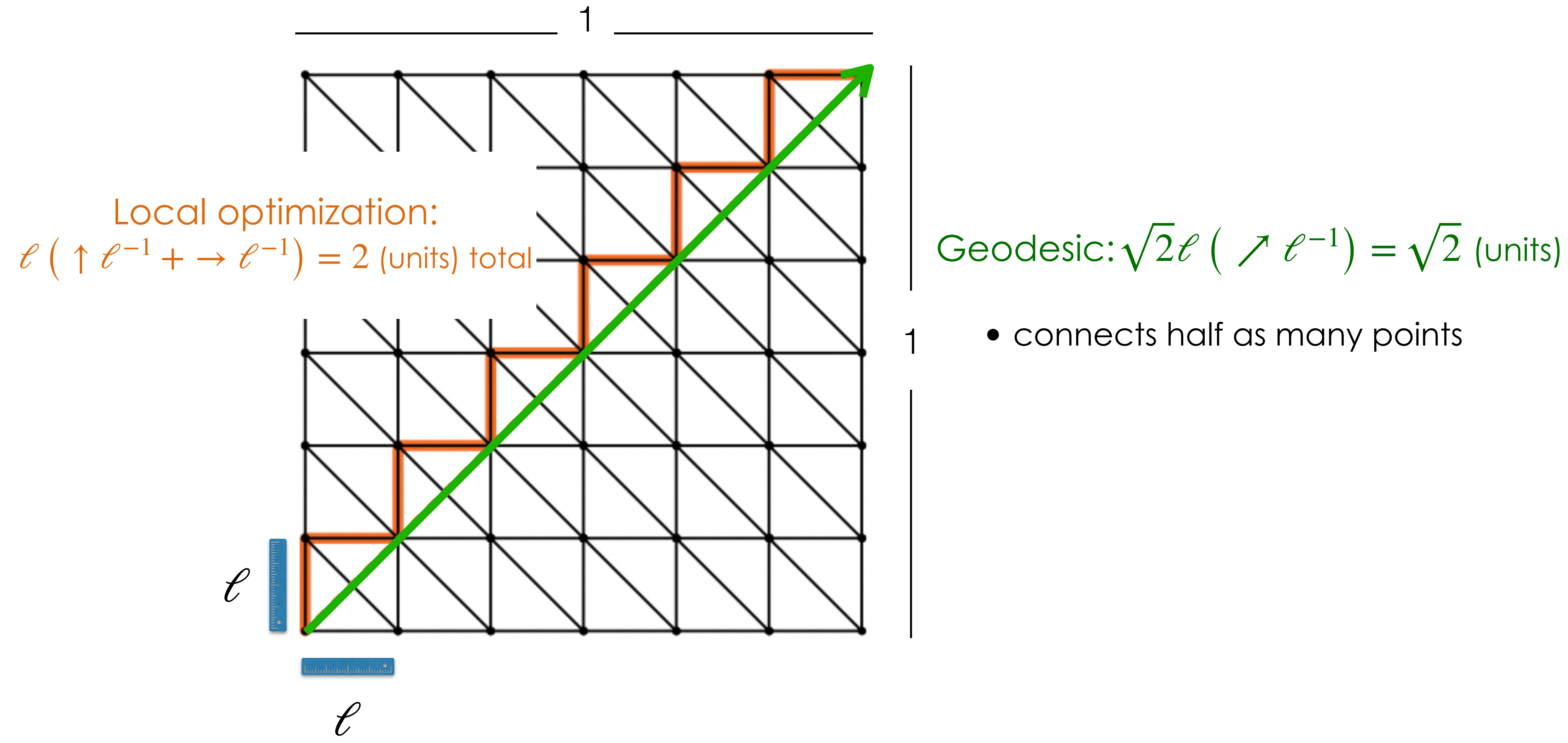
intrinsic local and **non-local** relationships between data points

How is data **interconnected**?



intrinsic local and **non-local** relationships between data points

How is data **interconnected**?



- connects half as many points

intrinsic local and **non-local** relationships between data points

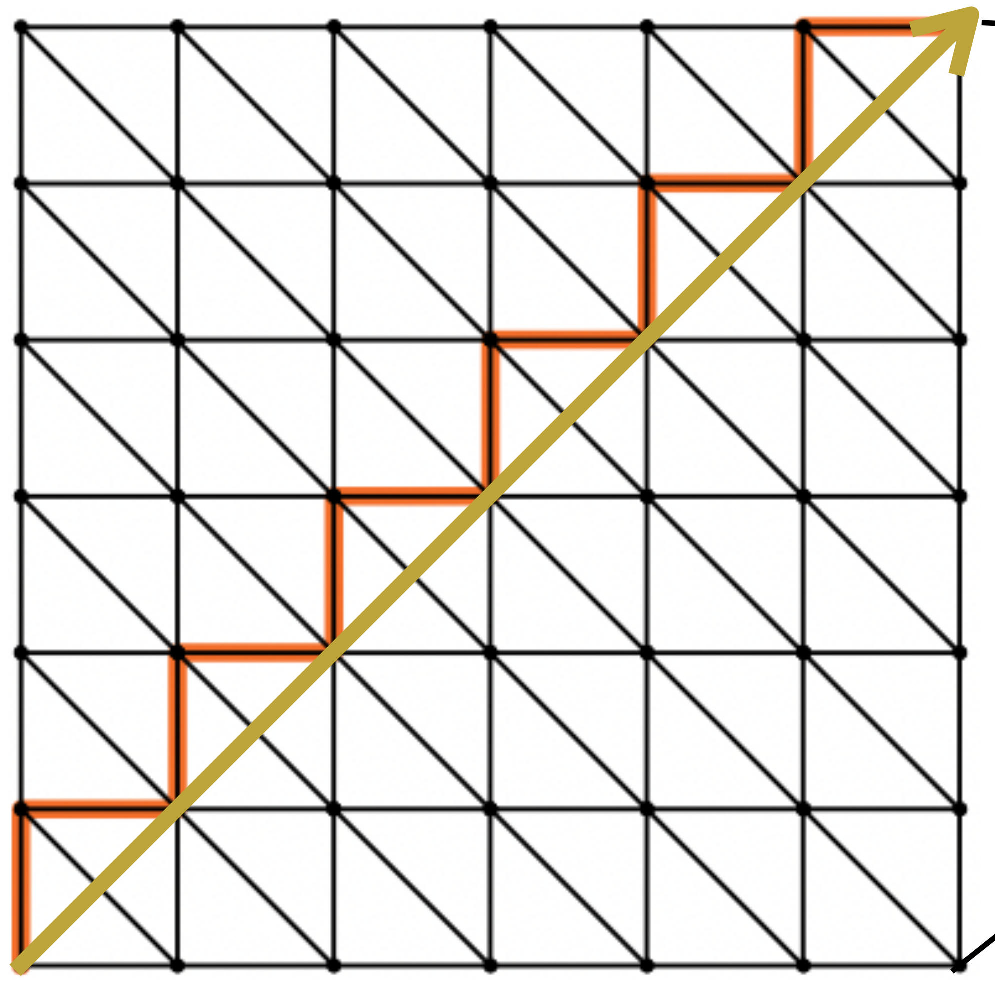
How is data **interconnected**?

Geodesic flow:

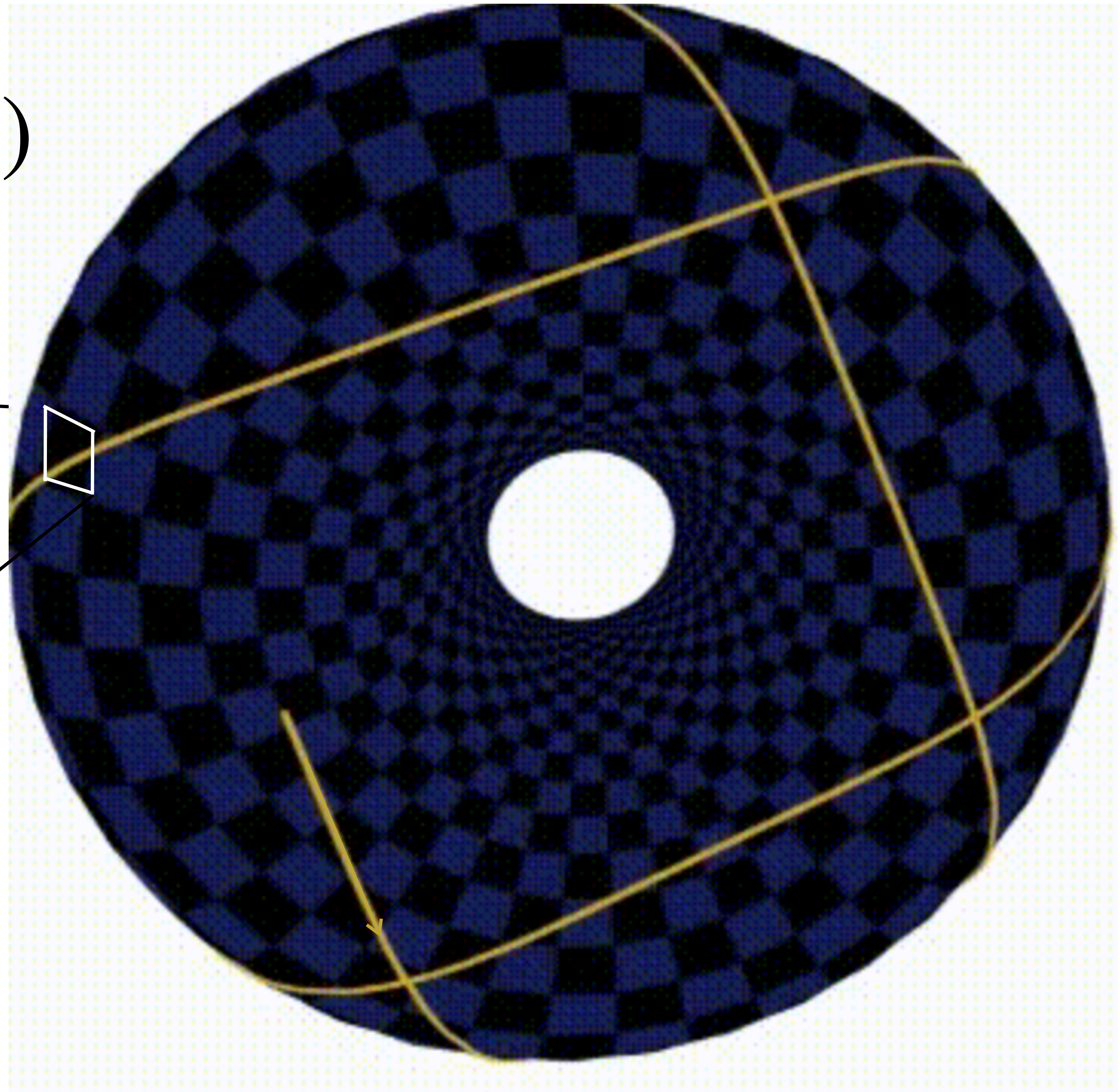
nonlinear in x, ξ

$$\begin{cases} \partial_t x_j &= \partial_{\xi_j} H = (g^{-1}(x) \cdot \xi)_j \\ \partial_t \xi_j &= -\partial_{x_j} H = \langle \partial_{x_j} g^{-1}(x) \cdot \xi, \xi \rangle \end{cases}$$

$$\hookrightarrow \Gamma^t(x_0, \xi_0)$$



Crane et al., *Comm. ACM* (2017)



Wikipedia: *Geodesics*

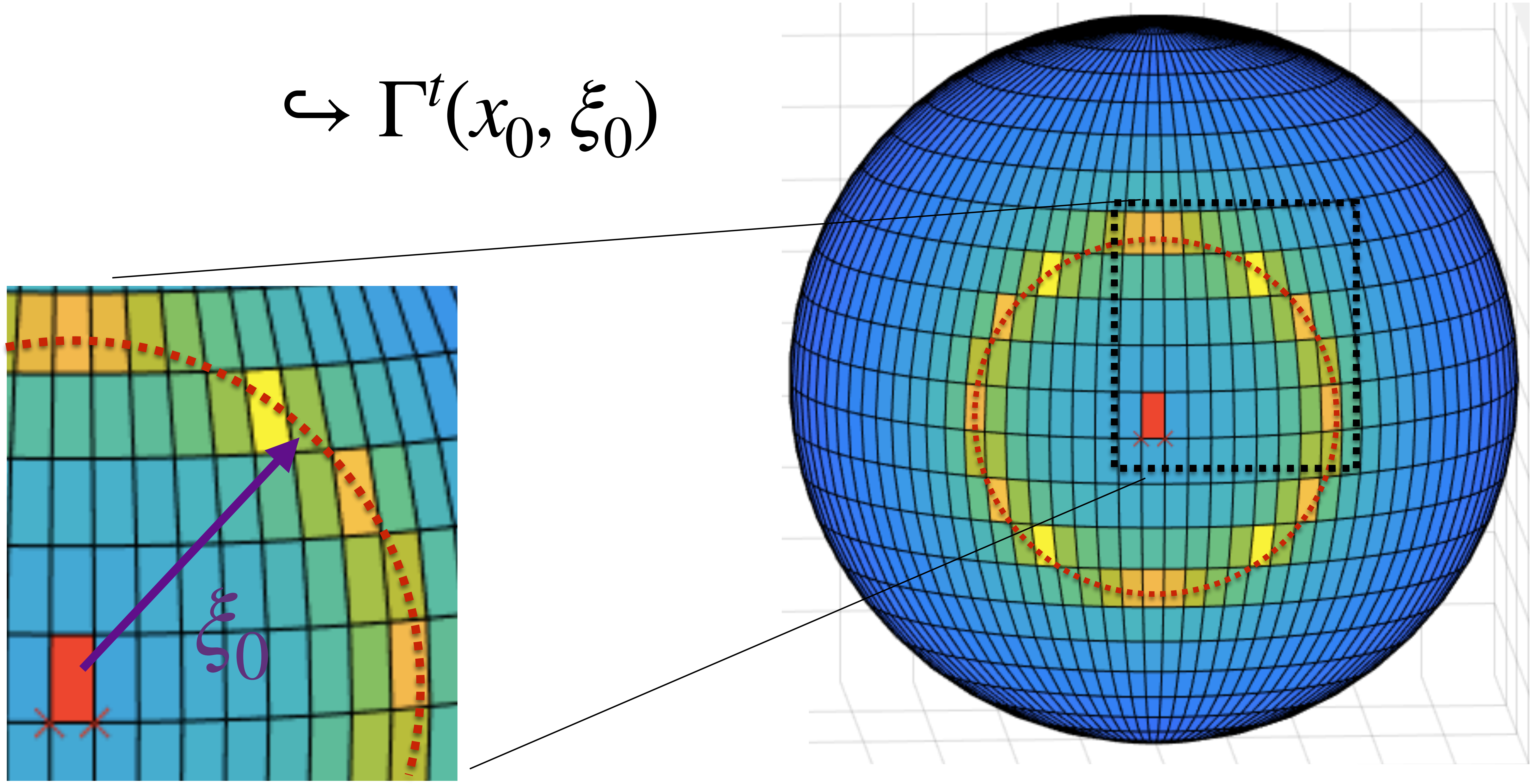
How is data **interconnected**?

Geodesic flow:

nonlinear in x, ξ

$$\begin{cases} \partial_t x_j & = \partial_{\xi_j} H = (g^{-1}(x) \cdot \xi)_j \\ \partial_t \xi_j & = -\partial_{x_j} H = \langle \partial_{x_j} g^{-1}(x) \cdot \xi, \xi \rangle \end{cases}$$

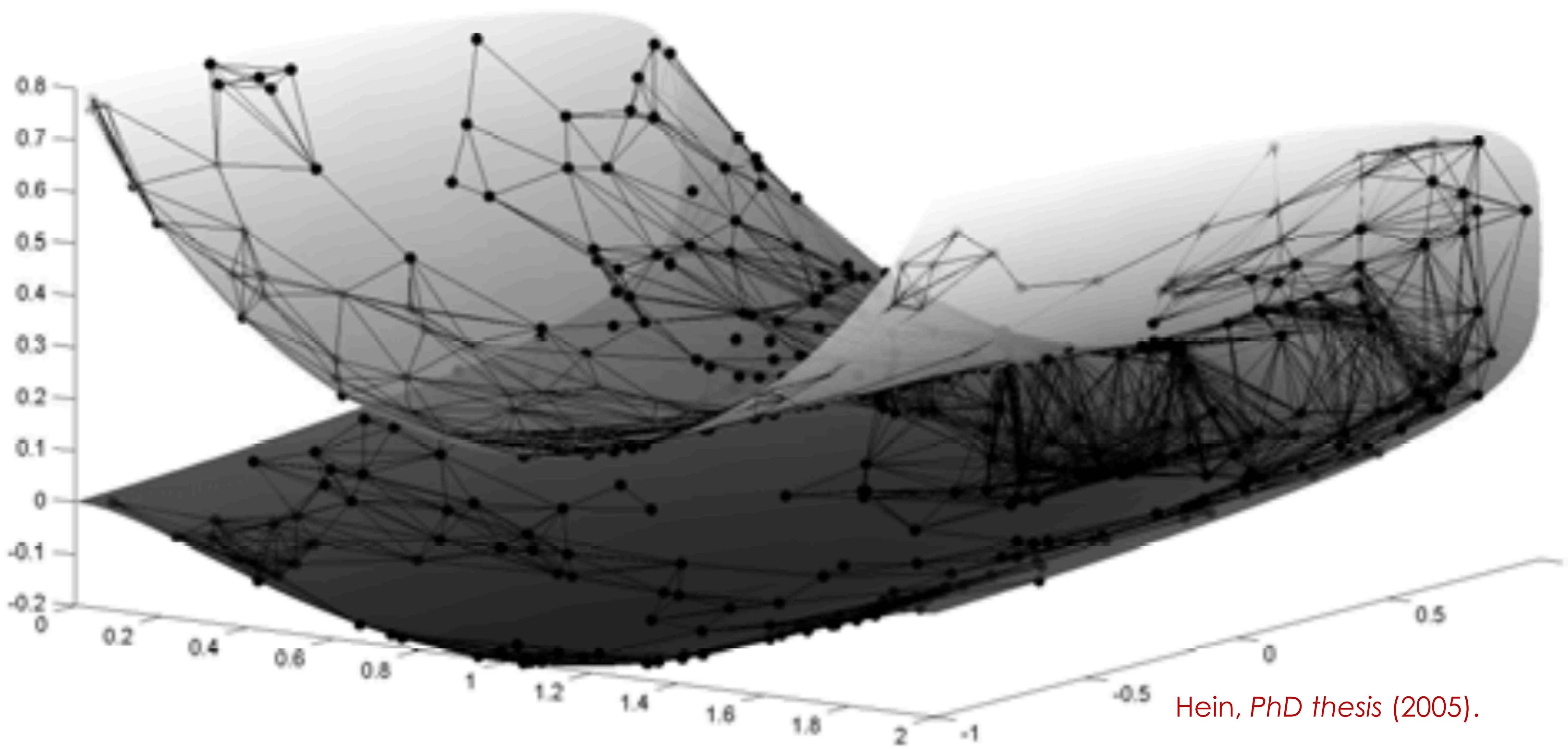
$$\hookrightarrow \Gamma^t(x_0, \xi_0)$$



The flow moves the point x_0 to the geodesic neighbour $x(t)$ at distance t in the direction ξ_0

The problem

Given *data* $X_N := \{v_1, \ldots, v_N\} \subset \mathcal{M} \subset \mathbb{R}^D$ sampled from a probability distribution confined to a Riemannian submanifold $\mathcal{M}$, **recover** the *intrinsic geometry* of $\mathcal{M}$



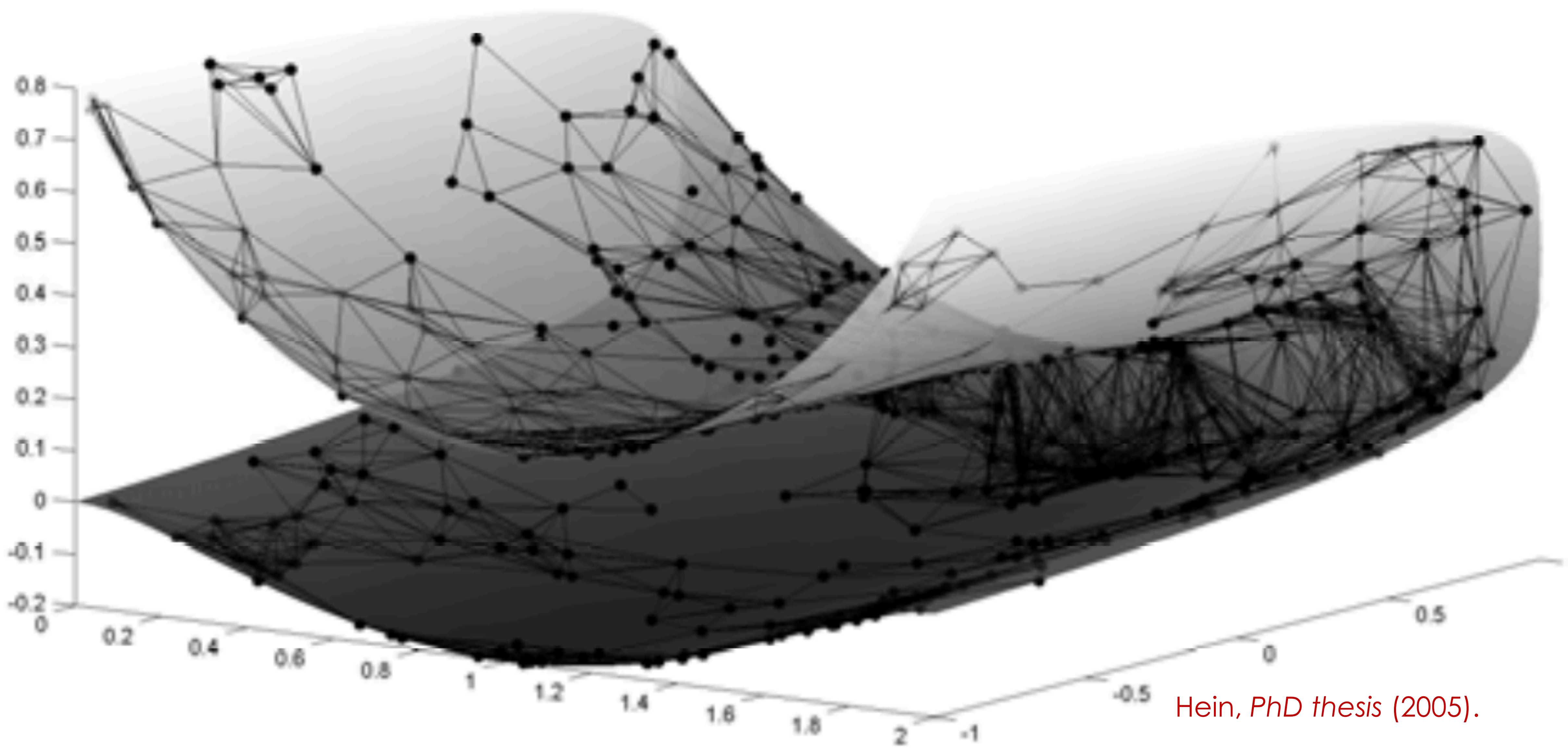
Hein, *PhD thesis* (2005).

The problem

Given *data* $X_N := \{v_1, \ldots, v_N\} \subset \mathscr{M} \subset \mathbb{R}^D$ sampled from a probability distribution confined to a Riemannian submanifold $\mathscr{M}$, **recover** the *intrinsic geometry* of $\mathscr{M}$

- intrinsic dimension, Riemannian metric, low-dimensional isometric embedding, heat diffusion, $\Delta_{\mathscr{M}}$

Belkin-Niyogi (2003), Hein (2005), Hein-Audibert-von Luxburg (2007), Trillos-Gerlach-Hein-Slepčev (2020)



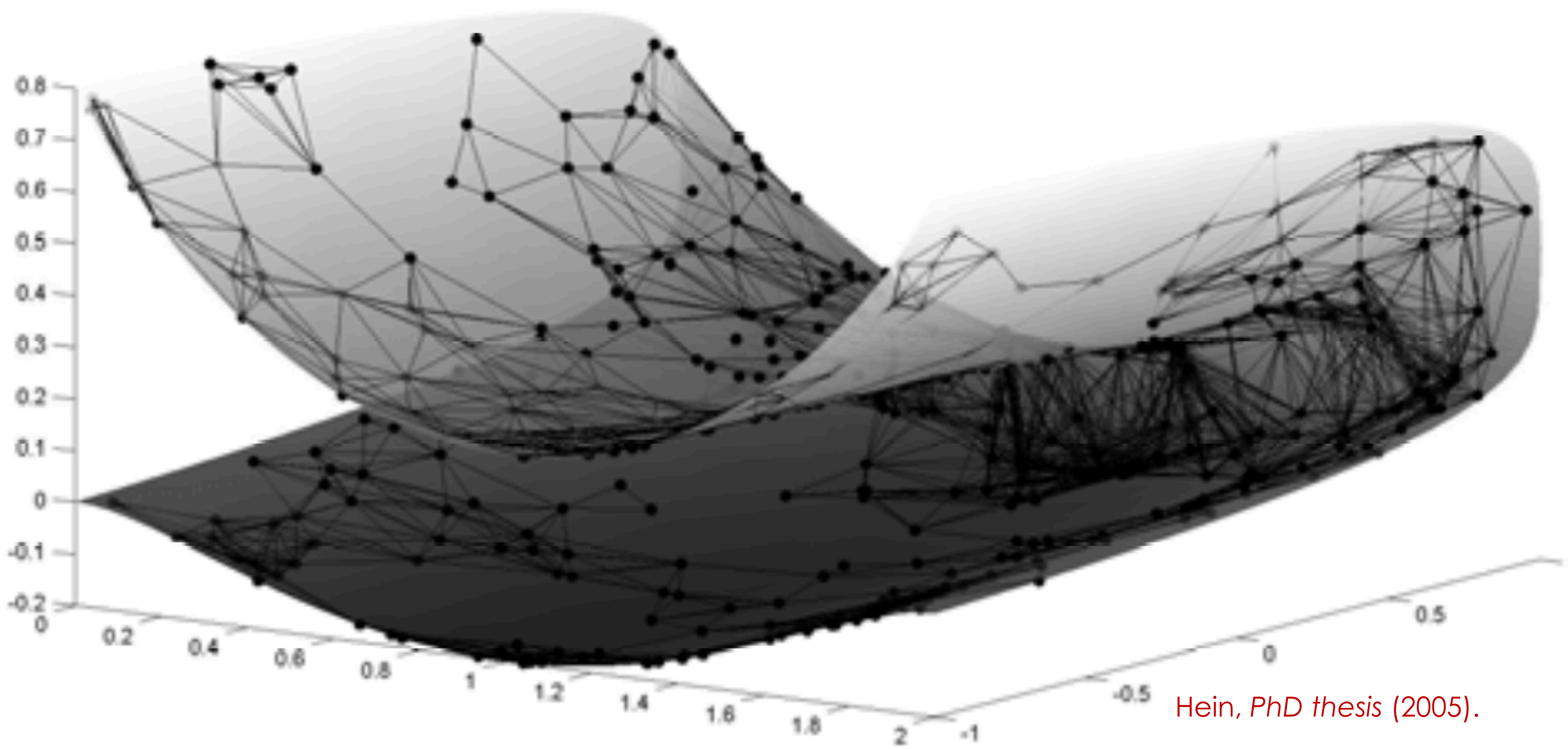
Hein, *PhD thesis* (2005).

The problem

Given *data* $X_N := \{v_1, \ldots, v_N\} \subset \mathscr{M} \subset \mathbb{R}^D$ sampled from a probability distribution confined to a Riemannian submanifold $\mathscr{M}$, **recover** the *intrinsic geometry* of $\mathscr{M}$

- intrinsic dimension, Riemannian metric, low-dimensional isometric embedding, heat diffusion, $\Delta_{\mathscr{M}}$ } *Belkin-Niyogi (2003), Hein (2005), Hein-Audibert-von Luxburg (2007), Trillos-Gerlach-Hein-Slepčev (2020)*

- **This talk:** Geometry-encoding quantum dynamics, geodesics ✦ Much harder!



Hein, *PhD thesis* (2005).

The problem

Given *data* $X_N := \{v_1, \ldots, v_N\} \subset \mathscr{M} \subset \mathbb{R}^D$ sampled from a probability distribution confined to a Rieman

), Hein (2005),
Luxburg (2007),
n-Slepčev (2020)

State of the art is to deploy *local* dynamics on data, *i.e.*, Markov processes

These yield *low-frequency* features of data, but miss *high-frequency* components (*i.e.*, geodesics)

Hein, *PhD thesis* (2005).

The solution

Graph Laplacian construction

Discrete quantized propagation: taking the position mean gives $O(h)$ approx. to $\iota(t)$

Inputs: $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$
Output: Propagated state $[\psi_h^\zeta](t)$
procedure PROPAGATE
  for $i, j = 1 : N$ **do** $[T_\epsilon]_{i,j} \leftarrow k(||v_i - v_j||^2/\epsilon)$
  $D_\epsilon \leftarrow$ diagonal matrix $\left(\sum_{j=1}^N [T_\epsilon]_{i,j}\right)_{1 \leq i \leq N}$
  $\Delta_{\epsilon,N} \leftarrow \frac{4(I_N - D_\epsilon^{-1} T_\epsilon)}{\epsilon}$
  $U_{\epsilon,N}^t \leftarrow \exp(-it\sqrt{\Delta_{\epsilon,N}})$
  $h \leftarrow \epsilon^{\frac{1}{(2+\alpha)}}$
  $p_0 \leftarrow v_j - v^*$ for v_j closest to point v^*
  while $1 \leq \ell \leq N$ **do**
    $[\psi_h^\zeta]_\ell \leftarrow e^{-||v_\ell - v^*||^2/2h} e^{\frac{i}{h}(v_\ell - v^*)^\top p_0/||p_0||}$
  return $[\psi_h^\zeta](t) = U_{\epsilon,N}^t [\psi_h^\zeta]$

The solution

Inputs: $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$

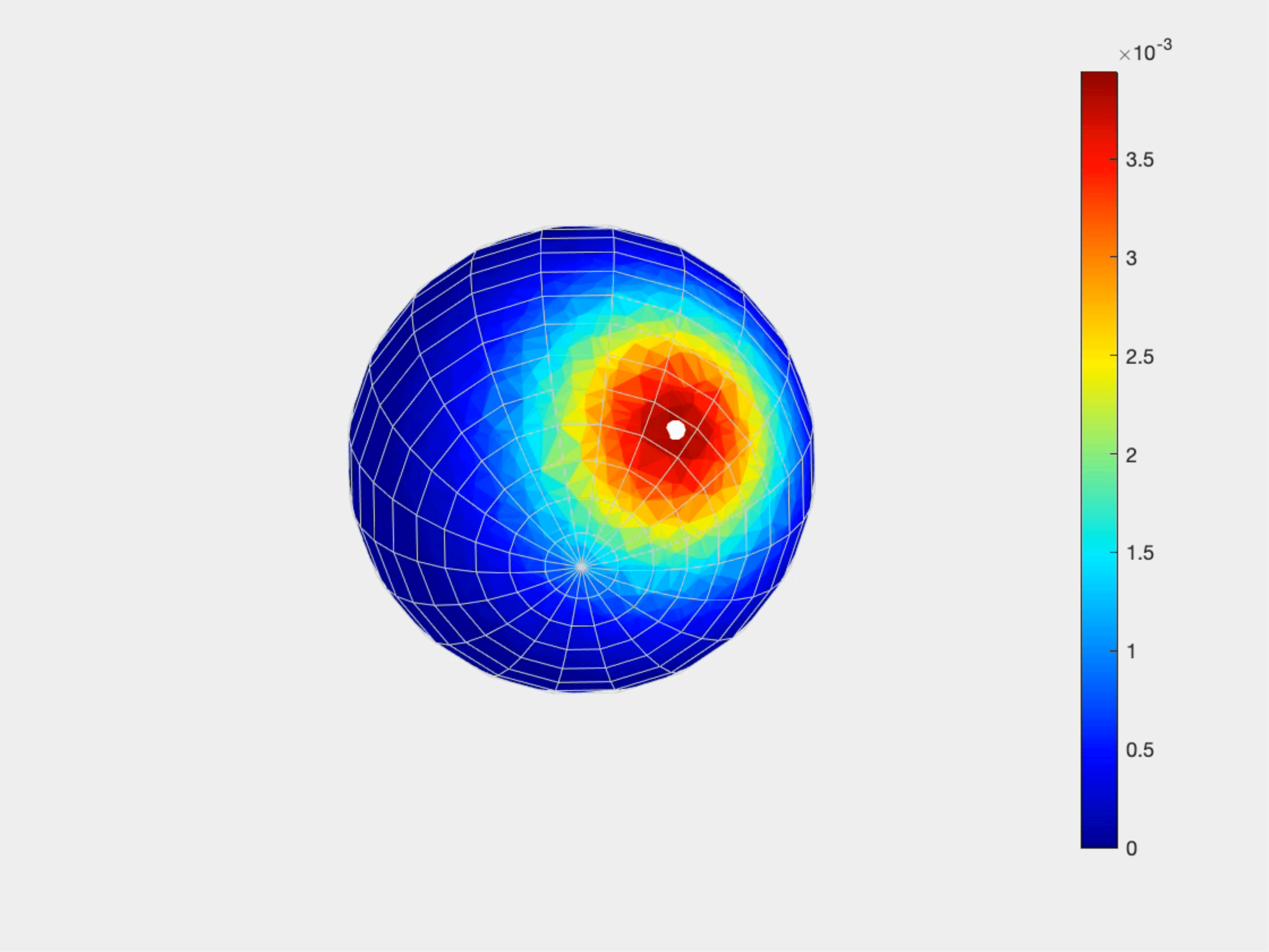
Matrix dynamics $e^{it\sqrt{\Delta}_{\epsilon,N}} \left| \psi_h^{(x_0,\xi_0)} \right\rangle$ on dataset

follows geodesic flow starting at x_0 in direction ξ_0 for distance t

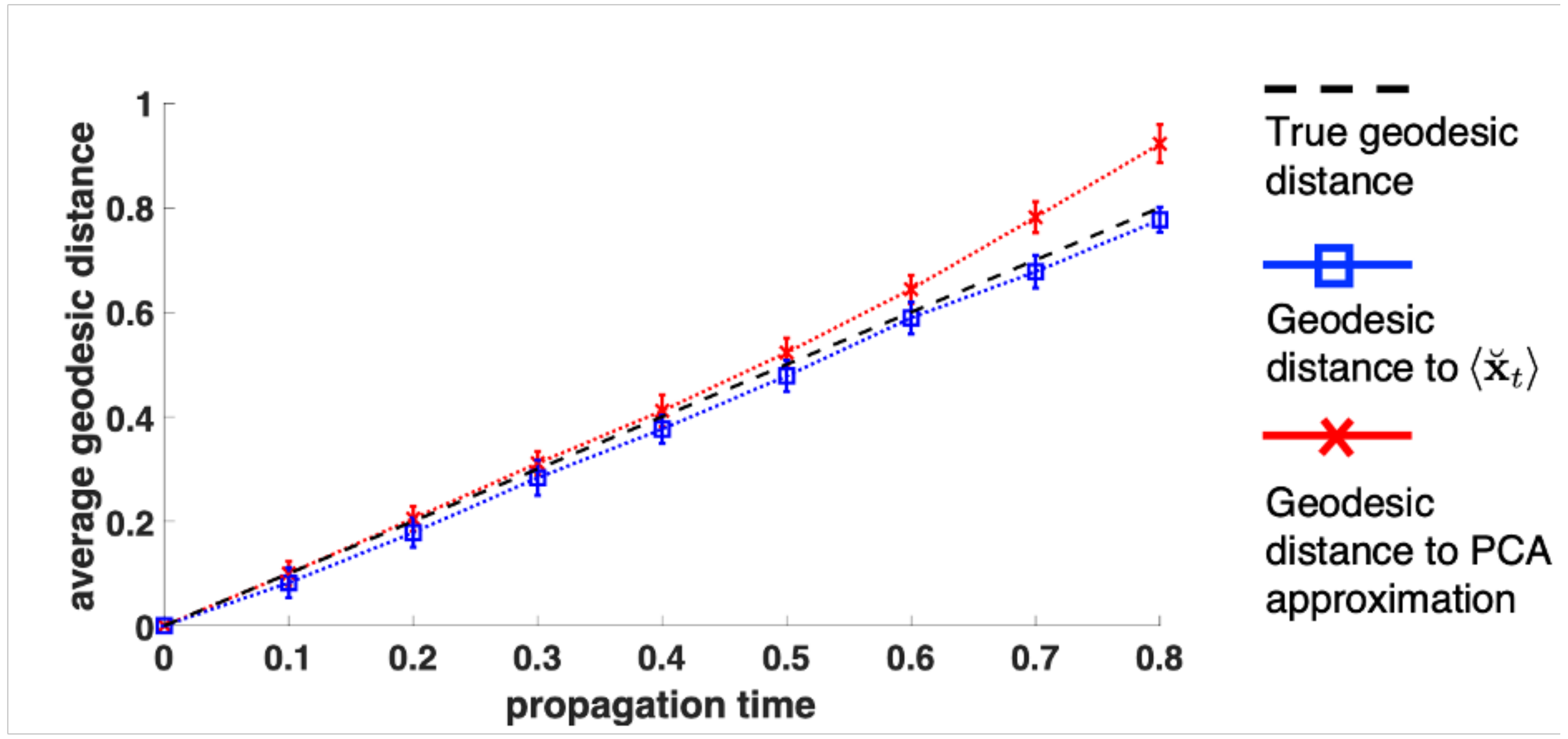
with $O(h)$ error

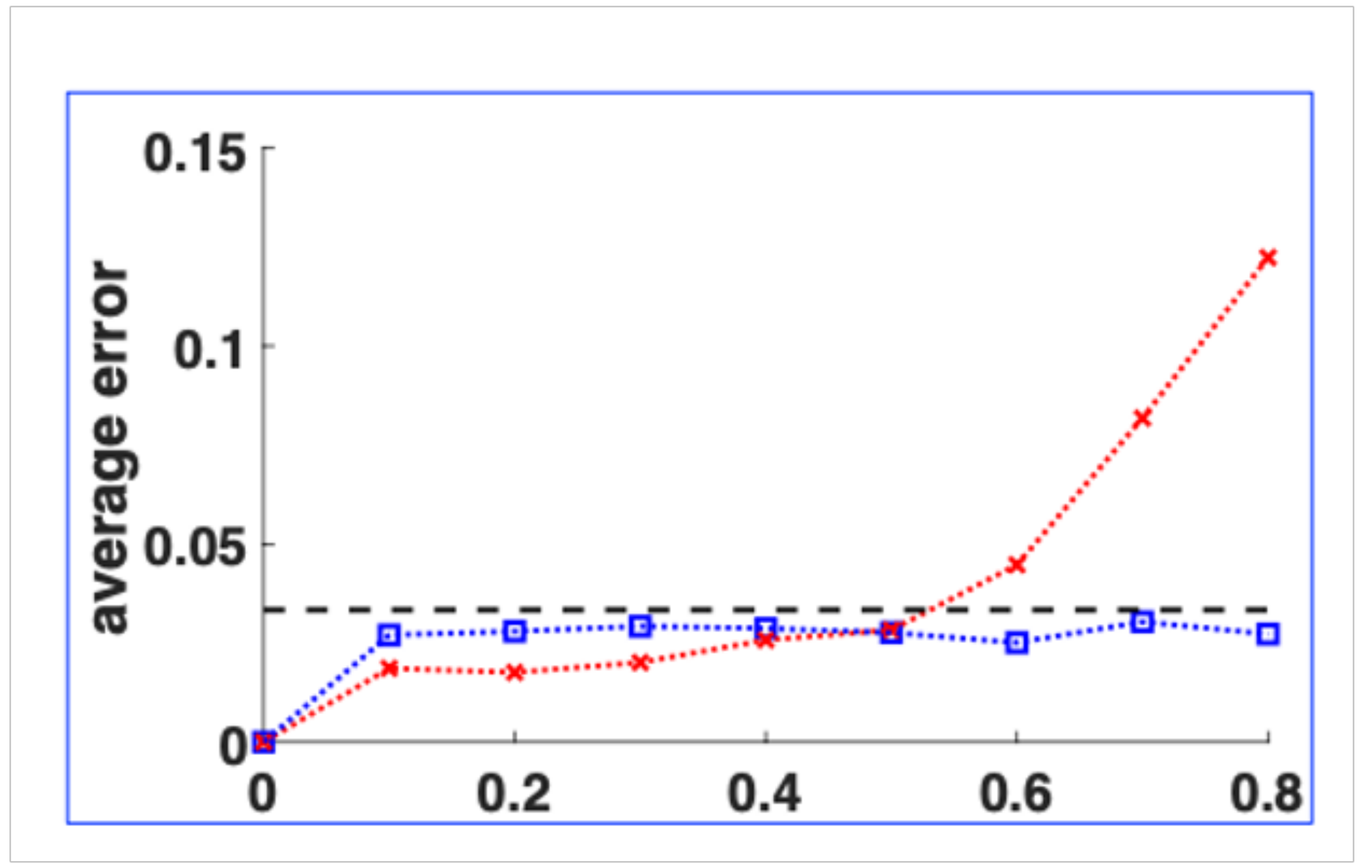
return $[\psi_h^{\zeta}](t) = U^t_{\epsilon,N}[\psi_h^{\zeta}]$

The solution

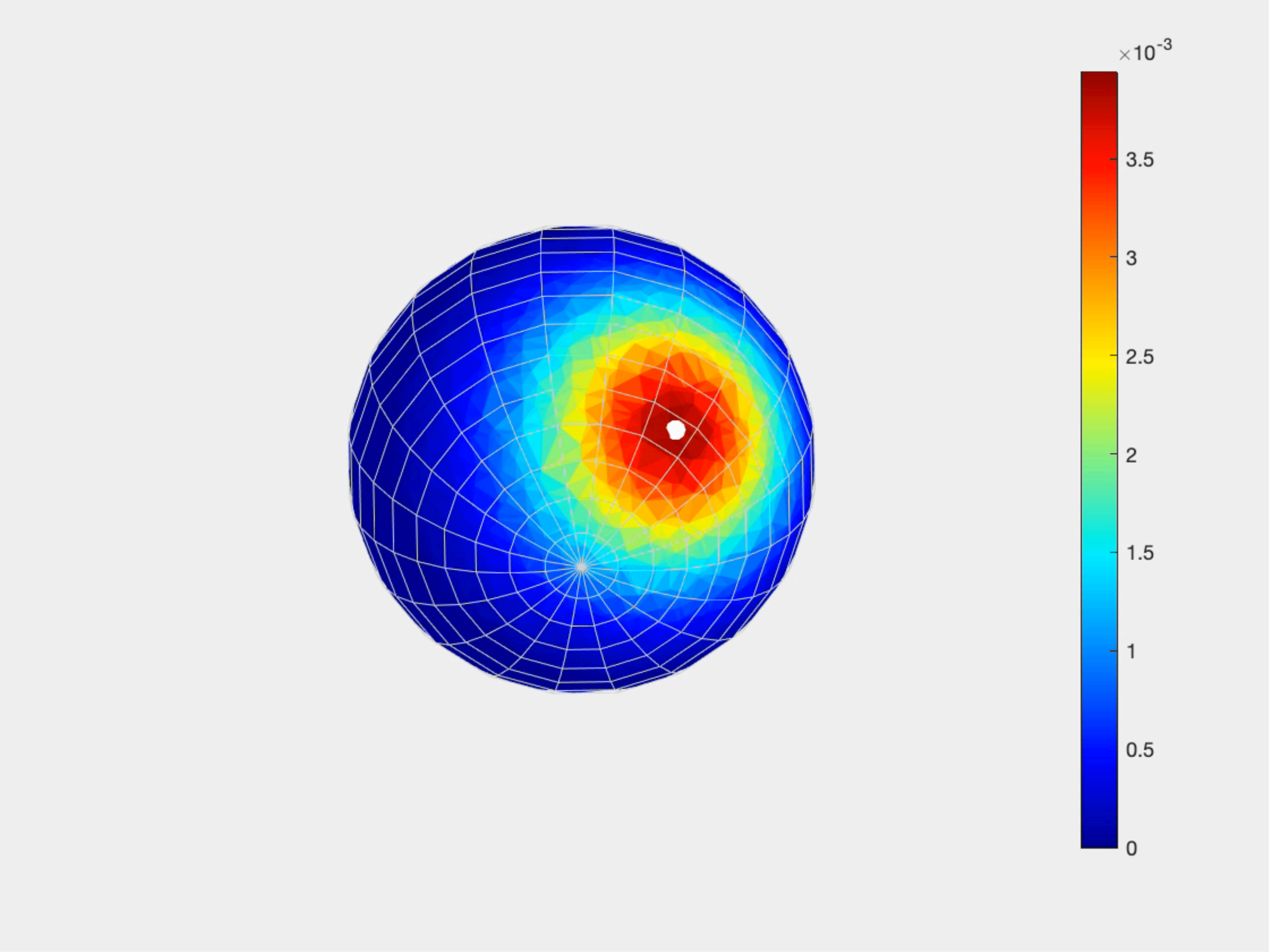


$N = 3000$ points
uniformly sampled on unit sphere

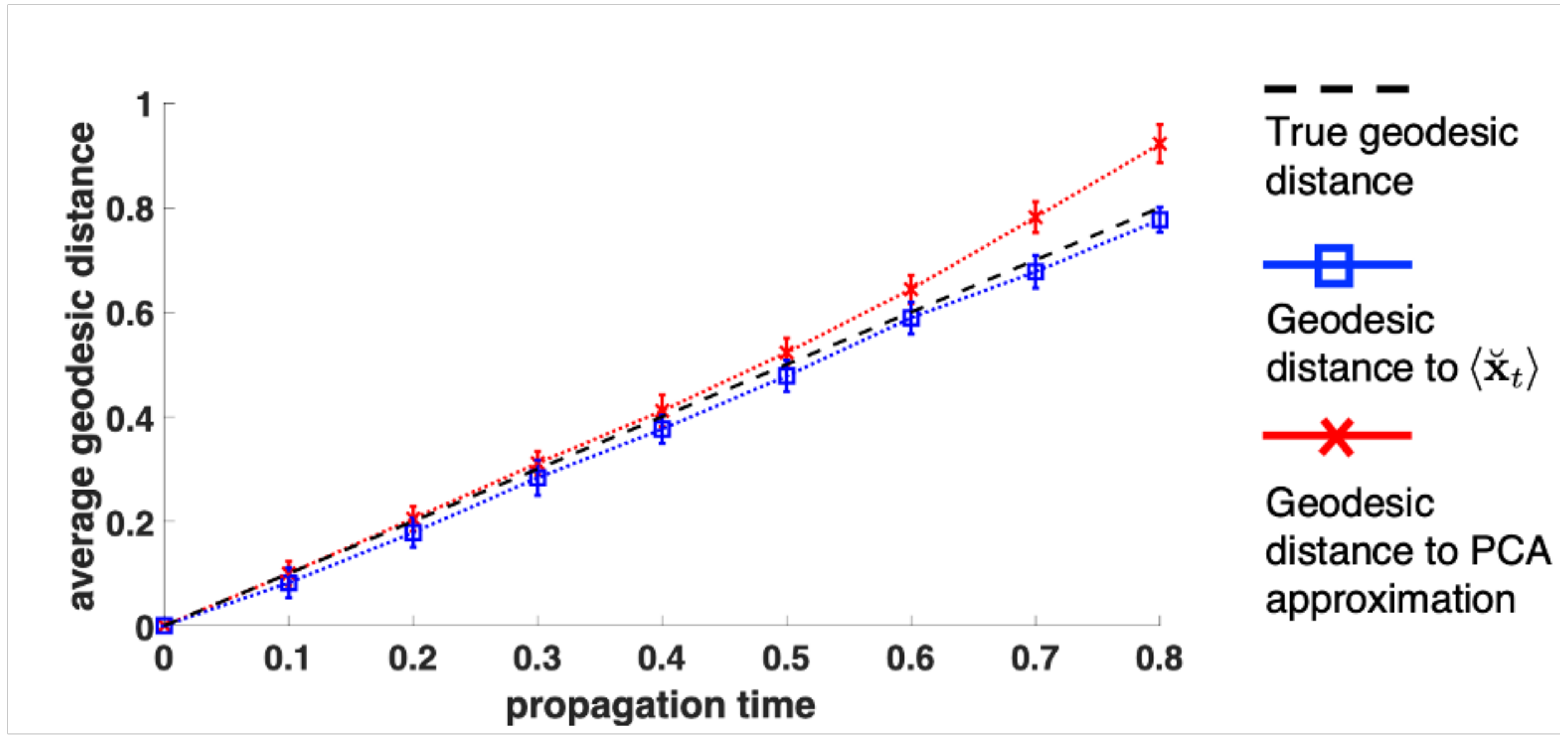


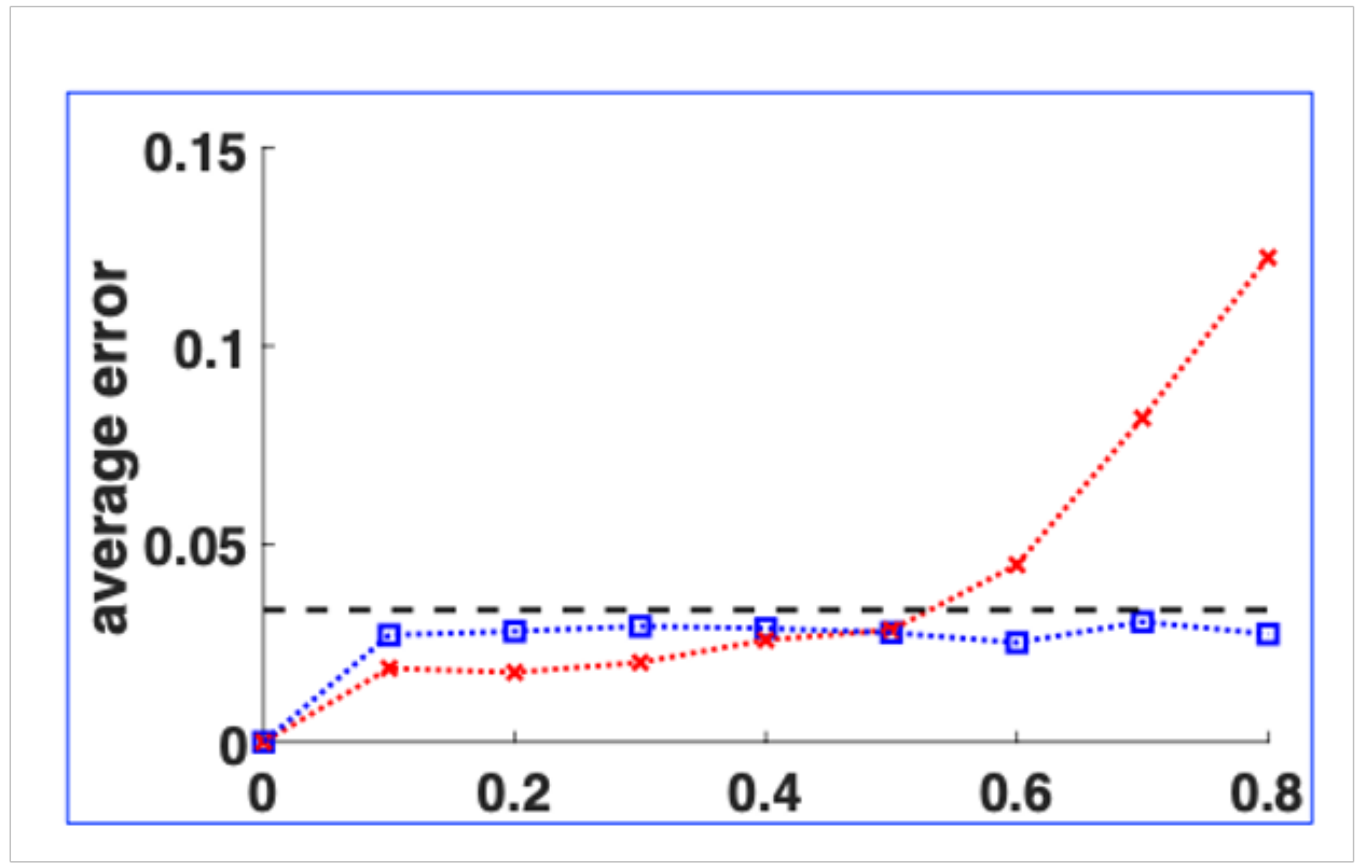


The solution



$N = 3000$ points
uniformly sampled on unit sphere

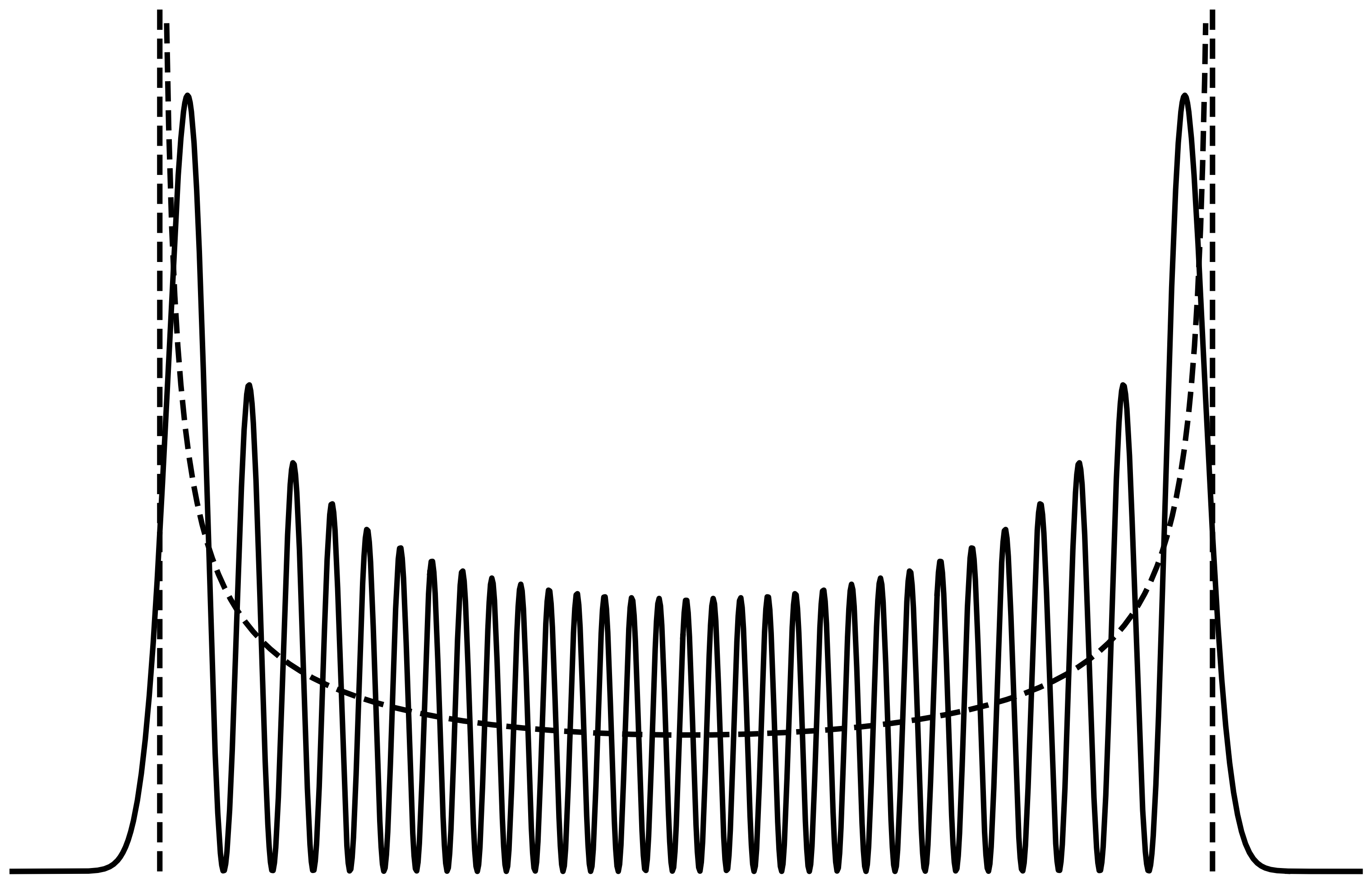


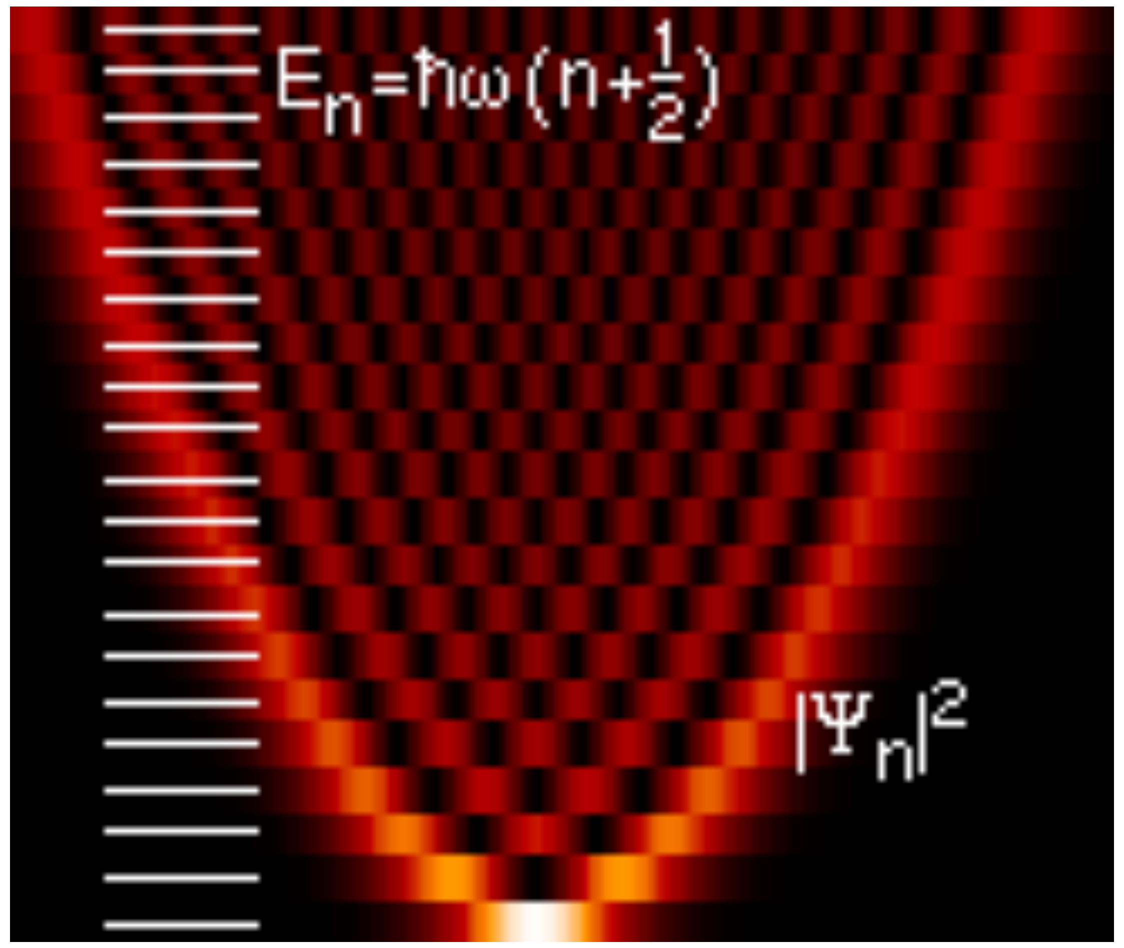


Quantum-classical correspondence

"classical mechanics emerges not directly, but only after averaging over phase-scrambling effects that can be ascribed to the environment […] — in modern parlance, decoherence effects."

Berry, *Quantum Mechanics: Scientific perspectives on divine action*, 41 (2001)

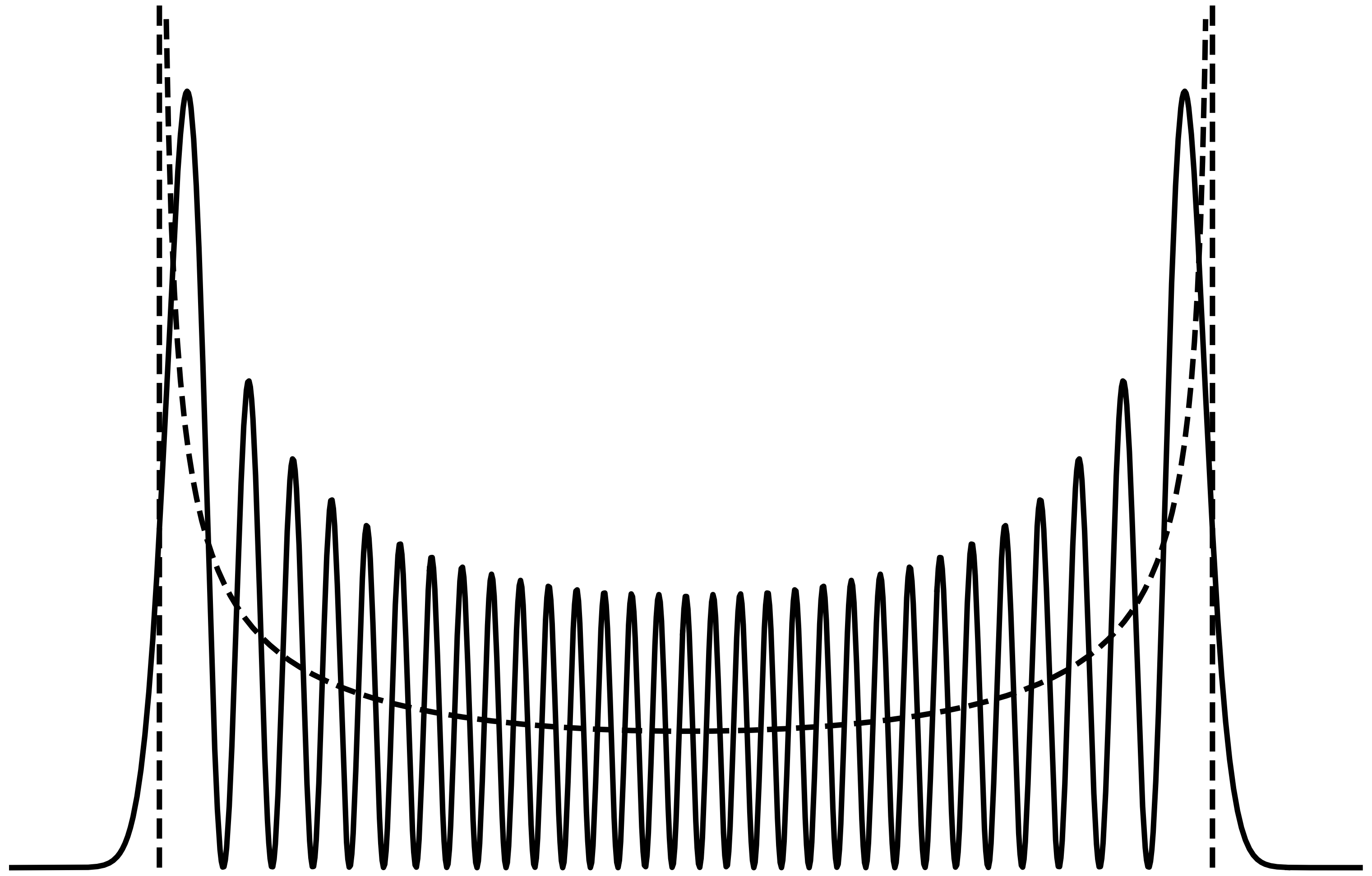


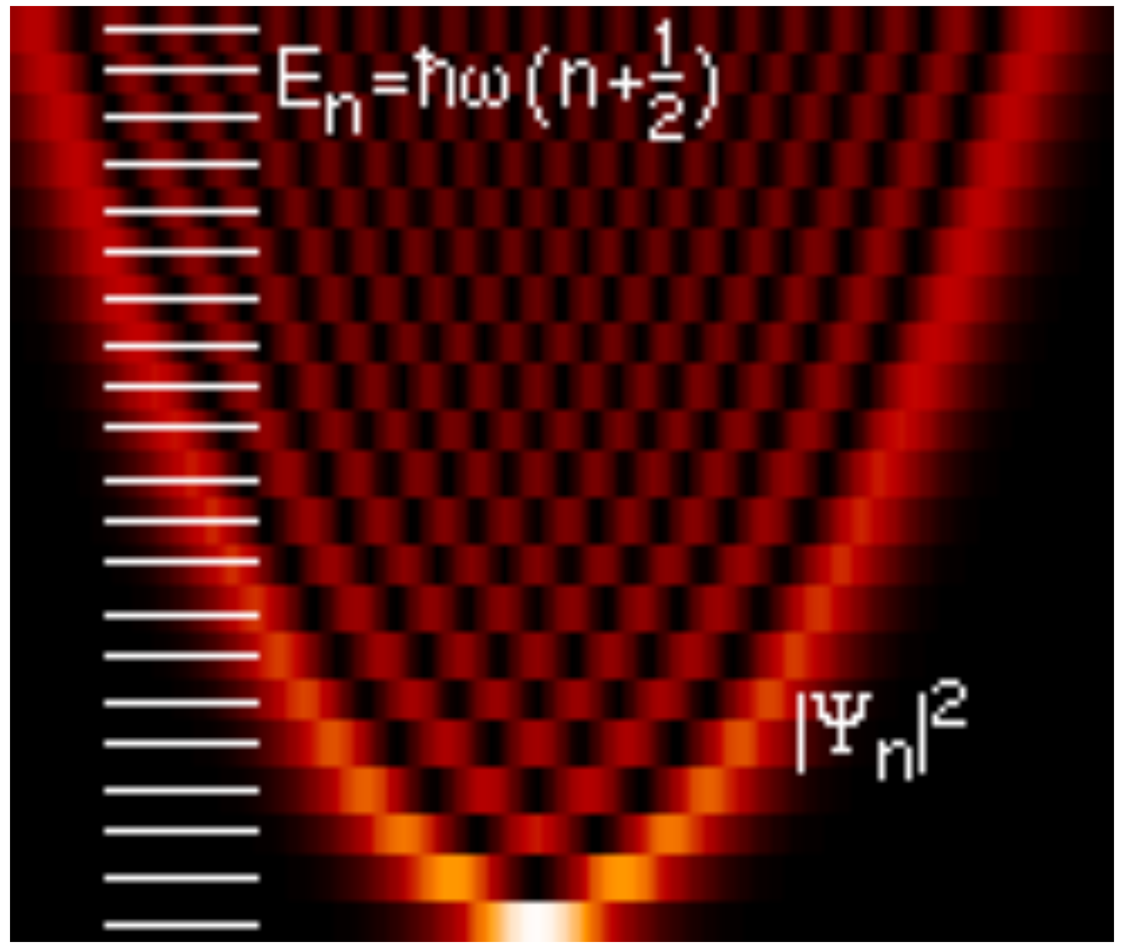


Quantum-classical correspondence

"classical mechanics emerges not directly, but only after averaging over phase-scrambling effects that can be ascribed to the environment […] — in modern parlance, decoherence effects."

Berry, *Quantum Mechanics: Scientific perspectives on divine action*, 41 (2001)





Quantum-classical correspondence

"classical mechanics emerges not directly, but only after averaging over phase-scrambling effects that can be ascribed to the environment [...] — in modern parlance, decoherence effects."

Berry, *Quantum Mechanics: Scientific perspectives on divine action*, 41 (2001)

Quantum dynamics:

linear in $|\psi_h\rangle$

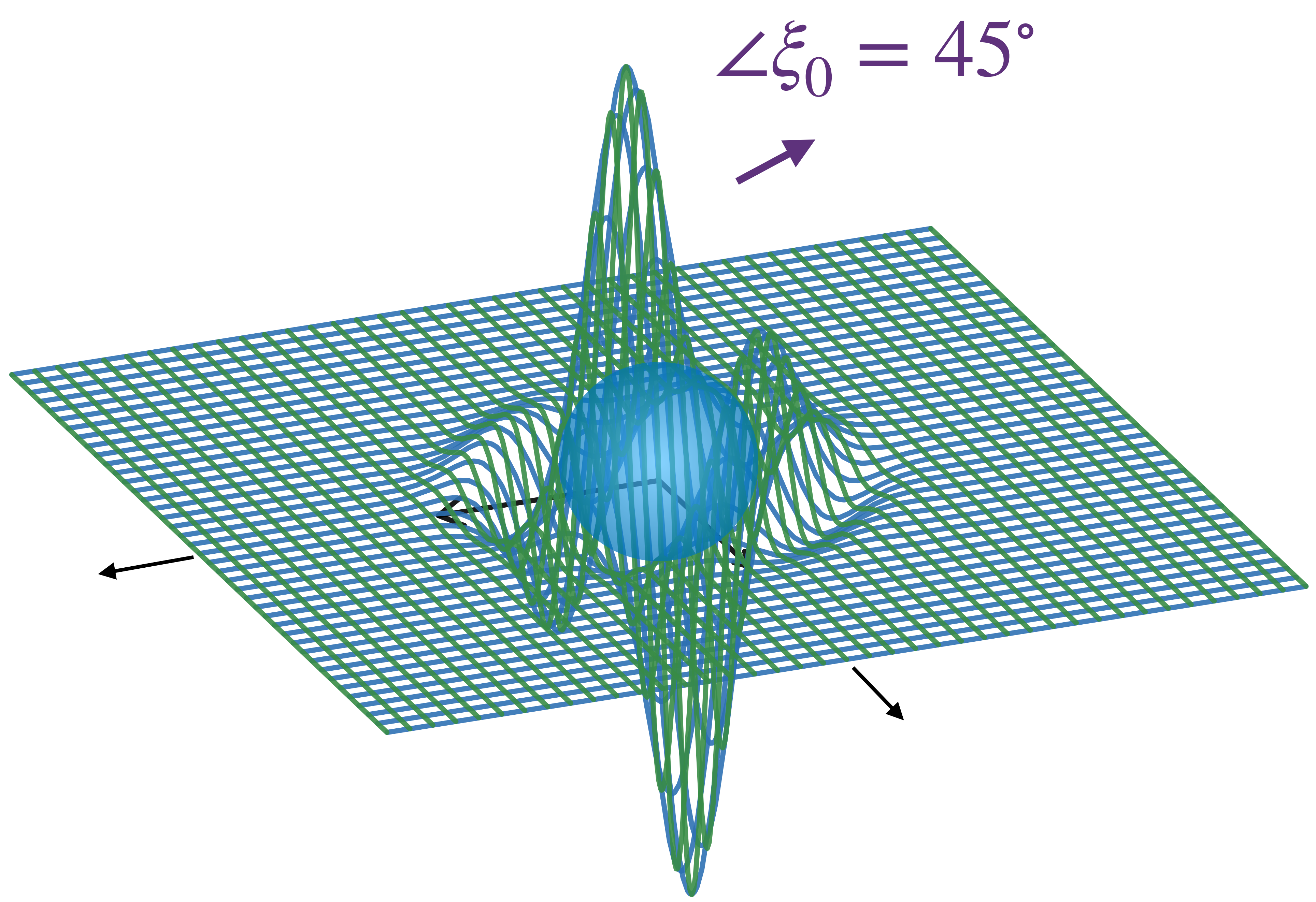
$$ih\partial_t|\psi_h\rangle = \sqrt{\hat{H}_h}\,|\psi_h\rangle$$

Geodesic flow:

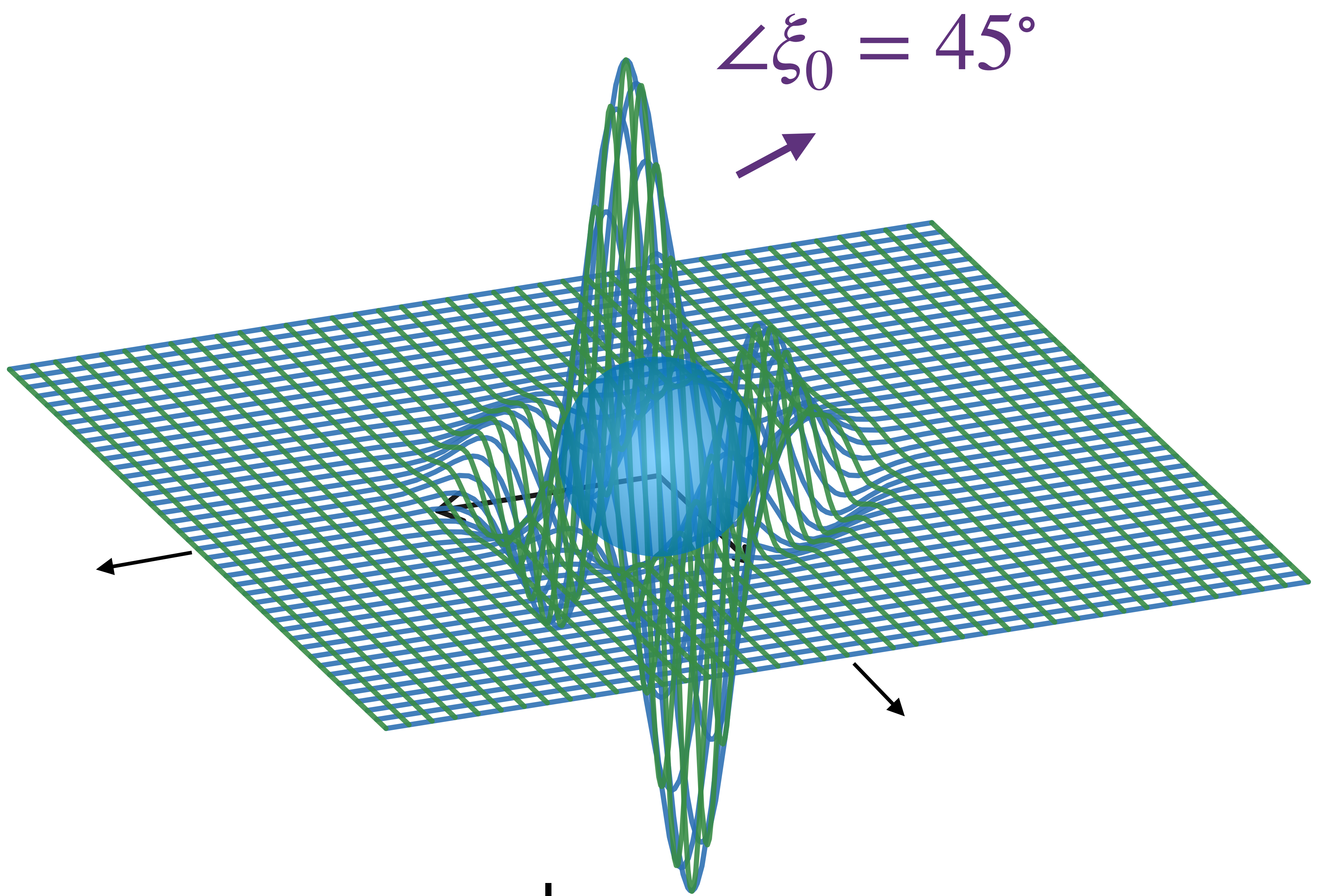
nonlinear in x, ξ

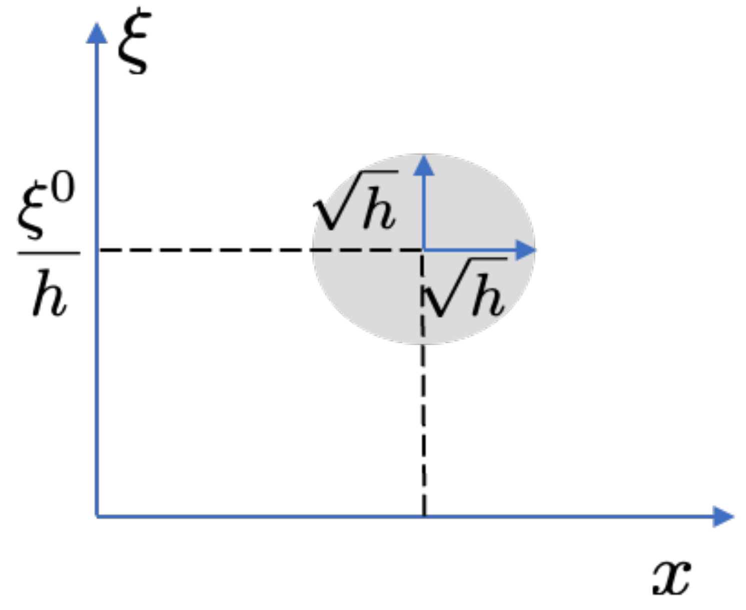
$$\begin{cases} \partial_t x_j &= \partial_{\xi_j} H = (g^{-1}(x)\cdot\xi)_j \\ \partial_t \xi_j &= -\partial_{x_j} H = \langle \partial_{x_j} g^{-1}(x)\cdot\xi, \xi\rangle \end{cases}$$

Quantizing particles



Quantizing particles

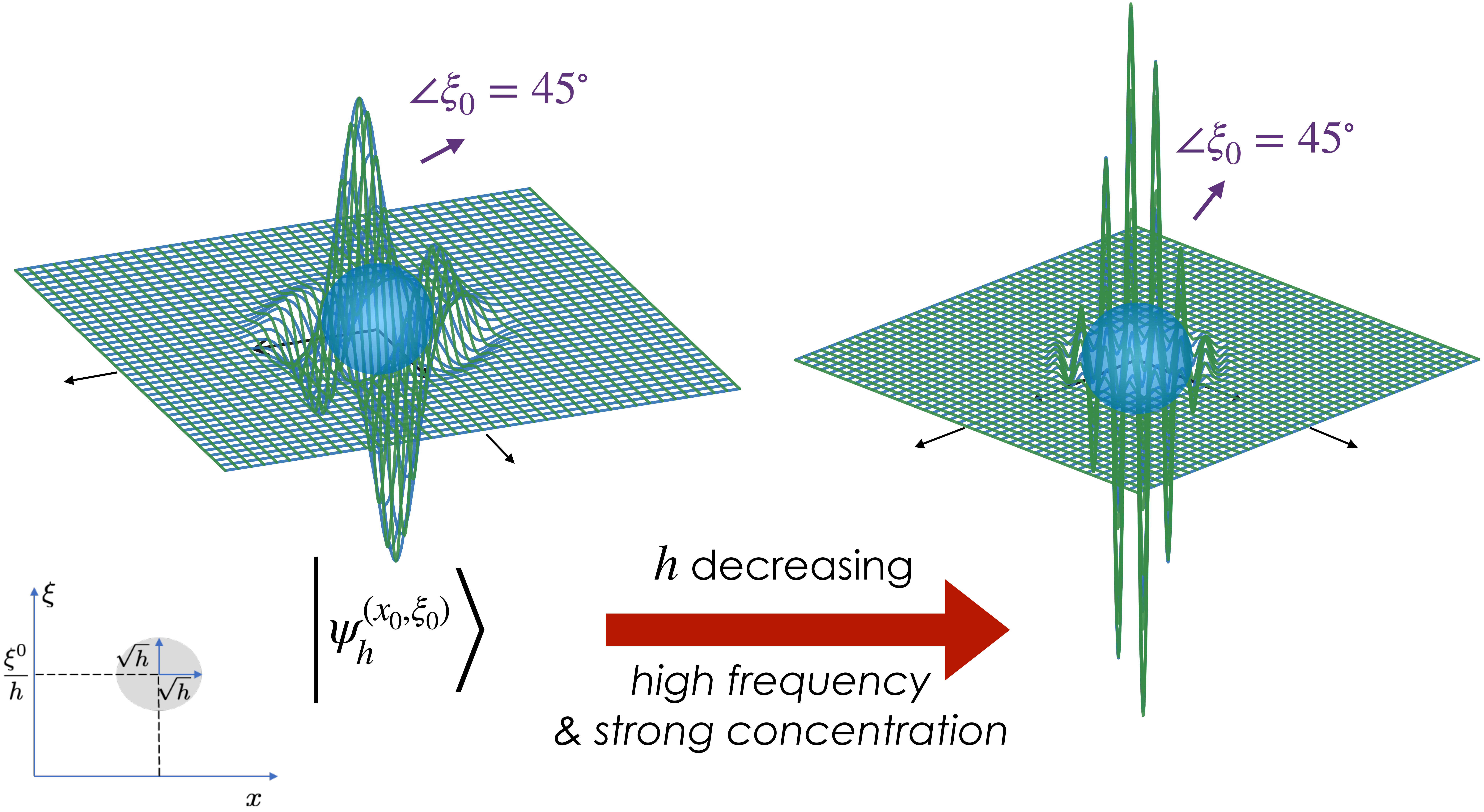




$$\left| \psi_h^{(x_0,\xi_0)} \right\rangle$$

$$\psi_h^{(x_0,\xi_0)}(x) \propto e^{-\frac{1}{2h}||x-x_0||^2-\frac{i}{h}(x-x_0)\cdot\xi_0}$$

Quantizing particles



$$\psi_h^{(x_0,\xi_0)}(x) \propto e^{-\frac{1}{2h}||x-x_0||^2-\frac{i}{h}(x-x_0)\cdot\xi_0}$$

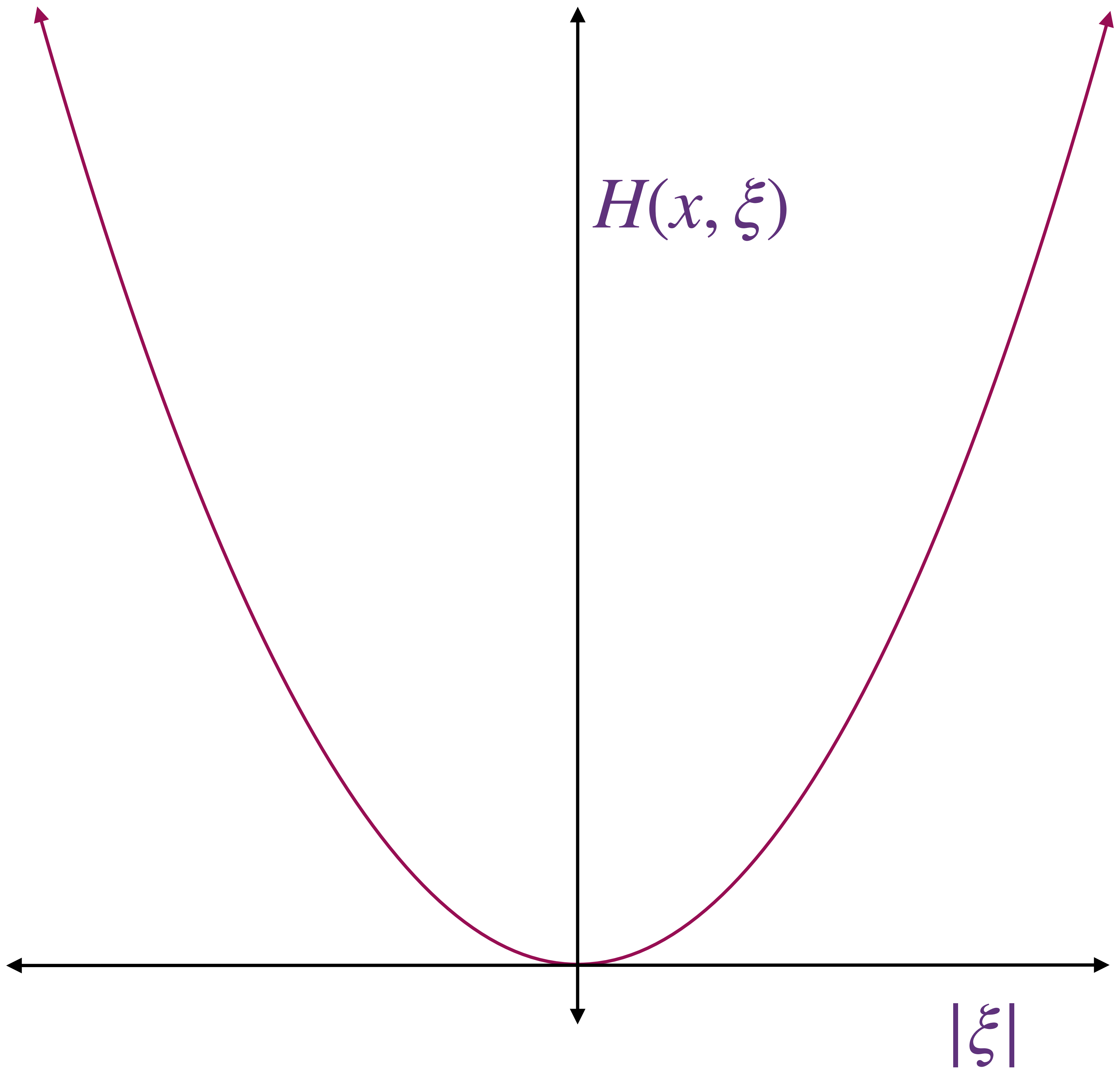
Quantizing particles

$$\psi_h^{(x_0,\xi_0)}(x) \propto e^{-\frac{1}{2h}||x-x_0||^2-\frac{i}{h}(x-x_0)\cdot\xi_0}$$

Quantizing the geodesic flow

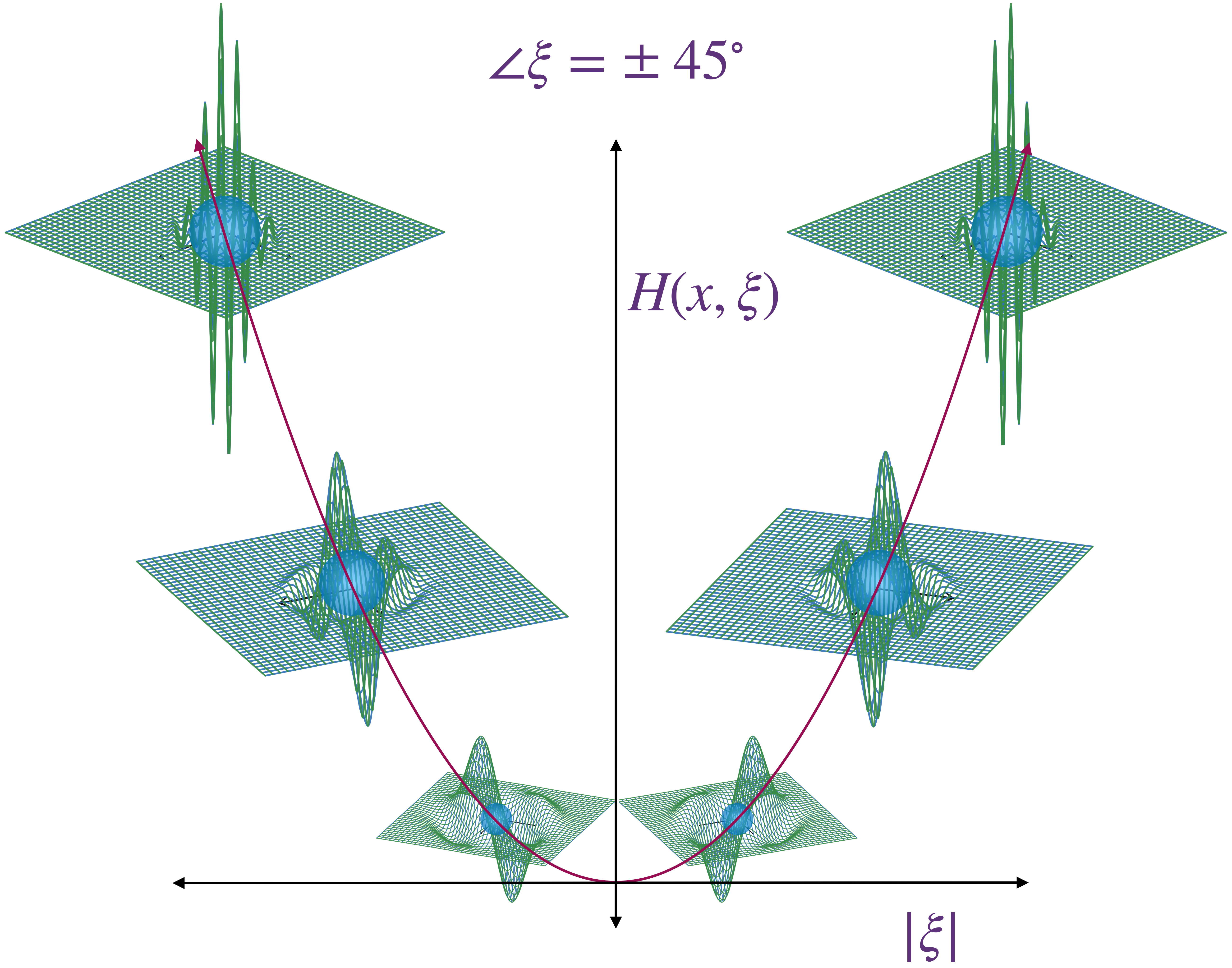
Hamiltonian flow governed by $H(x, \xi) := |\xi|^2_{g(x)}$

$\angle\xi = \pm 45^\circ$



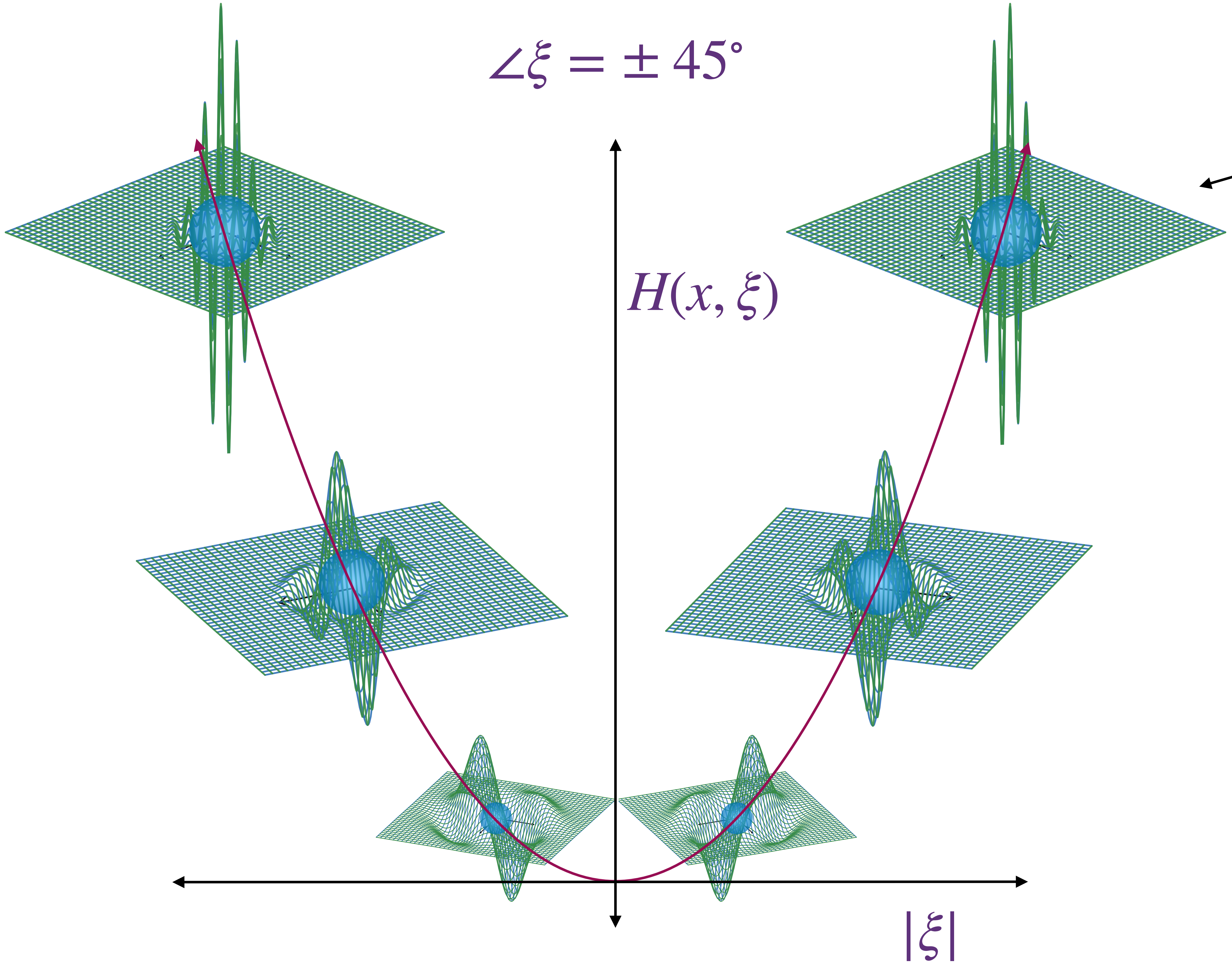
Quantizing the geodesic flow

Hamiltonian flow governed by $H(x,\xi) := |\xi|^2_{g(x)}$



Quantizing the geodesic flow

Hamiltonian flow governed by $H(x,\xi) := |\xi|^2_{g(x)}$



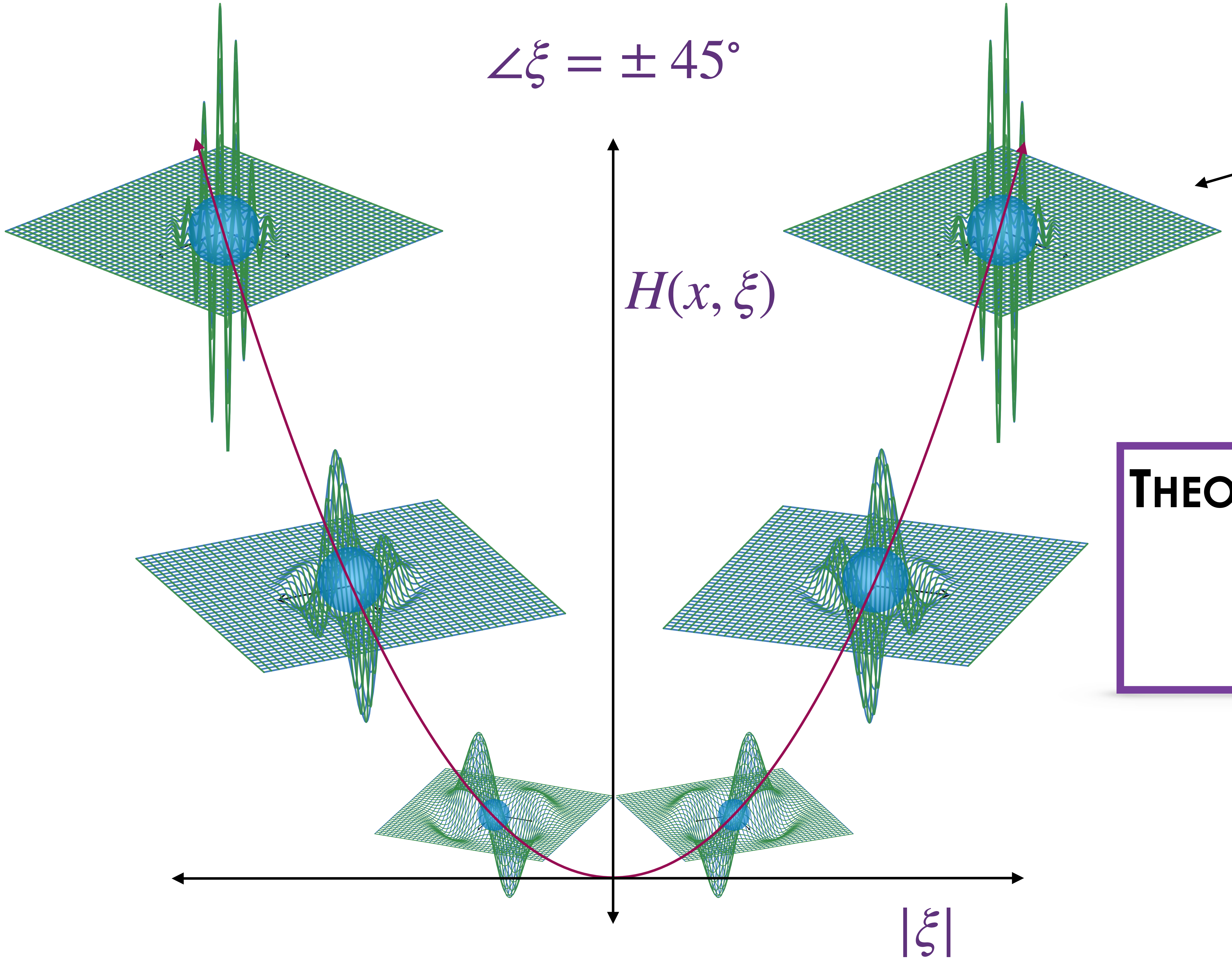
$$\hat{H}_h = \sum_{x,\xi} H(x,\xi) \left|\psi_h^{(x,\xi)}\right\rangle\left\langle\psi_h^{(x,\xi)}\right|$$

$$\langle\psi_h^{(x,\xi)}|\psi_h^{(x,\xi)}\rangle = 1 + O(h^\infty)$$

$$\implies \langle\psi_h^{(x,\xi)}|\hat{H}_h|\psi_h^{(x,\xi)}\rangle = H(x,\xi) + O(h)$$

Quantizing the geodesic flow

Hamiltonian flow governed by $H(x, \xi) := |\xi|^2_{g(x)}$



$$\hat{H}_h = \sum_{x,\xi} H(x,\xi) \left| \psi_h^{(x,\xi)} \right\rangle \left\langle \psi_h^{(x,\xi)} \right|$$

Theorem.

$$\hat{H}_h = h^2 \Delta_{\mathscr{M}} + \text{lower oder terms}$$

$$\langle \psi_h^{(x,\xi)} | \psi_h^{(x,\xi)} \rangle = 1 + O(h^\infty)$$

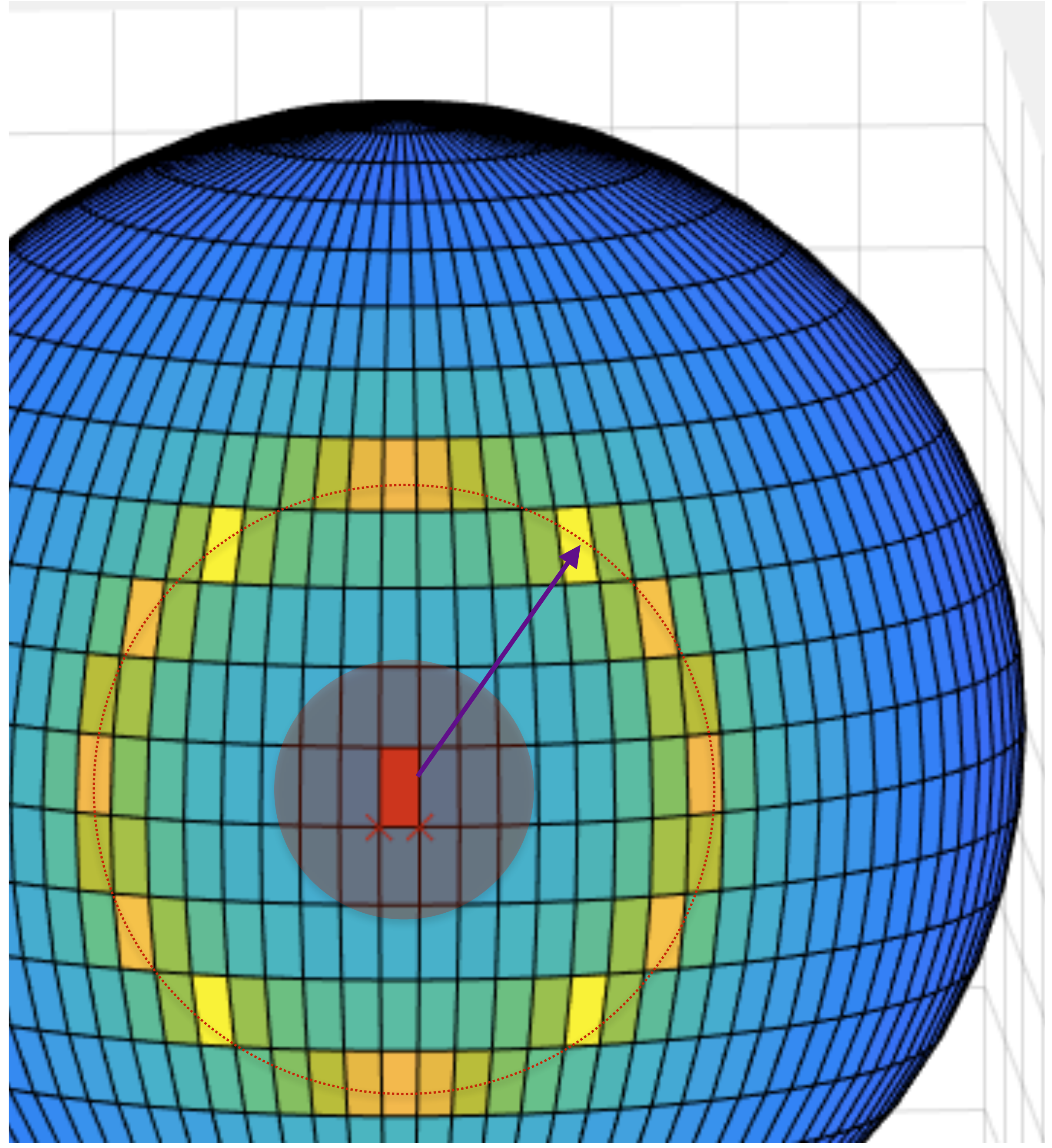
$$\implies \langle \psi_h^{(x,\xi)} | \hat{H}_h | \psi_h^{(x,\xi)} \rangle = H(x,\xi) + O(h)$$

Quantizing the geodesic flow

Schrödinger equation $ih\partial_t |\psi_h\rangle = \sqrt{\hat{H}_h}\,|\psi_h\rangle$

Quantizing the geodesic flow

Schrödinger equation $ih\partial_t|\psi_h\rangle = \sqrt{\hat{H}_h}\,|\psi_h\rangle$

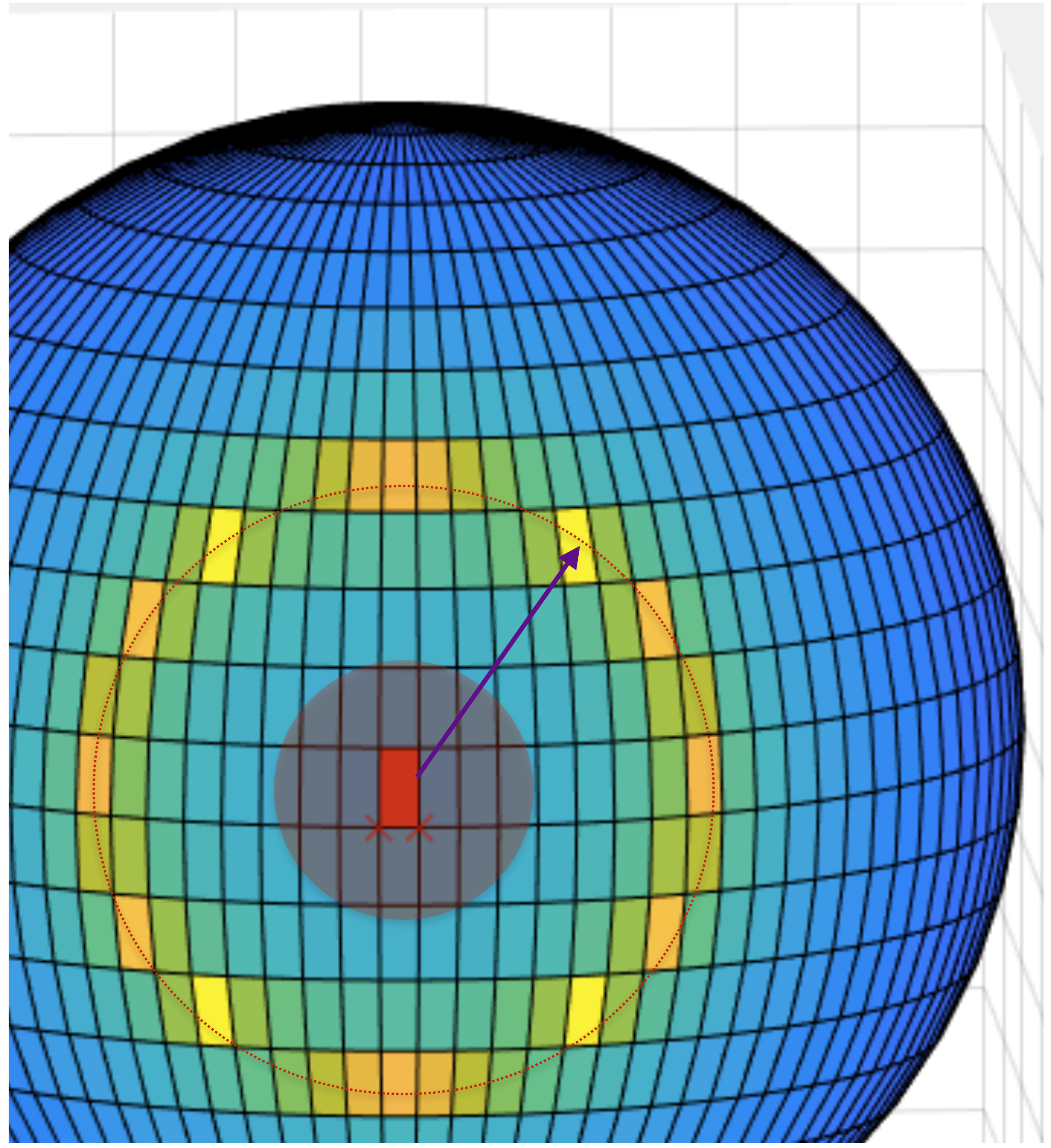


Initial state

$$\left|\psi_h^{(x_0,\xi_0)}\right\rangle$$

Quantizing the geodesic flow

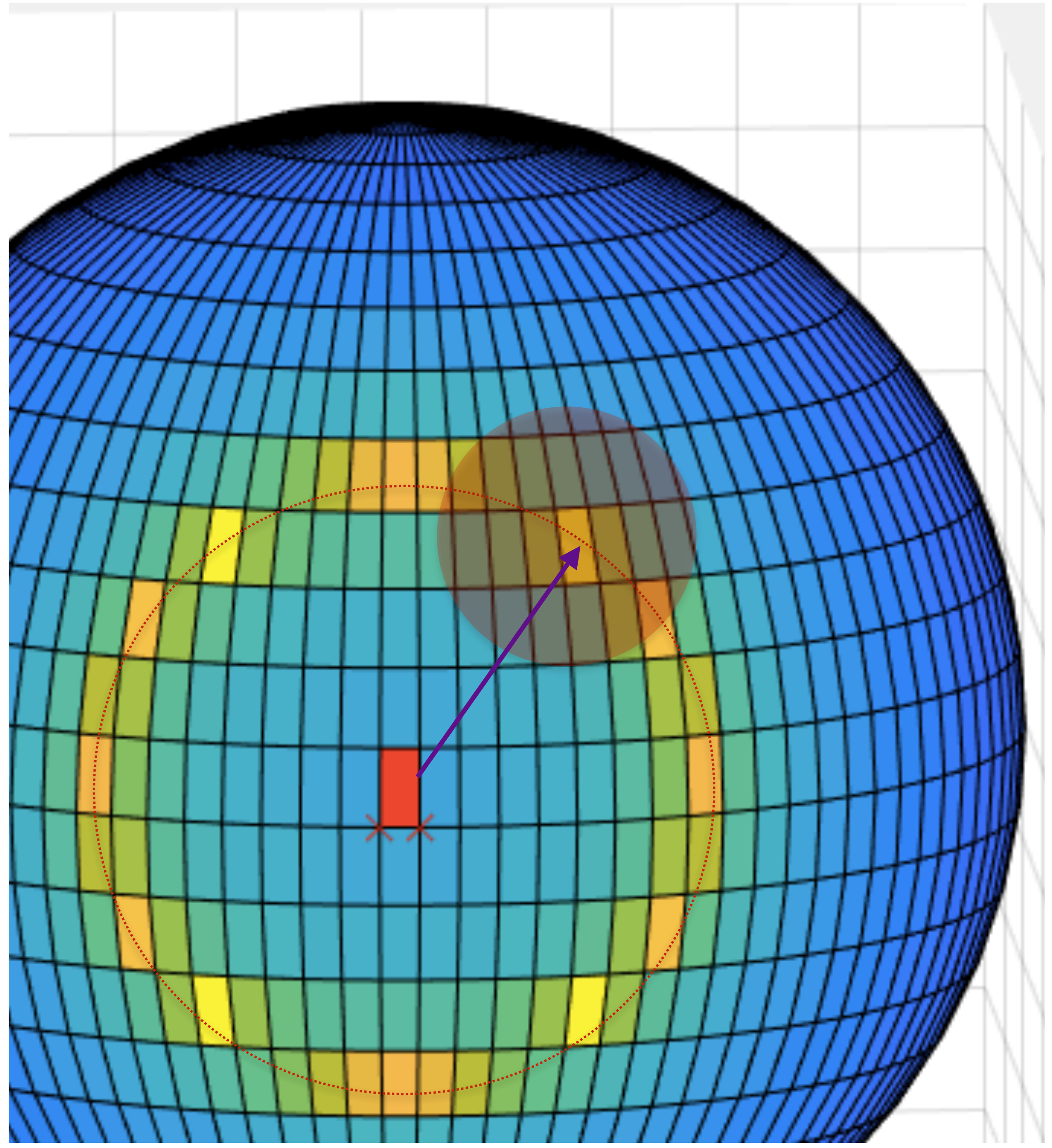
Schrödinger equation $ih\partial_t |\psi_h\rangle = \sqrt{\hat{H}_h} \left|\psi_h\right\rangle$



Initial state

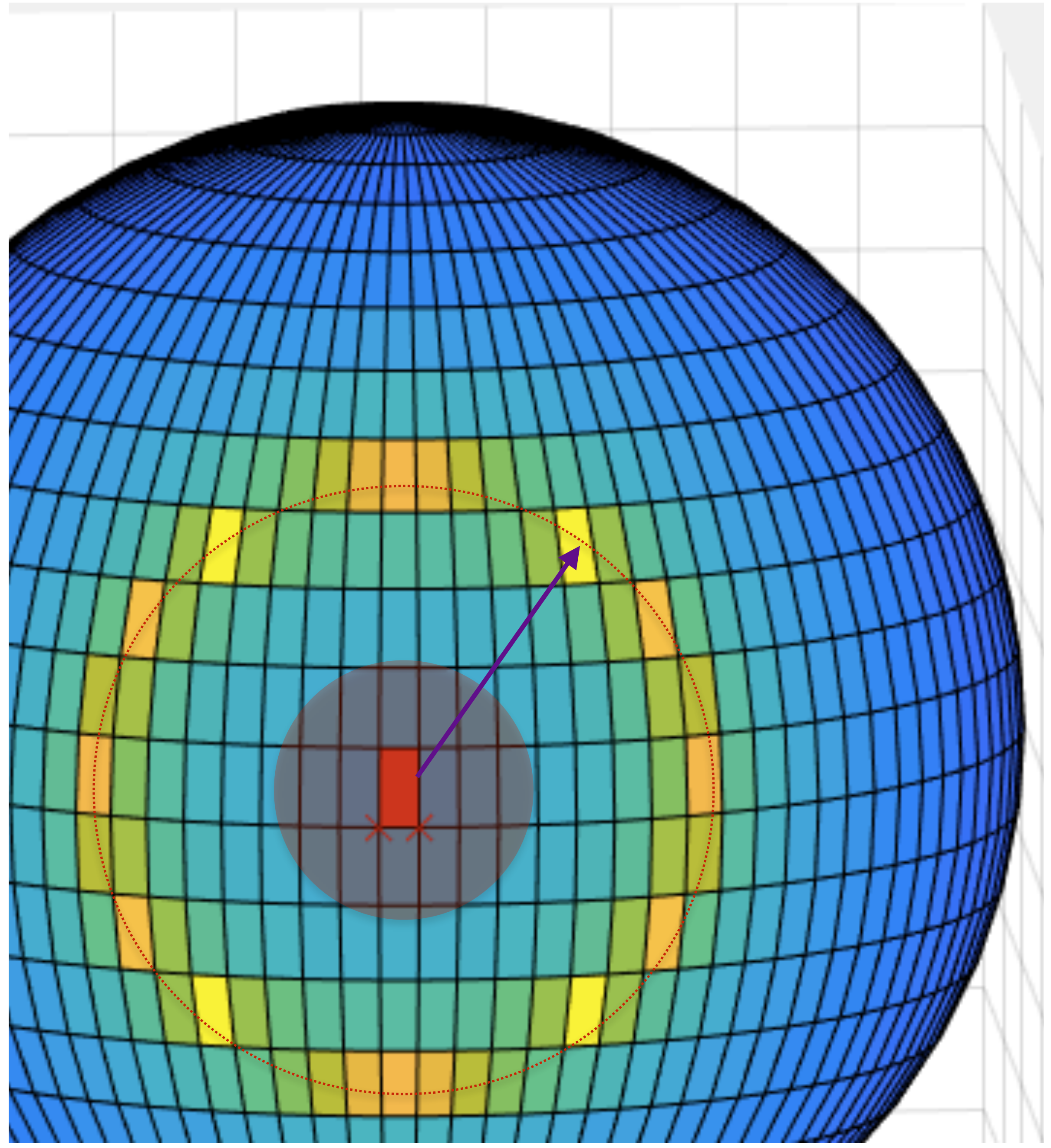
Propagated state

$$\left|\psi_h^{(x_0,\xi_0)}\right\rangle \longrightarrow \left|\psi_h(t)\right\rangle := e^{\frac{i}{h}t\sqrt{\hat{H}_h}} \left|\psi_h^{(x_0,\xi_0)}\right\rangle$$



Quantizing the geodesic flow

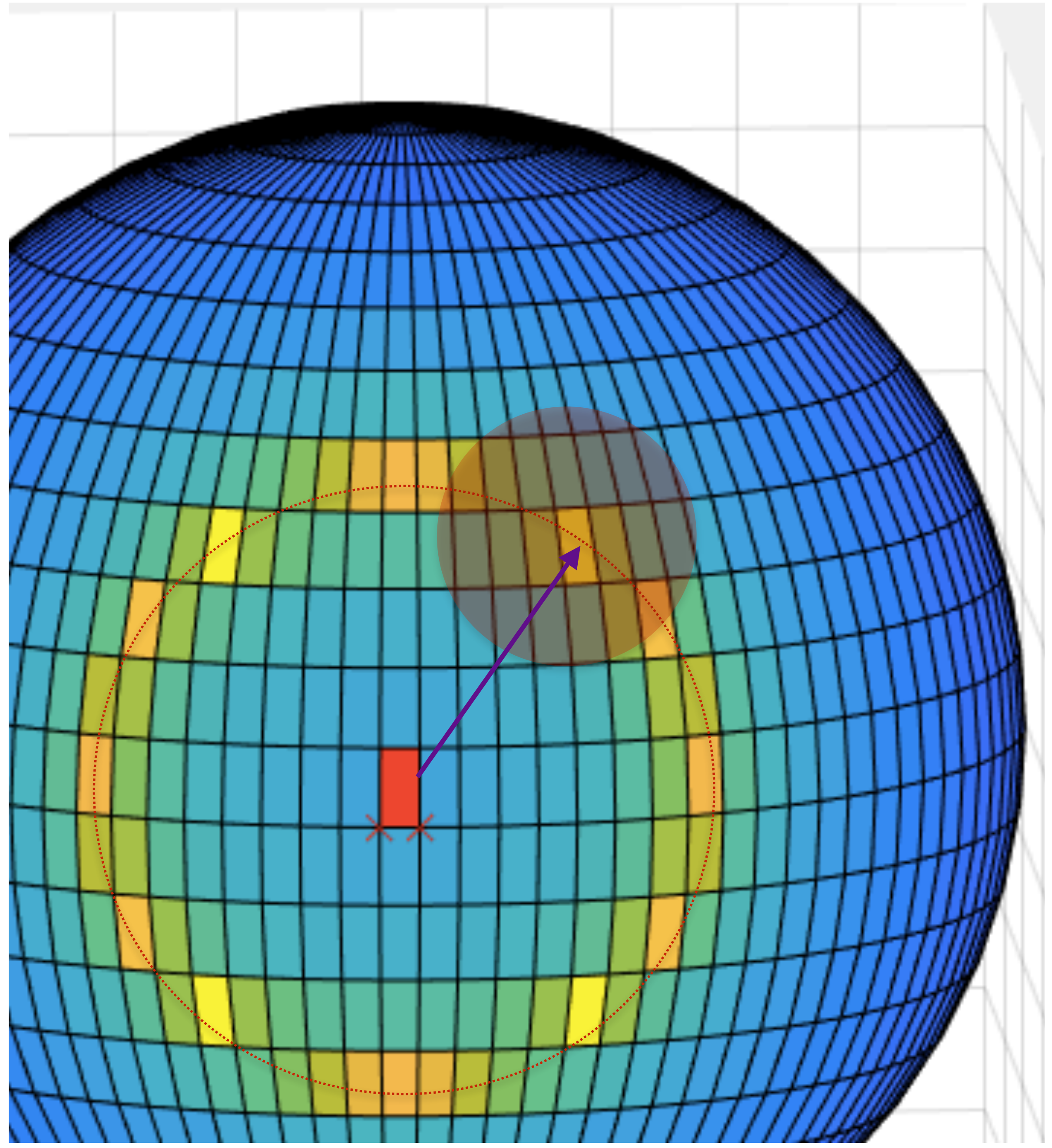
Schrödinger equation $ih\partial_t|\psi_h\rangle = \sqrt{\hat{H}_h}\,|\psi_h\rangle$



Initial state

$$\left|\psi_h^{(x_0,\xi_0)}\right\rangle \longrightarrow \left|\psi_h(t)\right\rangle := e^{\frac{i}{h}t\sqrt{\hat{H}_h}}\left|\psi_h^{(x_0,\xi_0)}\right\rangle$$

Propagated state



THEOREM.

The density $|\psi_h(t)|^2$ is concentrated in an $O(h)$ nbd. of $x(t) := \pi_{\mathscr{M}}\Gamma^t(x_0,\xi_0)$.

Discrete quantum-classical correspondence

“If you want to see something, you send waves in its general direction, you don’t throw heat at it.”

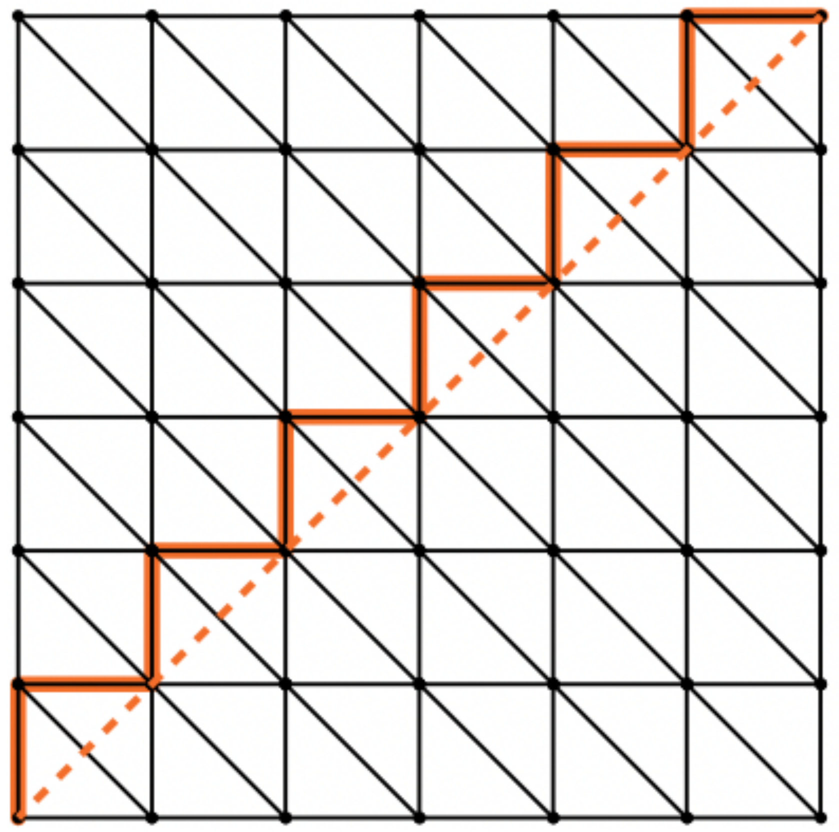
Attributed to Peter Lax (Cloninger & Steinerberger, *Applied & Computational Harmonic Analysis* (2017)

Random walks on data lead to diffusion

Random walk:

1. Nbd. graph:

$$\left[\mathscr{A}_{\epsilon,N}\right]_{i,j} : \begin{cases} \approx 1 & \text{if } |v_j - v_i| \ll \sqrt{\epsilon} \\ \approx 0 & \text{if } |v_j - v_i| \gg \sqrt{\epsilon} \end{cases}$$



Crane et al., Comm. ACM (2017)

2. Markov chain: $A_{\epsilon,N} := \frac{1}{\mathscr{A}_{\epsilon,N}[1]}\mathscr{A}_{\epsilon,N}$

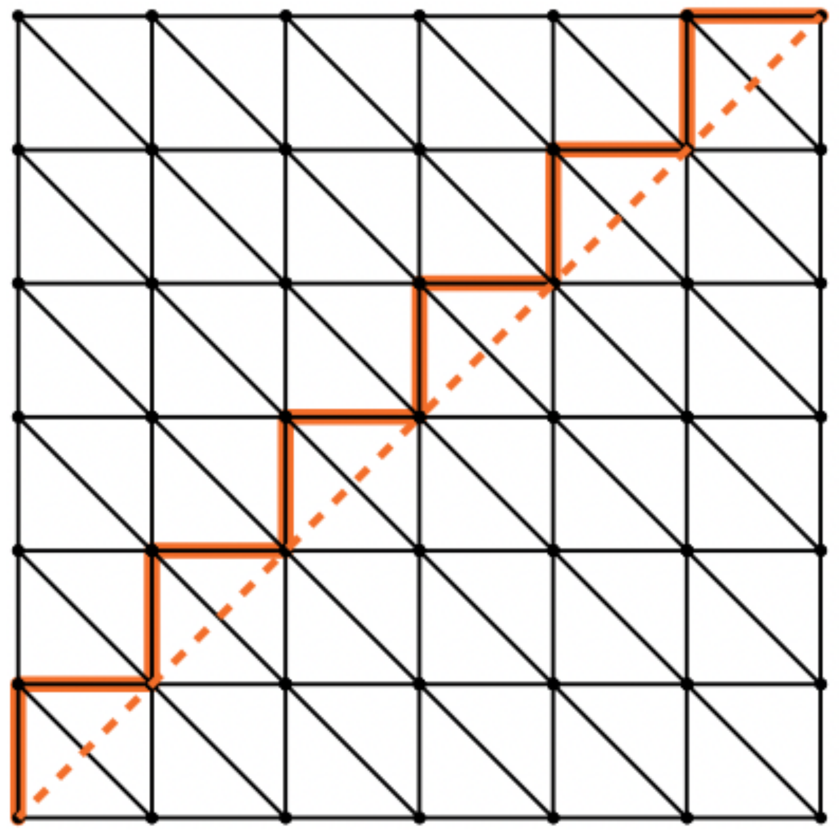
Random walks on data lead to diffusion

Random walk:

1. Nbd. graph:

$$[\mathscr{A}_{\epsilon,N}]_{i,j} : \begin{cases} \approx 1 & \text{if } |v_j - v_i| \ll \sqrt{\epsilon} \\ \approx 0 & \text{if } |v_j - v_i| \gg \sqrt{\epsilon} \end{cases}$$

2. Markov chain: $A_{\epsilon,N} := \frac{1}{\mathscr{A}_{\epsilon,N}[1]}\mathscr{A}_{\epsilon,N}$



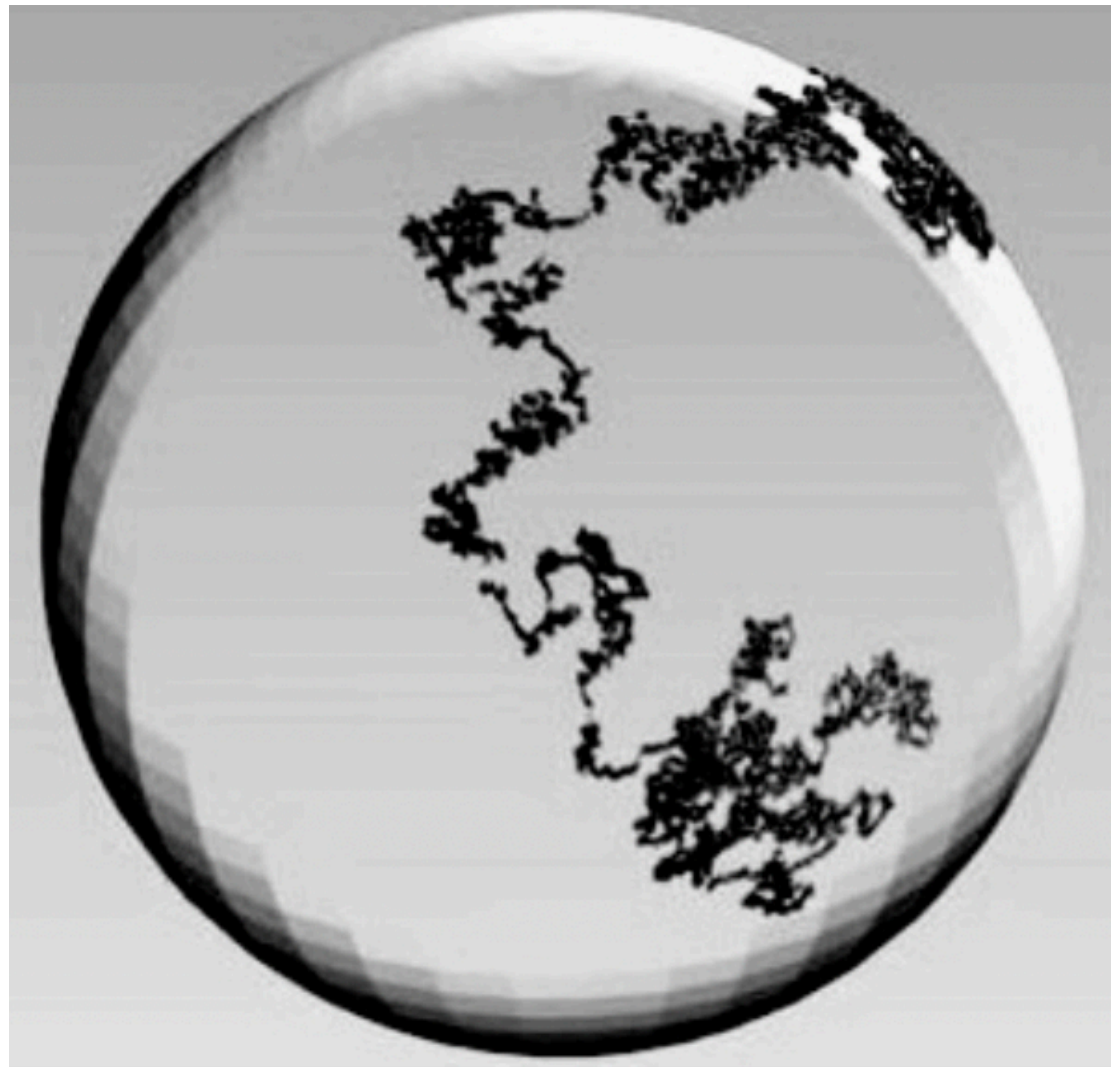
Crane et al., Comm. ACM (2017)

large N

with high prob.,

$\sqrt{\epsilon} \ll 1$

Geodesic random walk



Ginkel & Redig, *Journal of Statistical Physics* (2020).

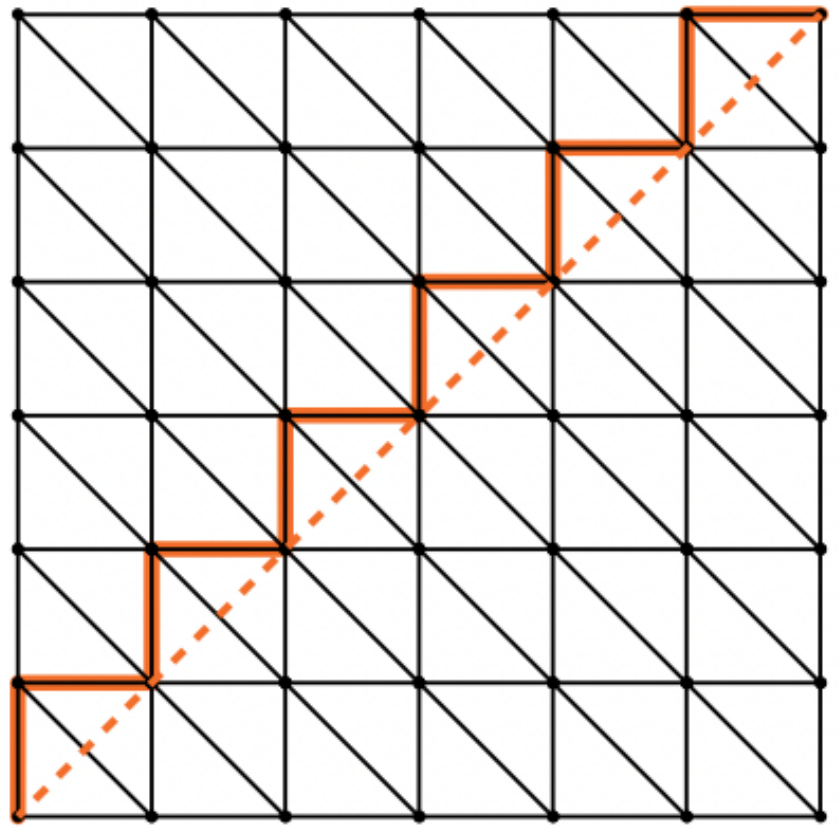
Random walks on data lead to diffusion

Random walk:

1. Nbd. graph:

$$[\mathscr{A}_{\epsilon,N}]_{i,j} : \begin{cases} \approx 1 & \text{if } |v_j - v_i| \ll \sqrt{\epsilon} \\ \approx 0 & \text{if } |v_j - v_i| \gg \sqrt{\epsilon} \end{cases}$$

2. Markov chain: $A_{\epsilon,N} := \dfrac{1}{\mathscr{A}_{\epsilon,N}[1]} \mathscr{A}_{\epsilon,N}$



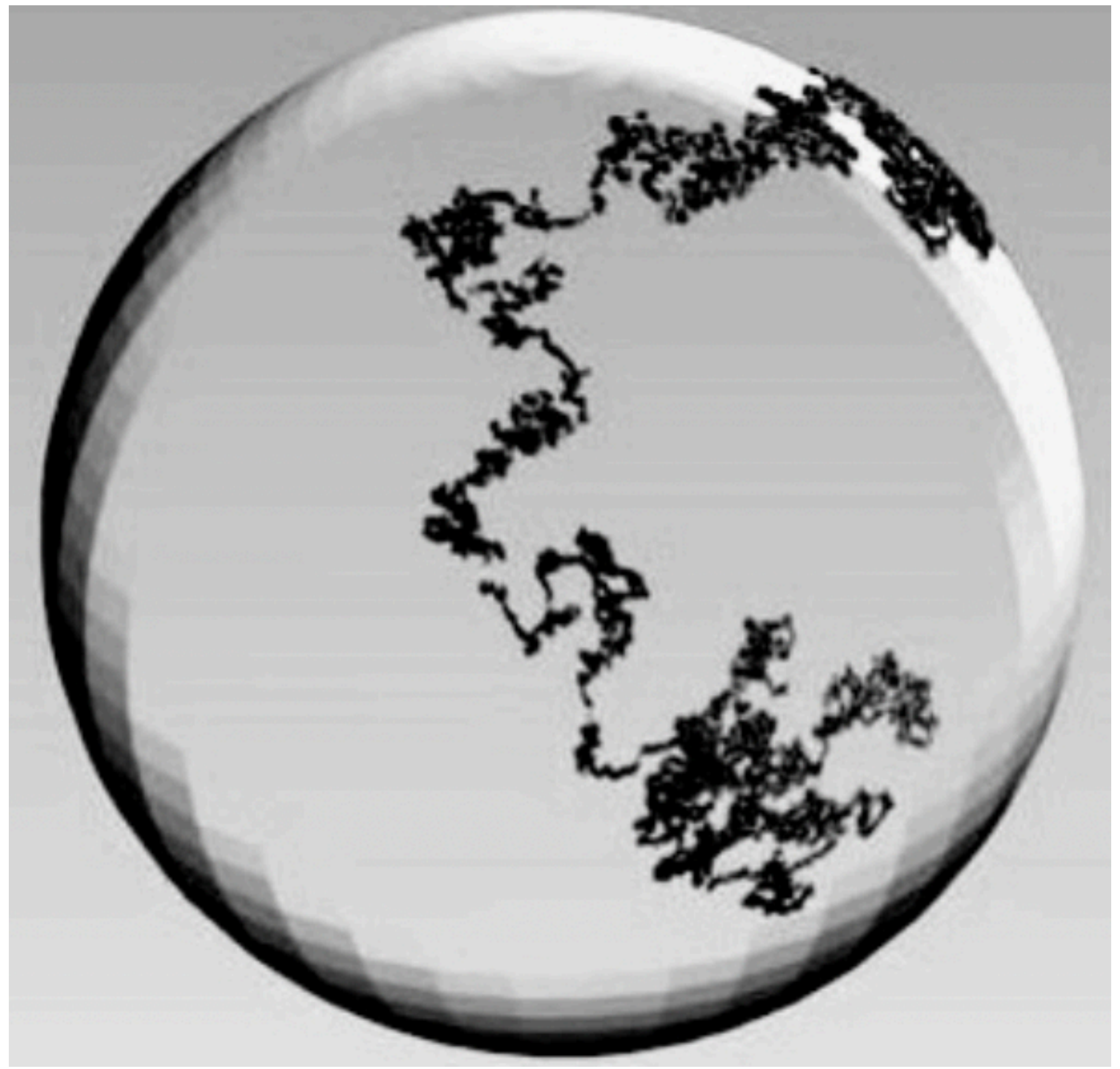
Crane et al., Comm. ACM (2017)

large N

with high prob.,

$\sqrt{\epsilon} \ll 1$

Geodesic random walk



Ginkel & Redig, *Journal of Statistical Physics* (2020).

DISCRETE GENERATOR: $\Delta_{\epsilon,N} := c(I - A_{\epsilon,N})/\epsilon$ IS THE *RANDOM WALK GRAPH* LAPLACIAN.

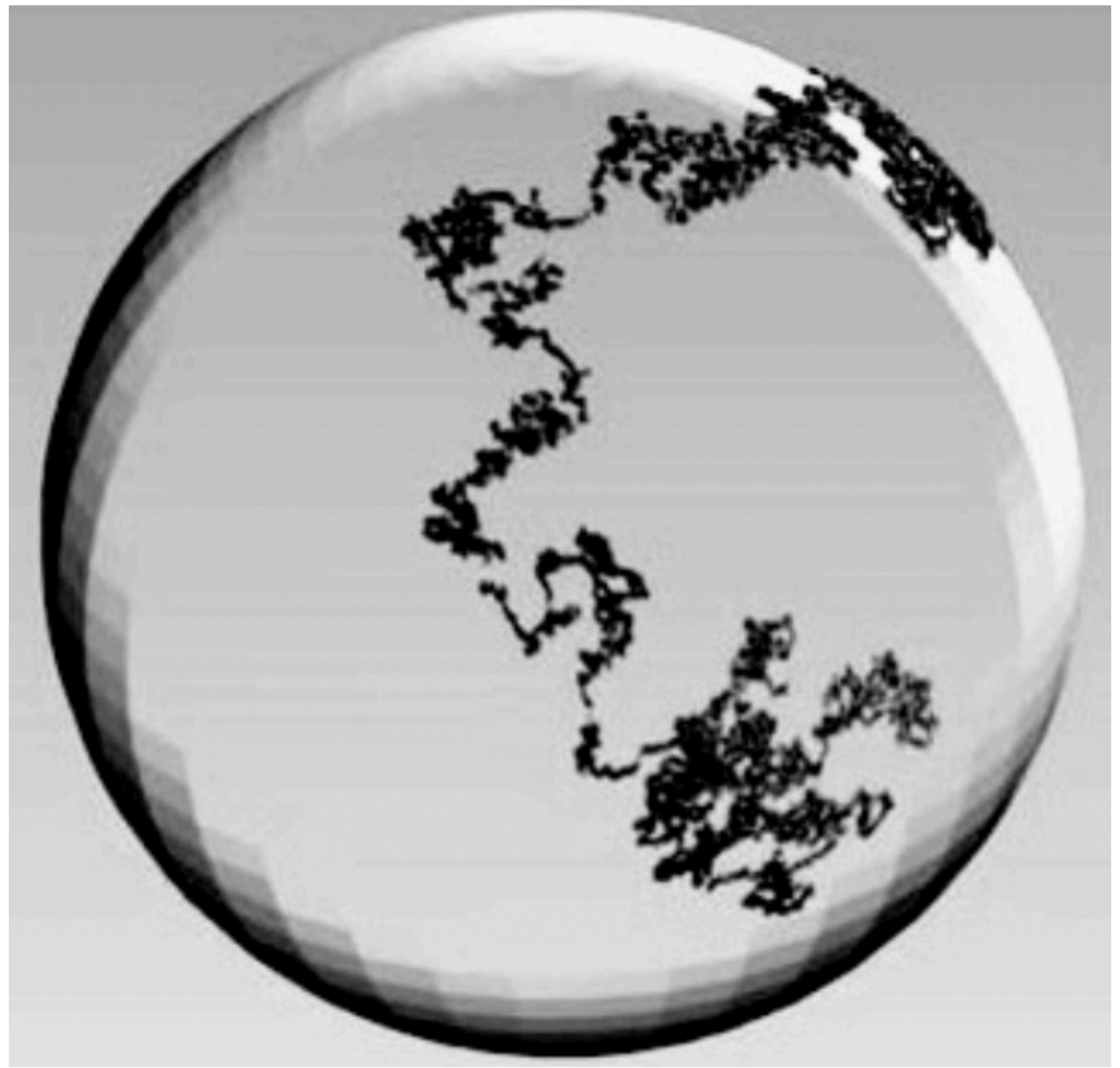
With high probability,

Hein, *PhD thesis* (2005).

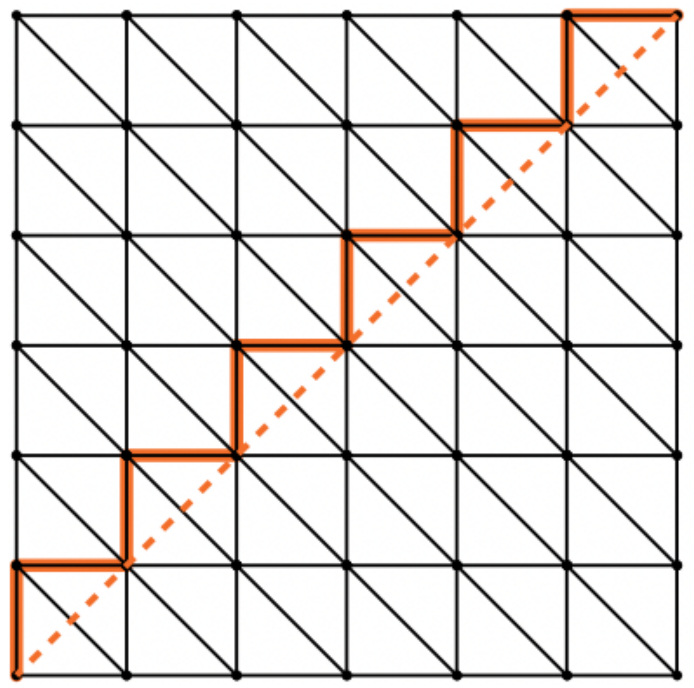
$$\Delta_{\epsilon,N}|\psi\rangle = \Delta_{\mathscr{M}}|\psi\rangle + \underbrace{O(\partial^1)}_{\text{sampling density}}|\psi\rangle + \underbrace{O(\epsilon)}_{\textit{diffusion terms}}$$

From random walks to quantum dynamics on data

Schrödinger equation $ih\partial_t|\psi_h\rangle = \sqrt{h^2\Delta_{\epsilon,N}}\,|\psi_h\rangle$



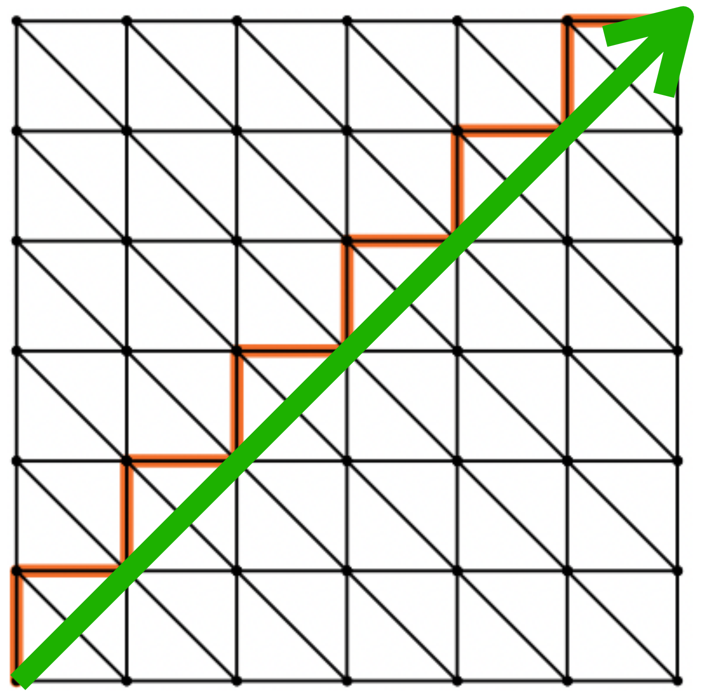
Ginkel & Redig, *Journal of Statistical Physics* (2020).

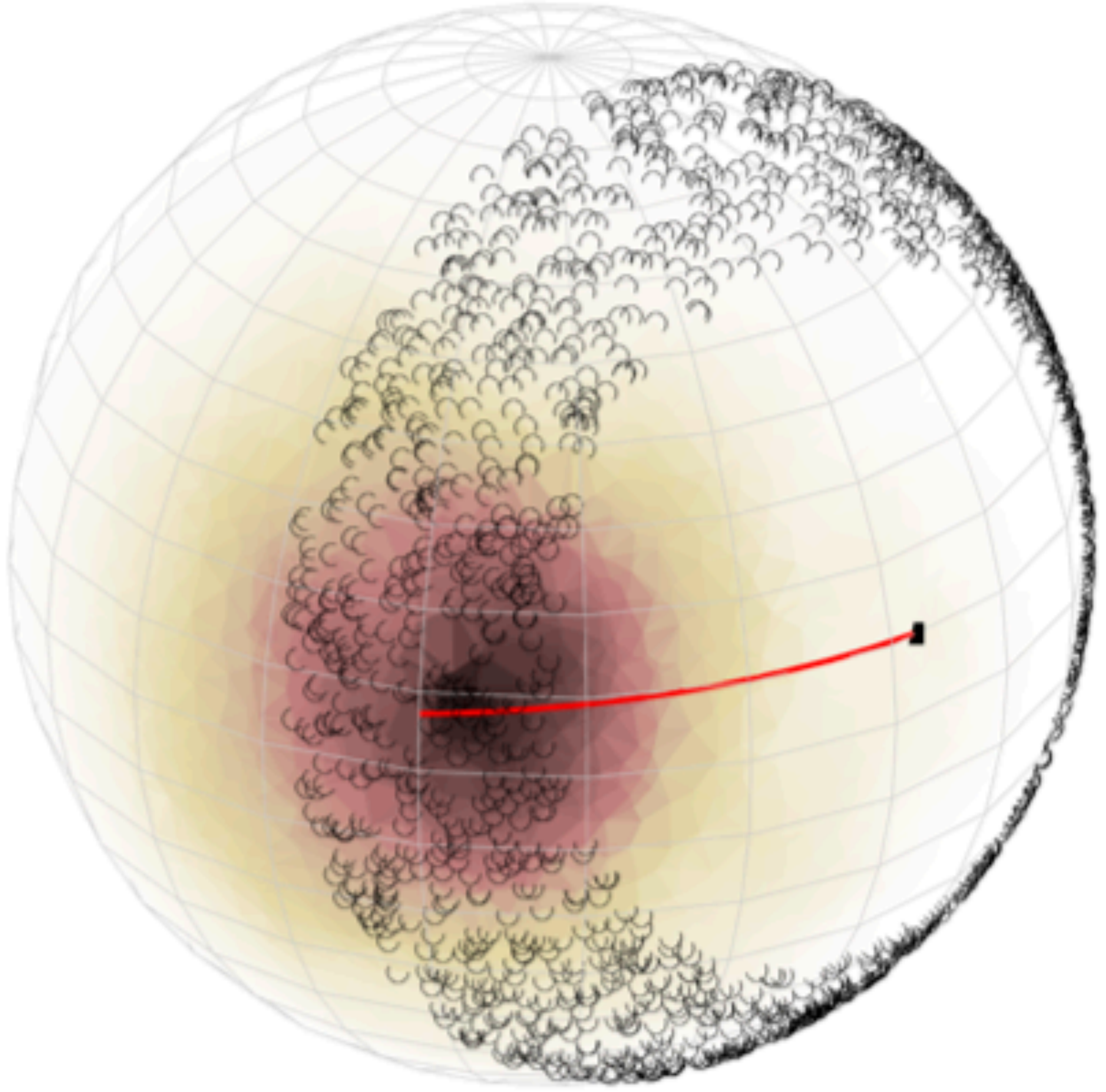


Crane et al., Comm. ACM (2017)

with high prob.,

$h \gg \sqrt{\epsilon}$

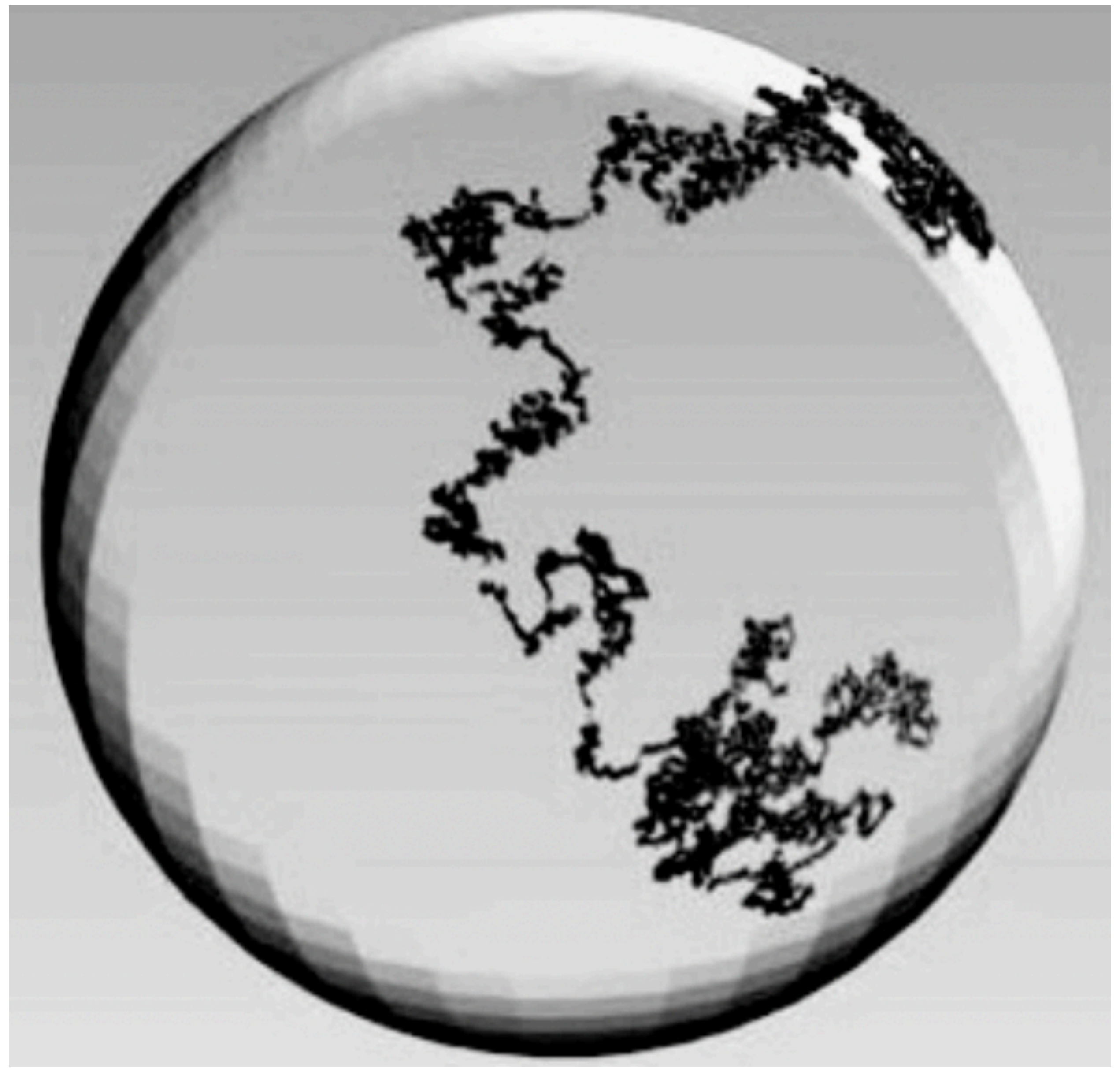




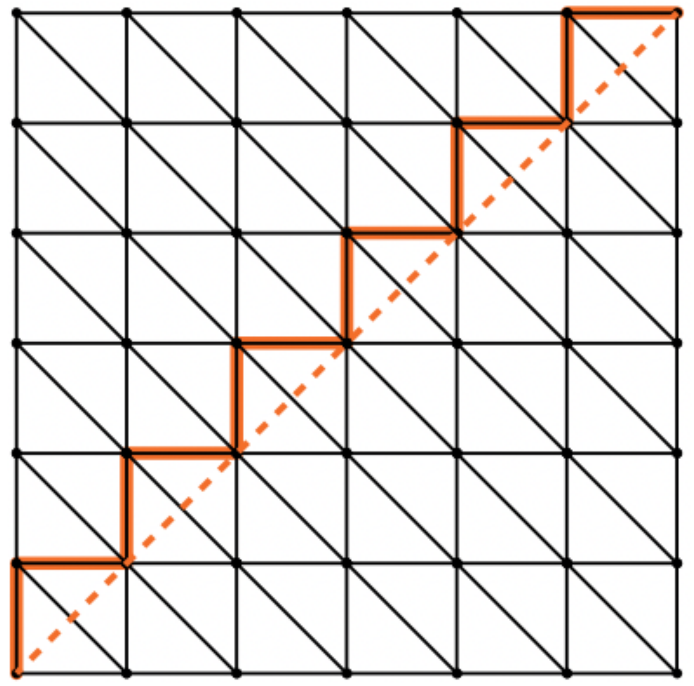
Kumar, arXiv:2112.10748 (2022)

From random walks to quantum dynamics on data

Schrödinger equation $ih\partial_t|\psi_h\rangle = \sqrt{h^2\Delta_{\epsilon,N}}\,|\psi_h\rangle$



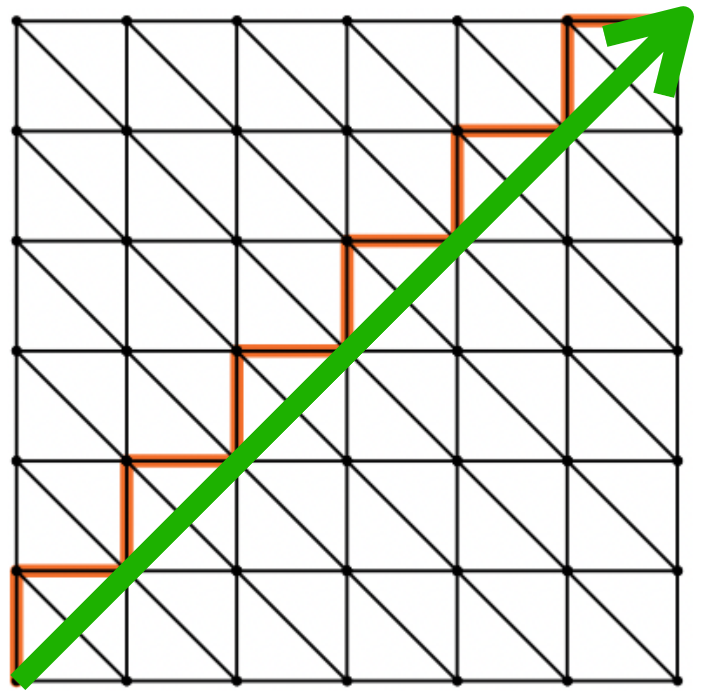
Ginkel & Redig, *Journal of Statistical Physics* (2020).

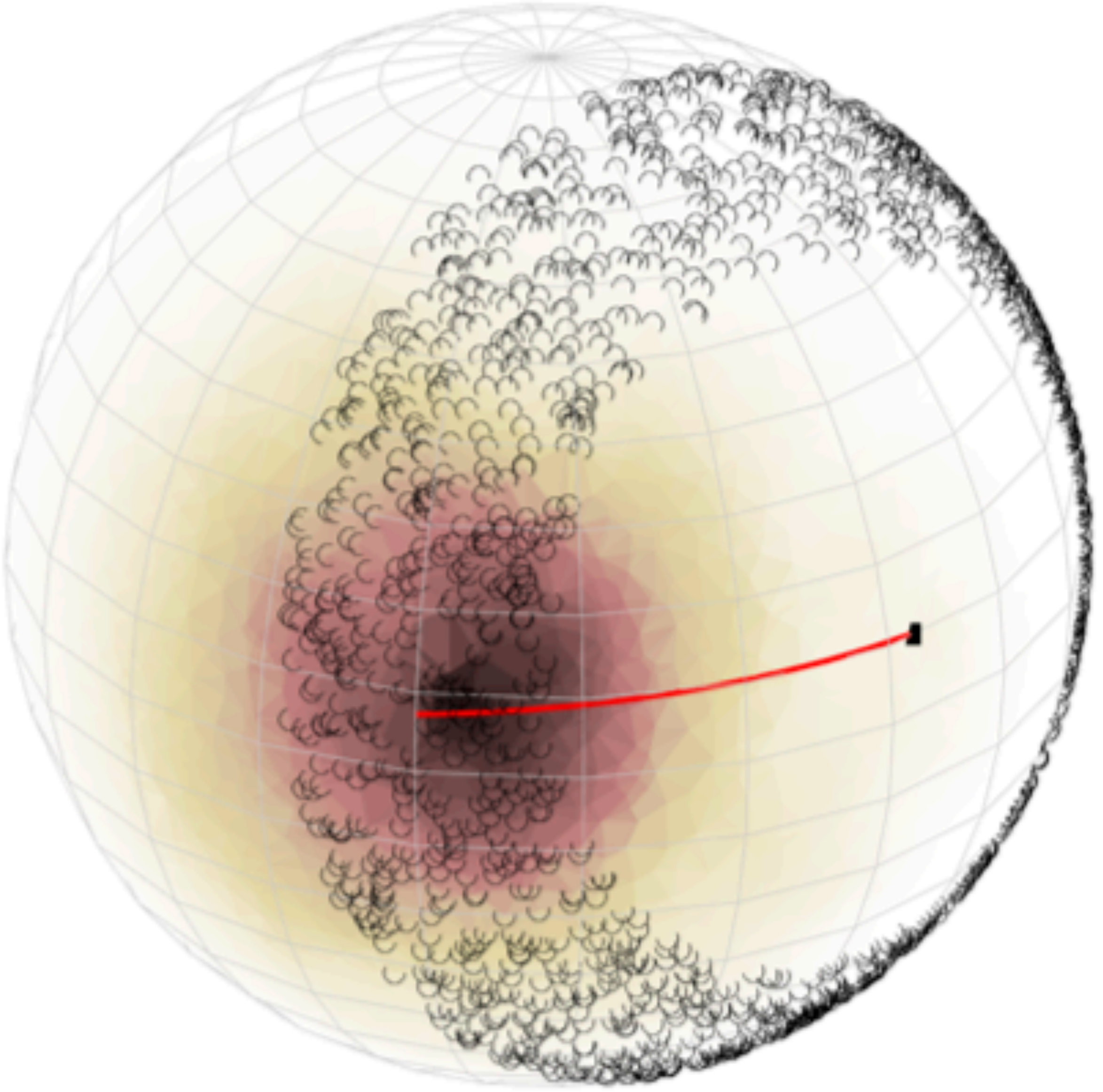


Crane et al., Comm. ACM (2017)

with high prob.,

$h \gg \sqrt{\epsilon}$



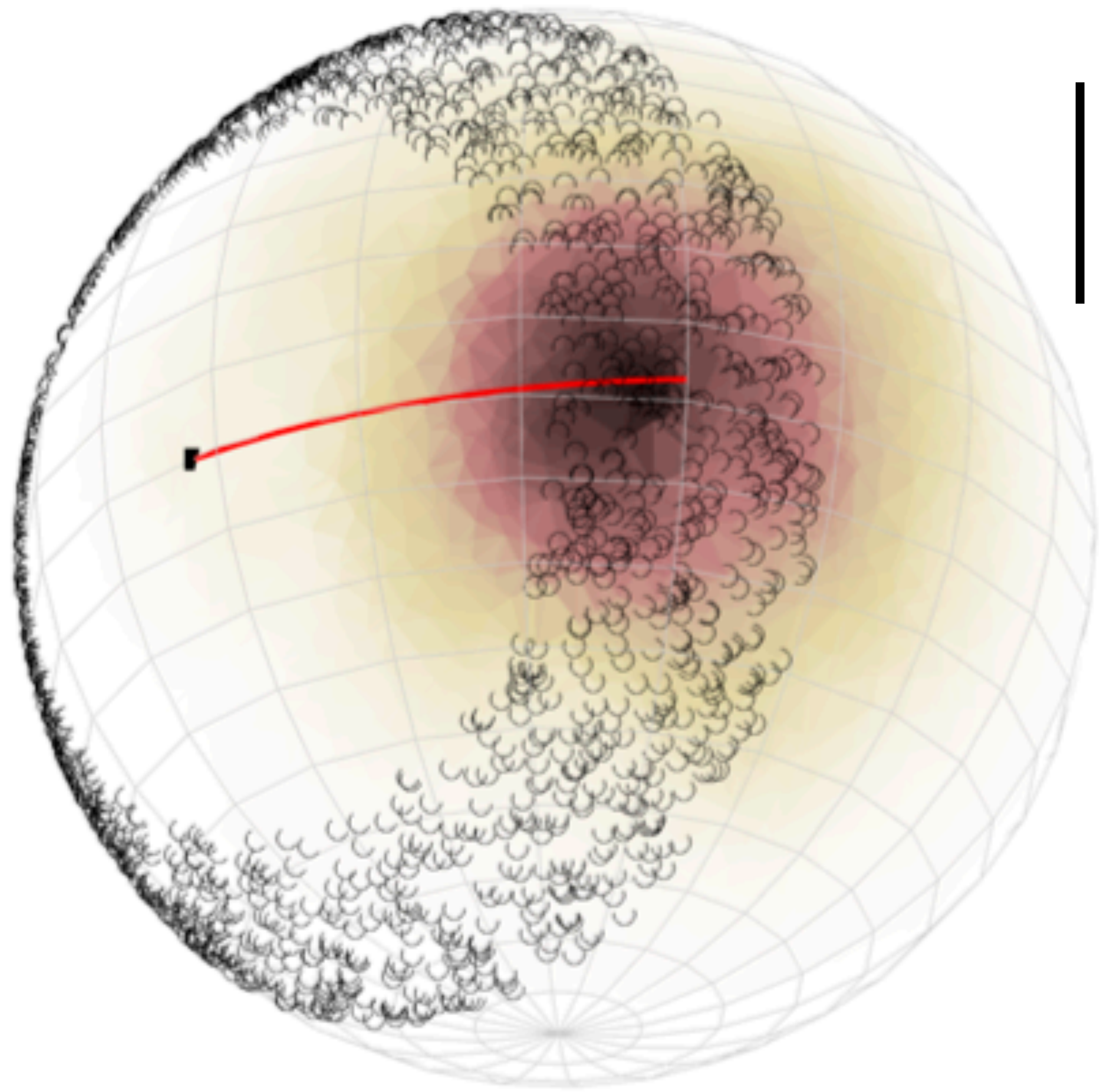


Theorem. If $h \gg \sqrt{\epsilon}$, then $\left| e^{it\sqrt{\Delta_{\epsilon,N}}}\,|\psi_h^{(x_0,\xi_0)}\rangle \right|^2 = \left| e^{\frac{i}{h}t\sqrt{\hat{H}_h}}\,|\psi_h^{(x_0,\xi_0)}\rangle \right|^2 + O(h)$

with probability at least $1 - e^{-\Omega(Nh^{\beta})}$ ($\beta > 0$ a constant). Kumar, arXiv:2112.10748 (2022)

From random walks to quantum dynamics of observables on data

Heisenberg equation $i\partial_t \operatorname{diag}(a|_{X_N}) = \left[\sqrt{\Delta}_{\epsilon,N}, \operatorname{diag}(a|_{X_N})\right]$



$\left|\psi_h(t)\right\rangle$

with high prob.,

$h \gg \sqrt{\epsilon}$

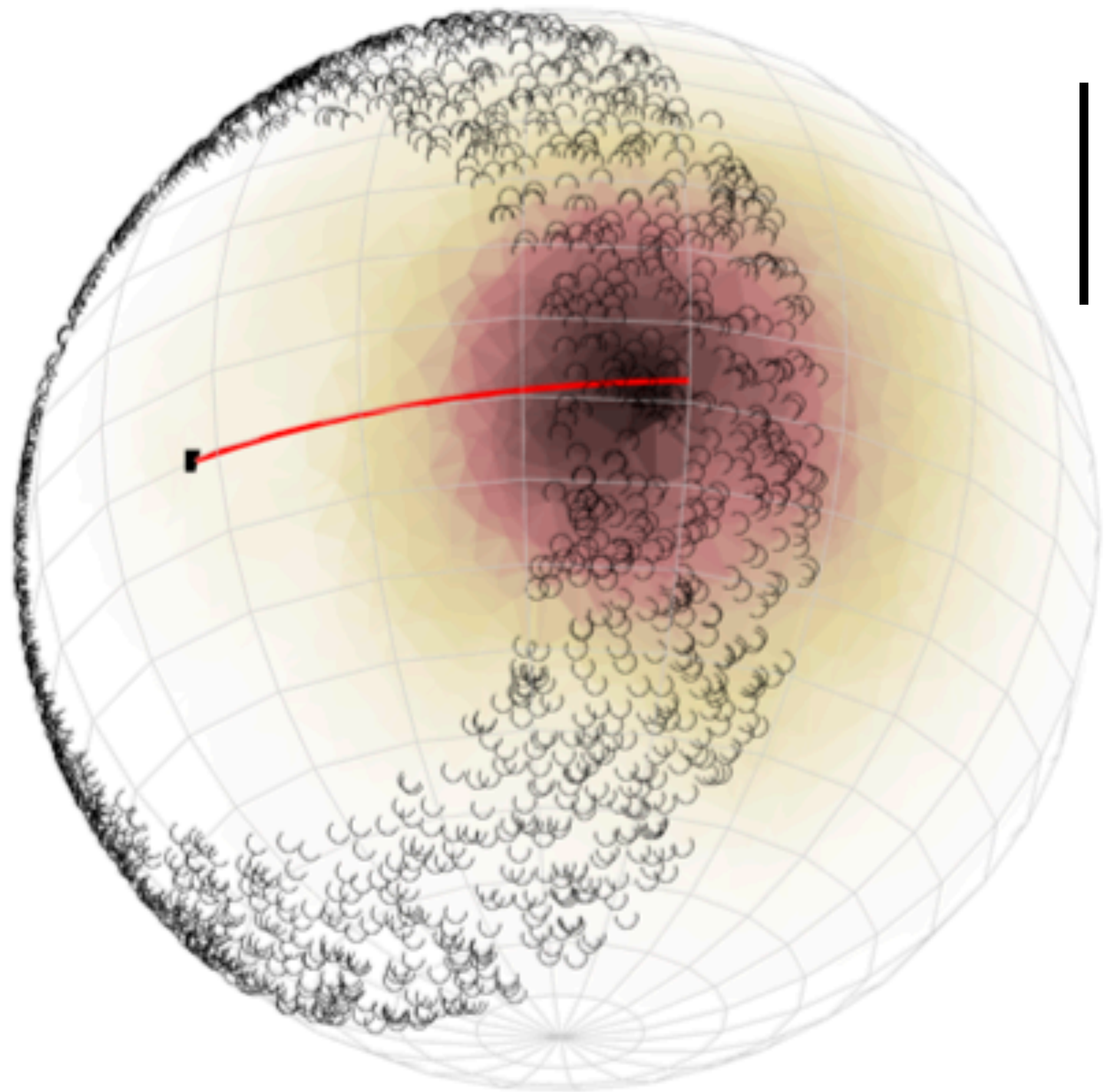
$$a \in C^\infty(\mathcal{M})$$

$$\left\langle \psi_h(t) \middle| \operatorname{diag}(a|_{X_N}) \middle| \psi_h(t) \right\rangle = a(x(t)) + O(h)$$

Kumar, arXiv:2112.10748 (2022)

From random walks to quantum dynamics of observables on data

Heisenberg equation $i\partial_t \operatorname{diag}(a|_{X_N}) = \left[\sqrt{\Delta}_{\epsilon,N}, \operatorname{diag}(a|_{X_N})\right]$



$\left|\psi_h(t)\right\rangle$

with high prob.,

$h \gg \sqrt{\epsilon}$

$a \in C^\infty(\mathcal{M})$

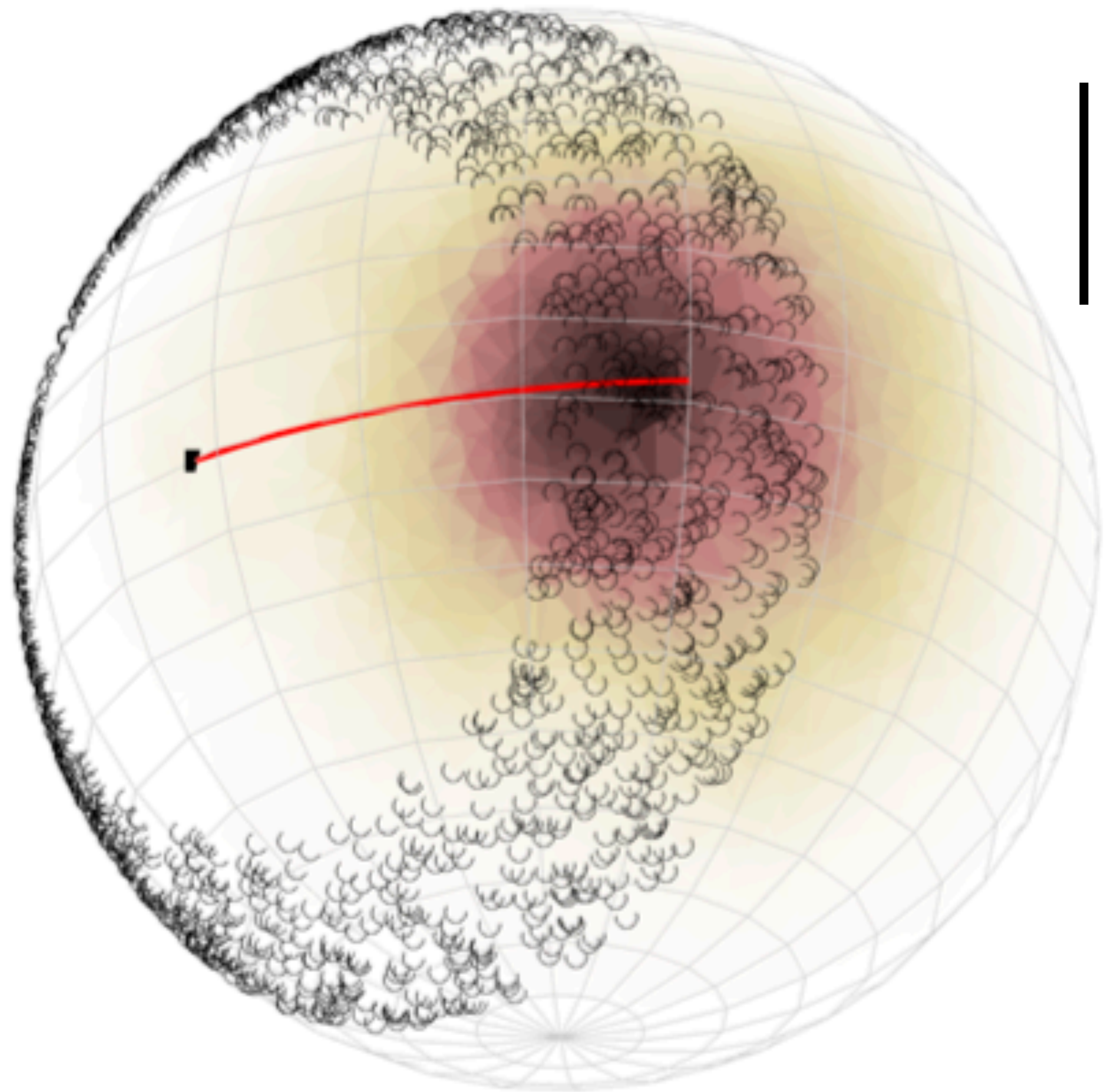
$\left\langle \psi_h(t) \middle| \operatorname{diag}(a|_{X_N}) \middle| \psi_h(t) \right\rangle = a(x(t)) + O(h)$

Theorem. If $a \in C^\infty(\mathcal{M})$ and $h \gg \sqrt{\epsilon}$, then $\left\langle \psi_h(t) \middle| \operatorname{diag}(a|_{X_N}) \middle| \psi_h(t) \right\rangle = a(x(t)) + O(h)$ with probability at least $1 - e^{-\Omega(Nh^\beta)}$ ($\beta > 0$ a constant).

Kumar, arXiv:2112.10748 (2022)

From random walks to quantum dynamics of observables on data

Heisenberg equation $i\partial_t \operatorname{diag}(a|_{X_N}) = \left[\sqrt{\Delta}_{\epsilon,N}, \operatorname{diag}(a|_{X_N})\right]$



$\left|\psi_h(t)\right\rangle$

with high prob.,

$h \gg \sqrt{\epsilon}$

$a \in C^\infty(\mathscr{M})$

$\left\langle \psi_h(t) \middle| \operatorname{diag}(a|_{X_N}) \middle| \psi_h(t) \right\rangle = a(x(t)) + O(h)$

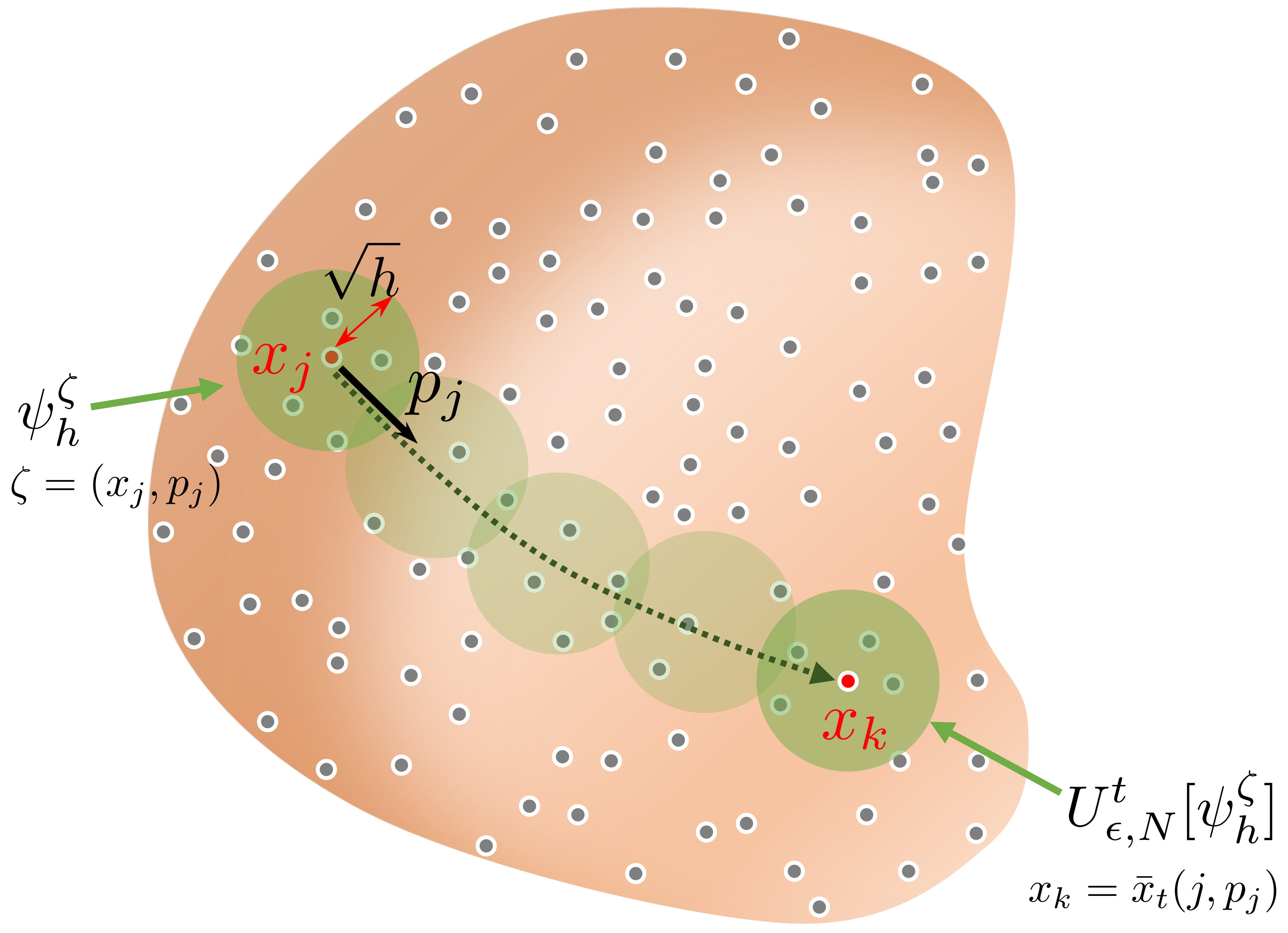
Microlocal analysis ∩ geometry ∩ markov processes ∩ probability/statistics ∩ quantum dynamics

Theorem. If $a \in C^\infty(\mathscr{M})$ and $h \gg \sqrt{\epsilon}$, then $\left\langle \psi_h(t) \middle| \operatorname{diag}(a|_{X_N}) \middle| \psi_h(t) \right\rangle = a(x(t)) + O(h)$ with probability at least $1 - e^{-\Omega(Nh^\beta)}$ ($\beta > 0$ a constant).

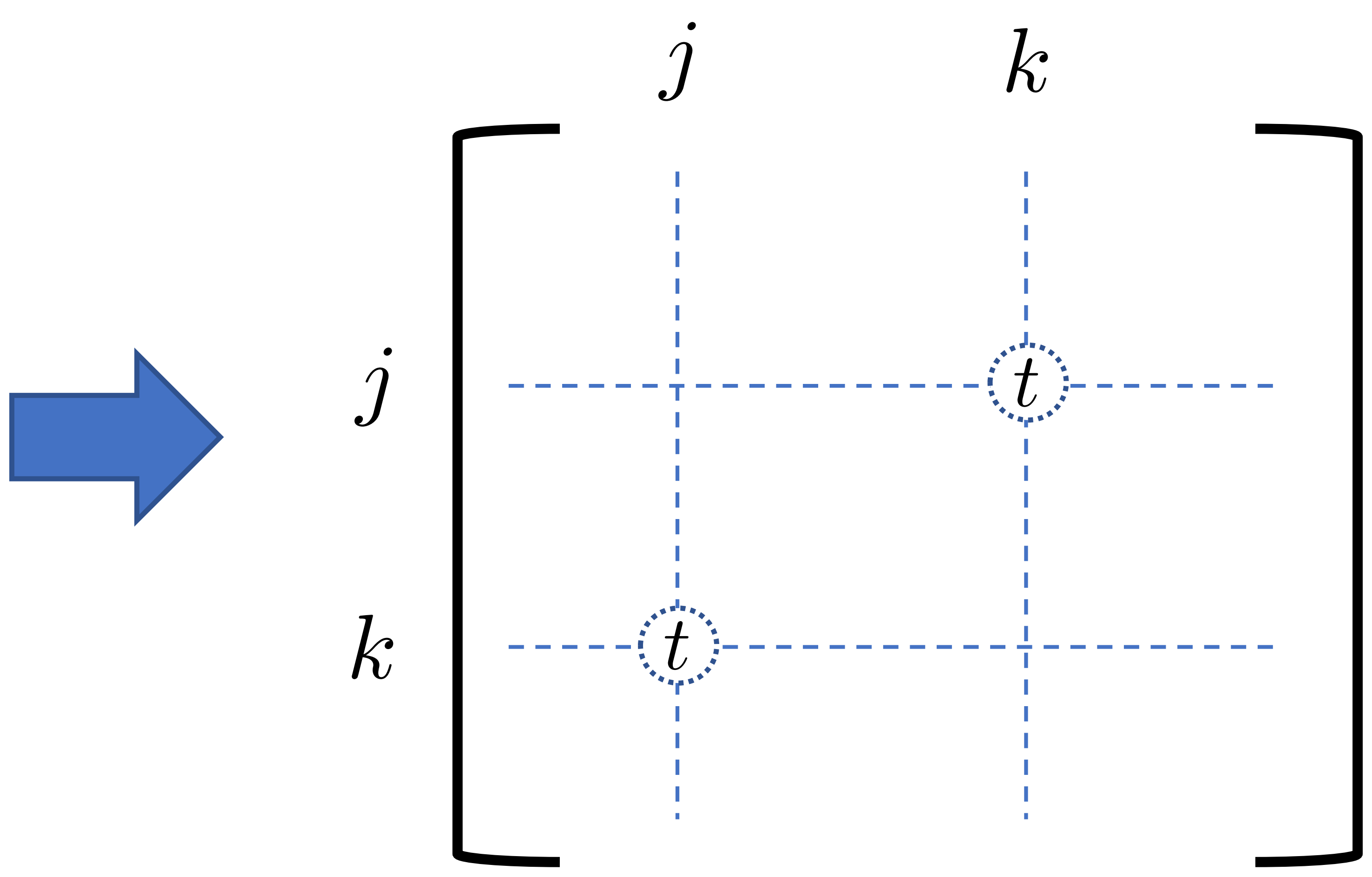
Kumar, arXiv:2112.10748 (2022)

Applications

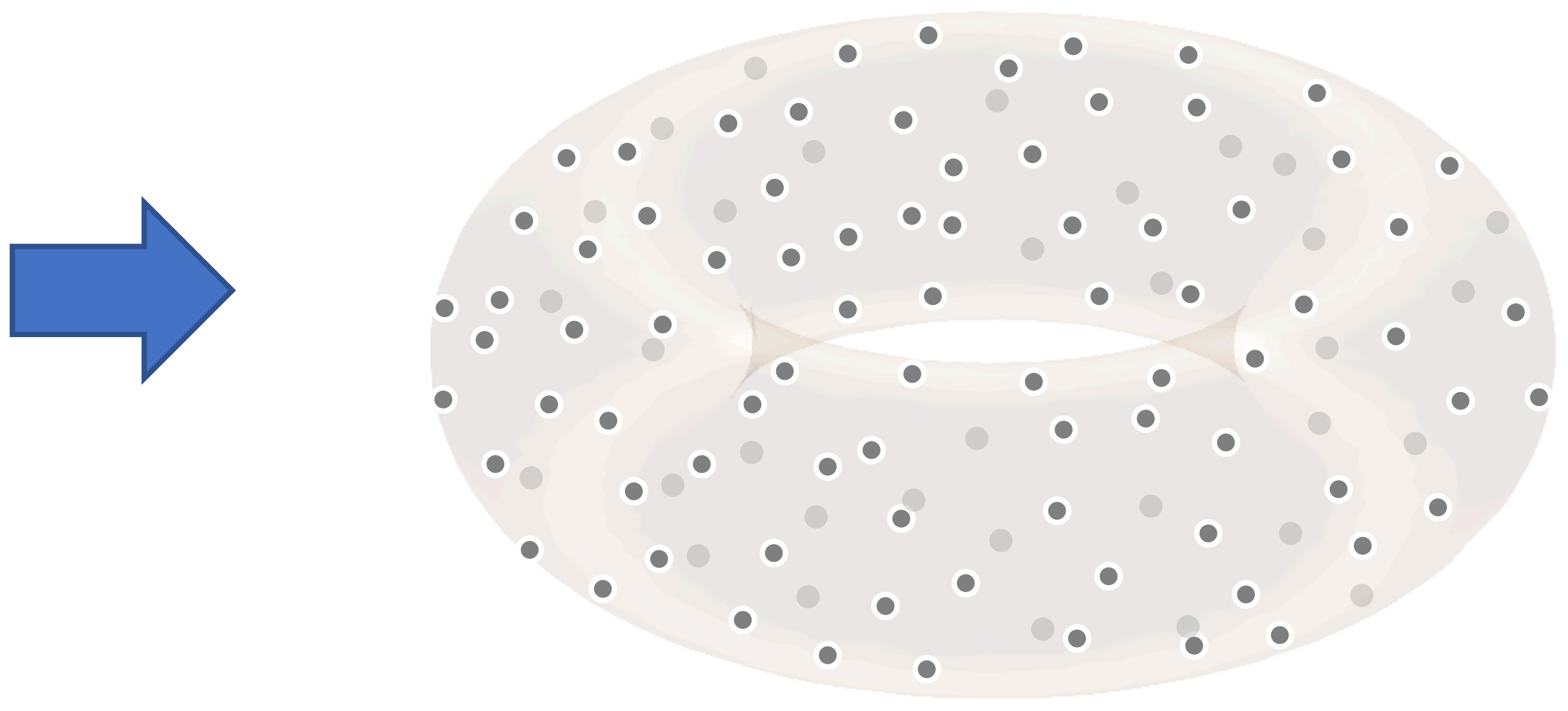
Quantum dynamical *reach-time* embedding of datasets



Cross-section of dataset of size N in $\mathbb{R}^n$



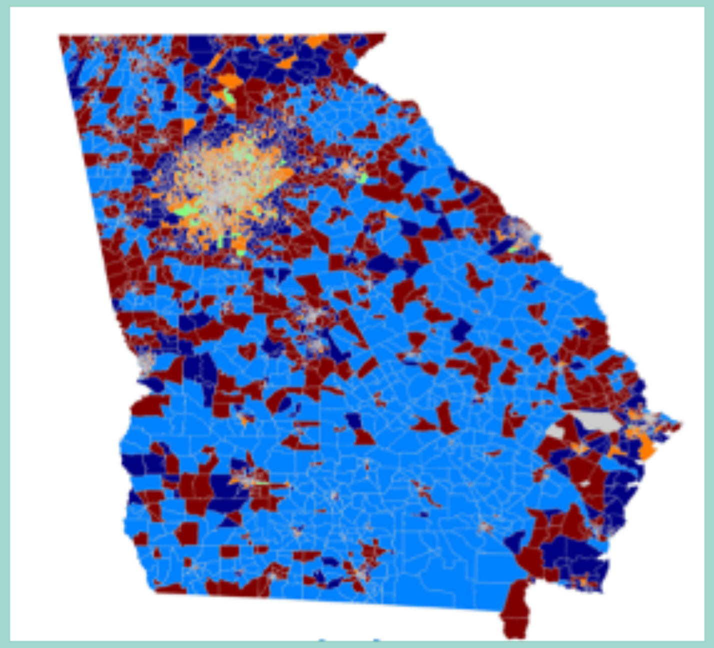
$\mathcal{G}$ $(N \times N \text{ matrix})$



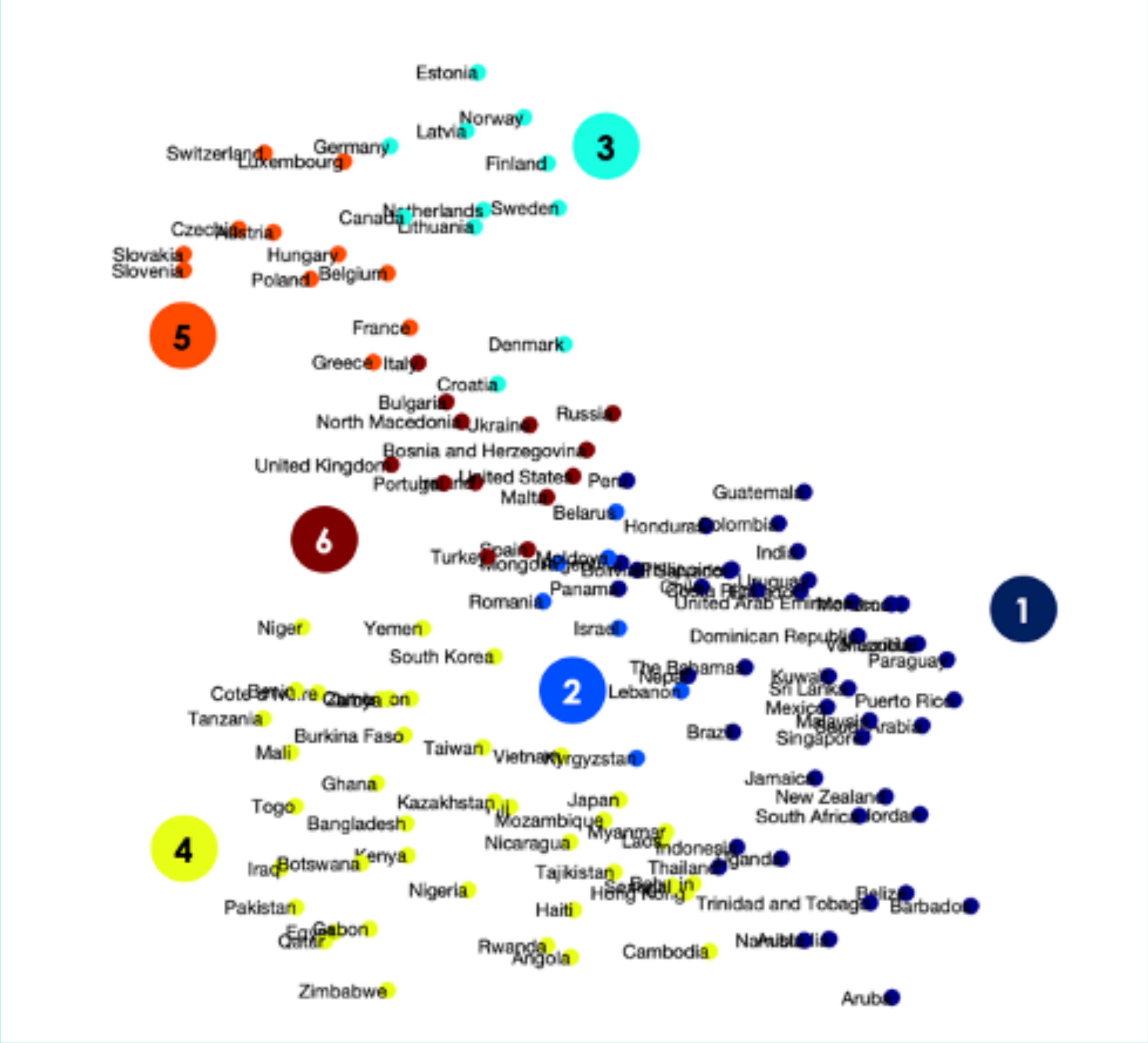
Graph embedding of dataset in $\mathbb{R}^3$ based on weighted adjacency matrix $\mathcal{G}$

Example applications

Clustering of geographic regions according to COVID-19 social distancing behavior

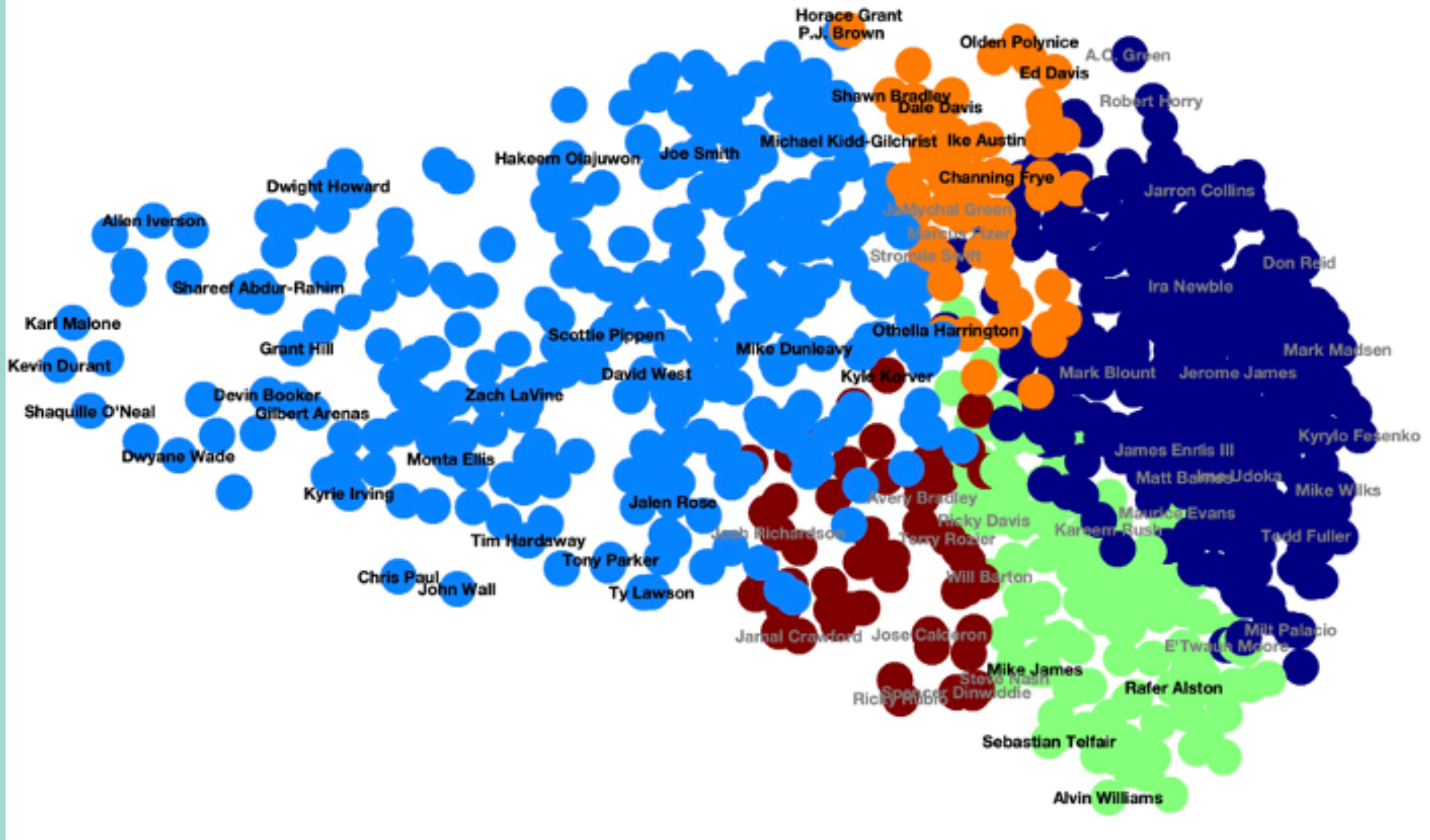


SafeGraph dataset for state of GA



Google dataset for worldwide mobility patterns

Clustering of NBA players based on performance



NOTES:

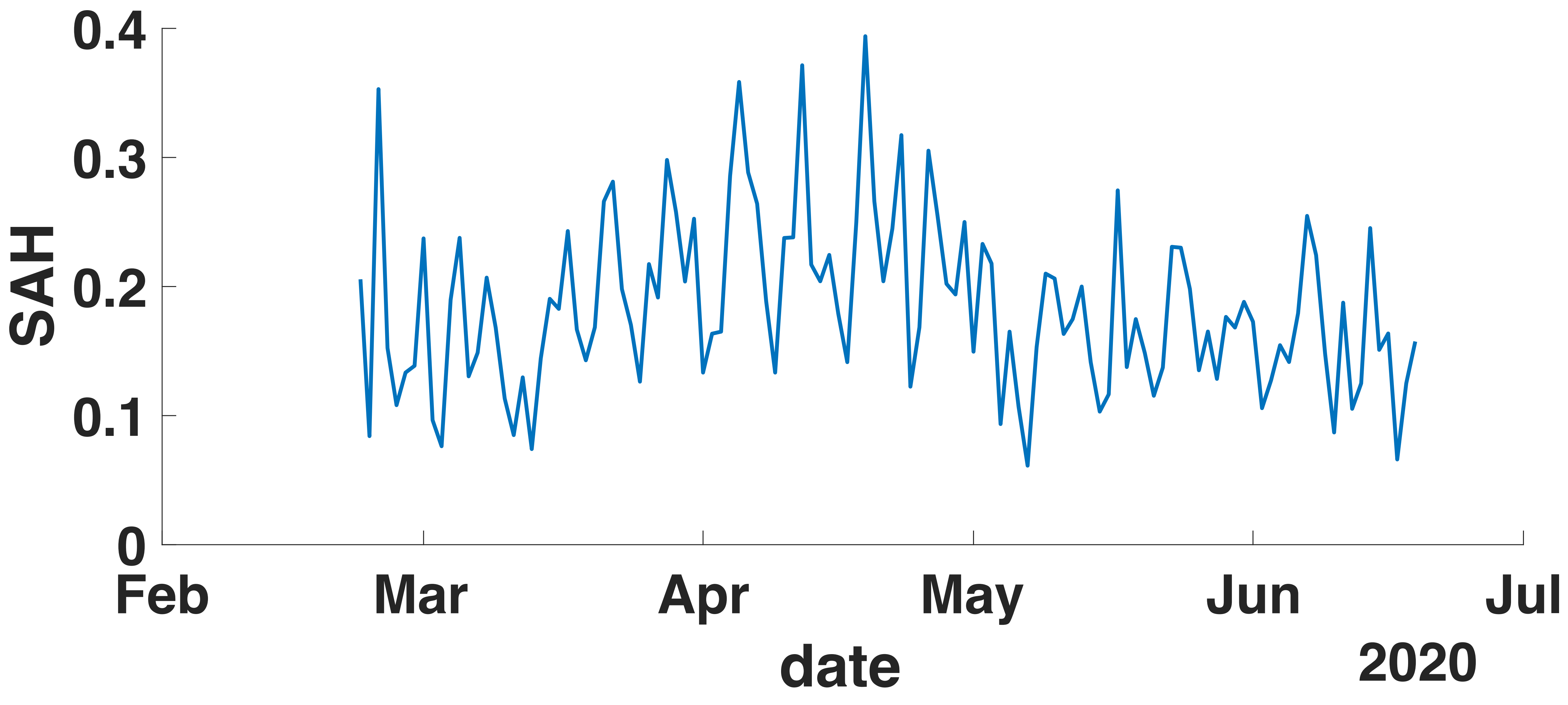
1. Manifold hypothesis not necessary for utility
2. Some of these datasets are small (e.g., $N = 126$)

See "Manifold learning via quantum dynamics." A. Kumar & M. Sarovar. arXiv:2112.11161 (2021)

Example: COVID-19 mobility data

Social Distancing Metric dataset from SafeGraph Inc.
https://docs.safegraph.com/docs/social-distancing-metrics

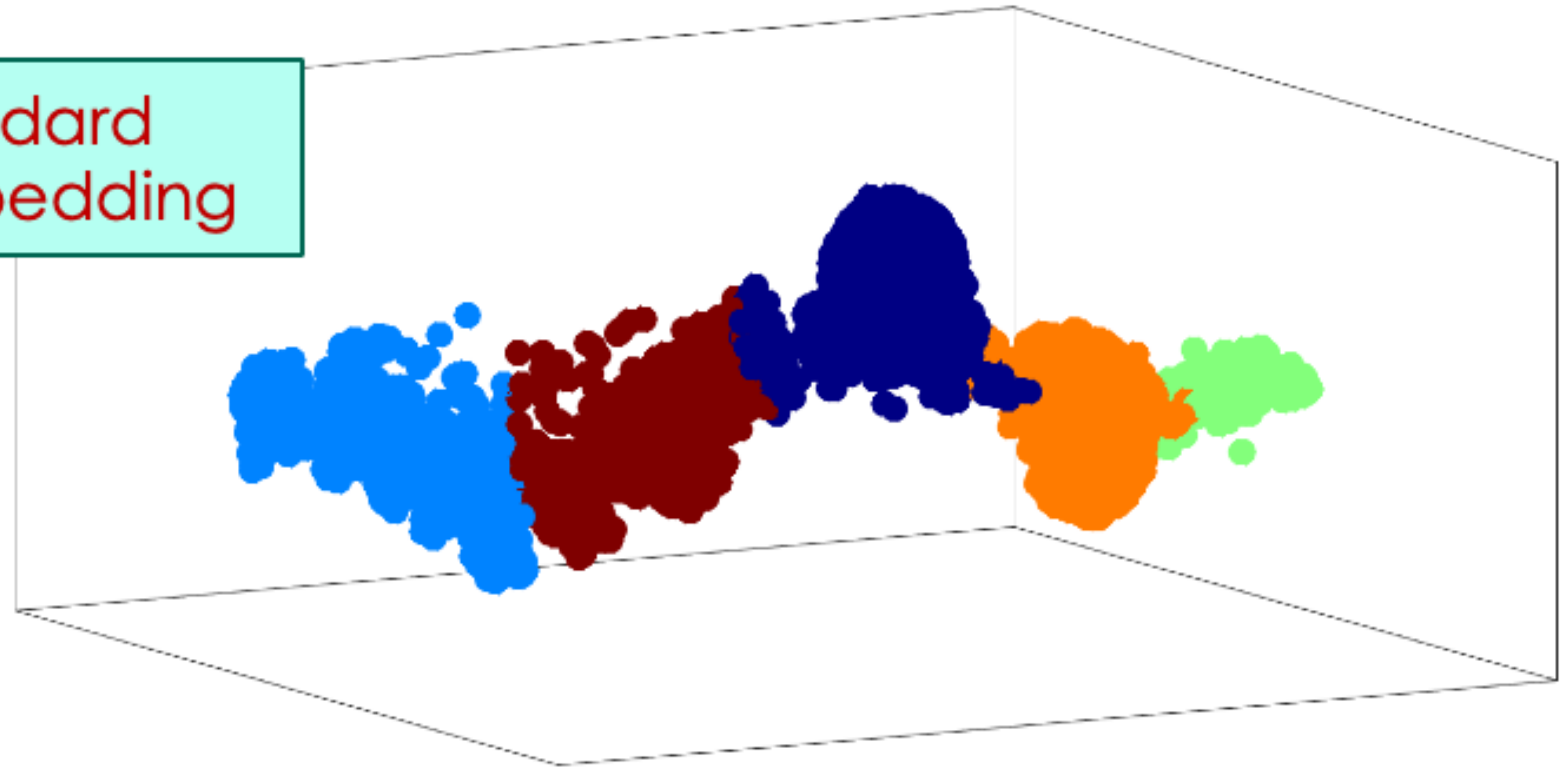
- Dataset collects user location information (from cellphone GPS data) over the course of the initial ~3 months of the COVID-19 pandemic (Feb 23, 2020 – June 19, 2020: 117 days).
- Data aggregated at the census block group (CBG) level.
- We compute a "stay-at-home" fraction for each CBG, which represents the fraction of devices that stayed at their home location on a day.
- We concentrate on the data for Georgia (GA), which has 5509 CBGs.
- Dataset: 5509 x 117



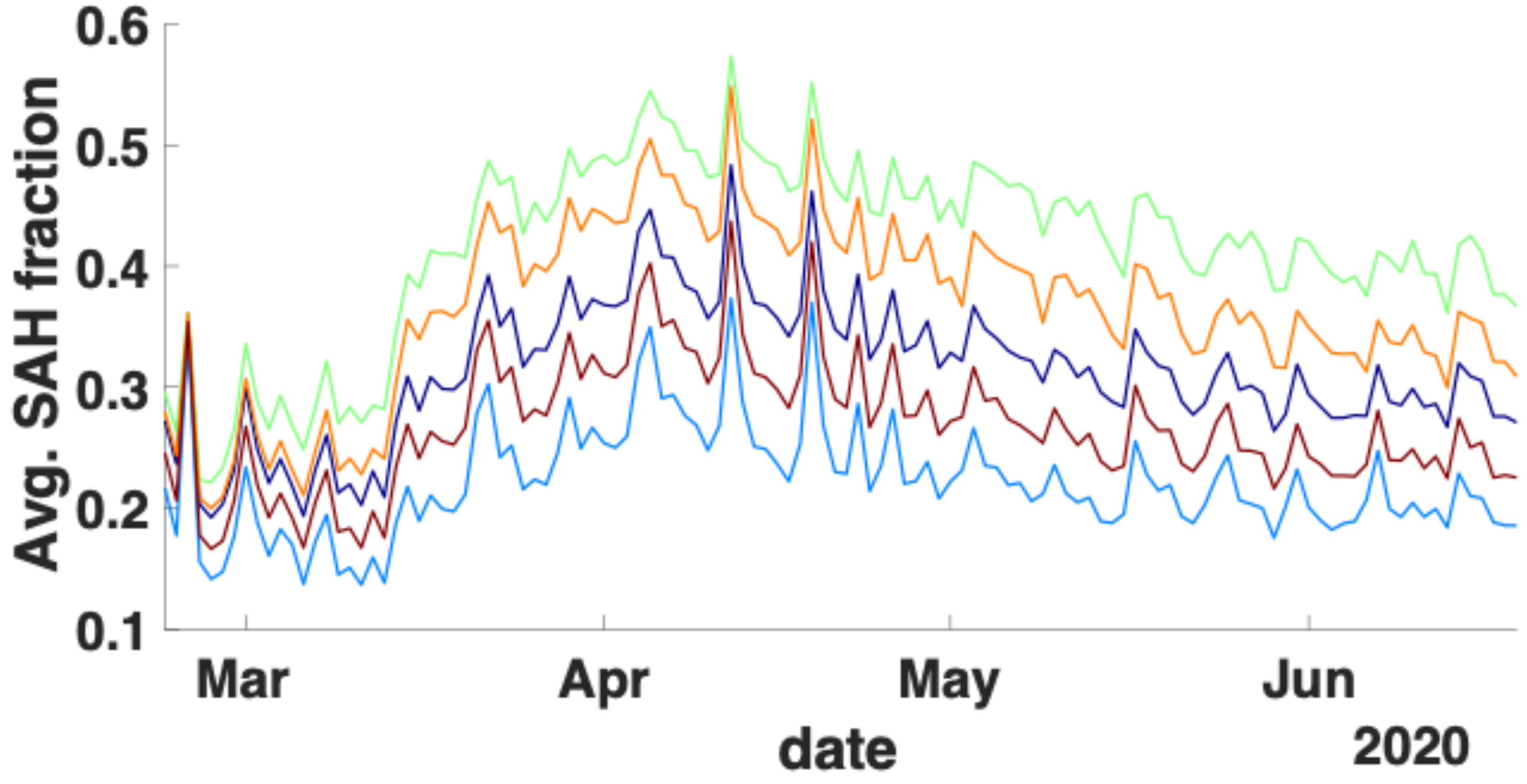
Example: COVID-19 mobility data

Social Distancing Metric dataset from SafeGraph Inc.
https://docs.safegraph.com/docs/social-distancing-metrics

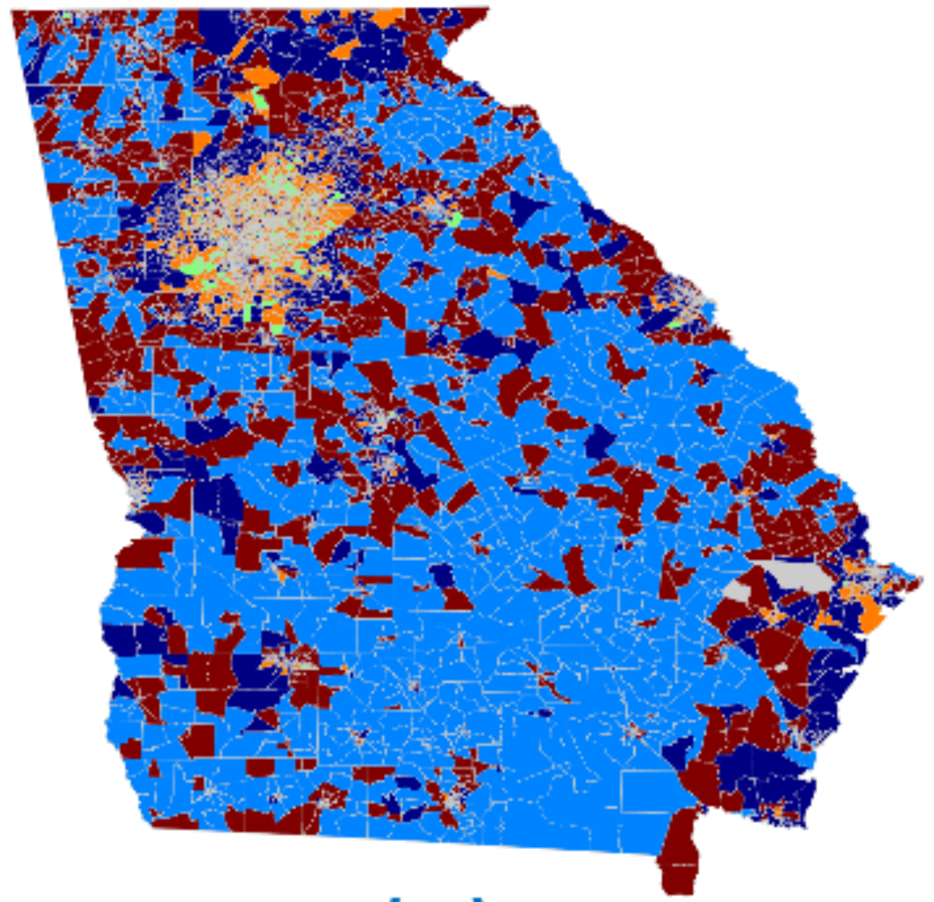
Standard embedding



Embedding in 3D and k-means using these 3D coordinates

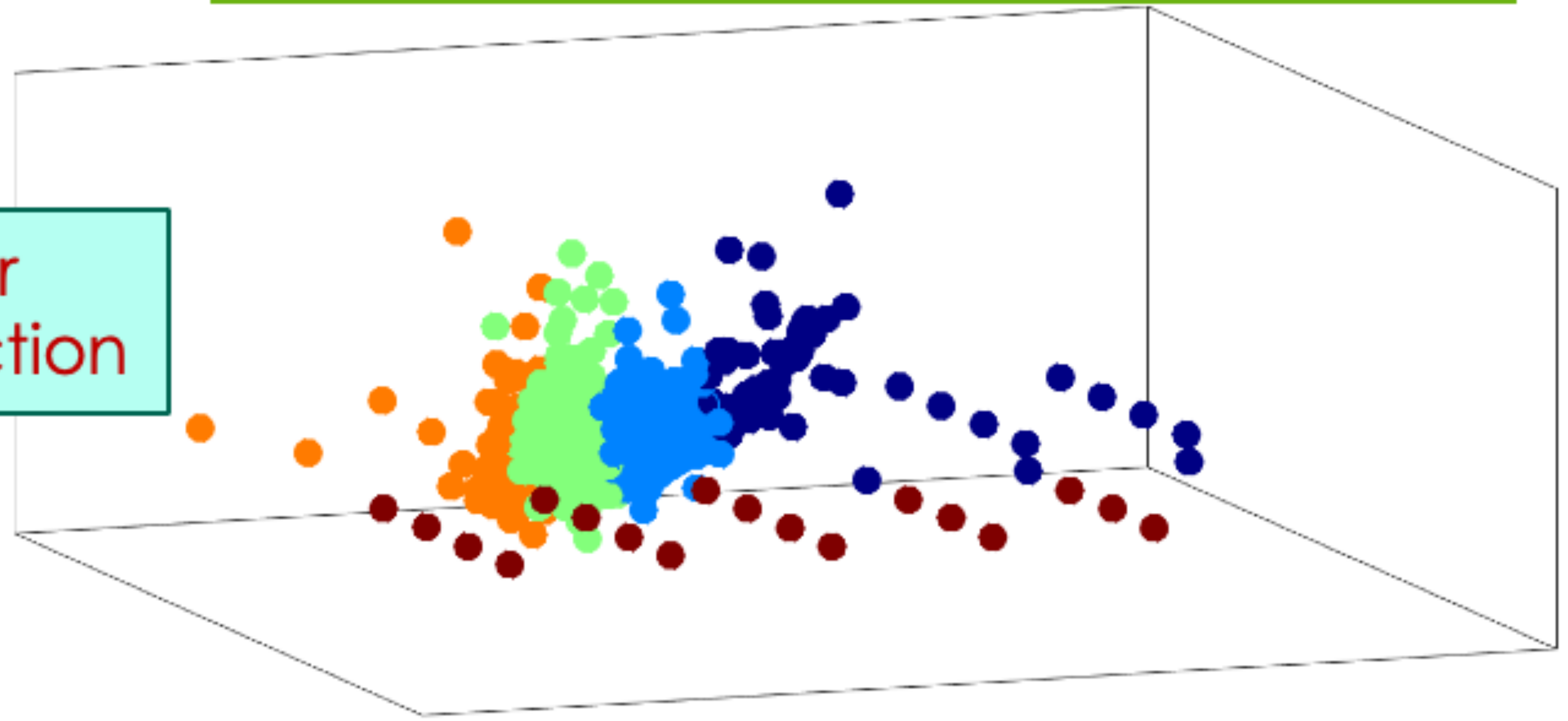


Average SAH fraction time series for each cluster

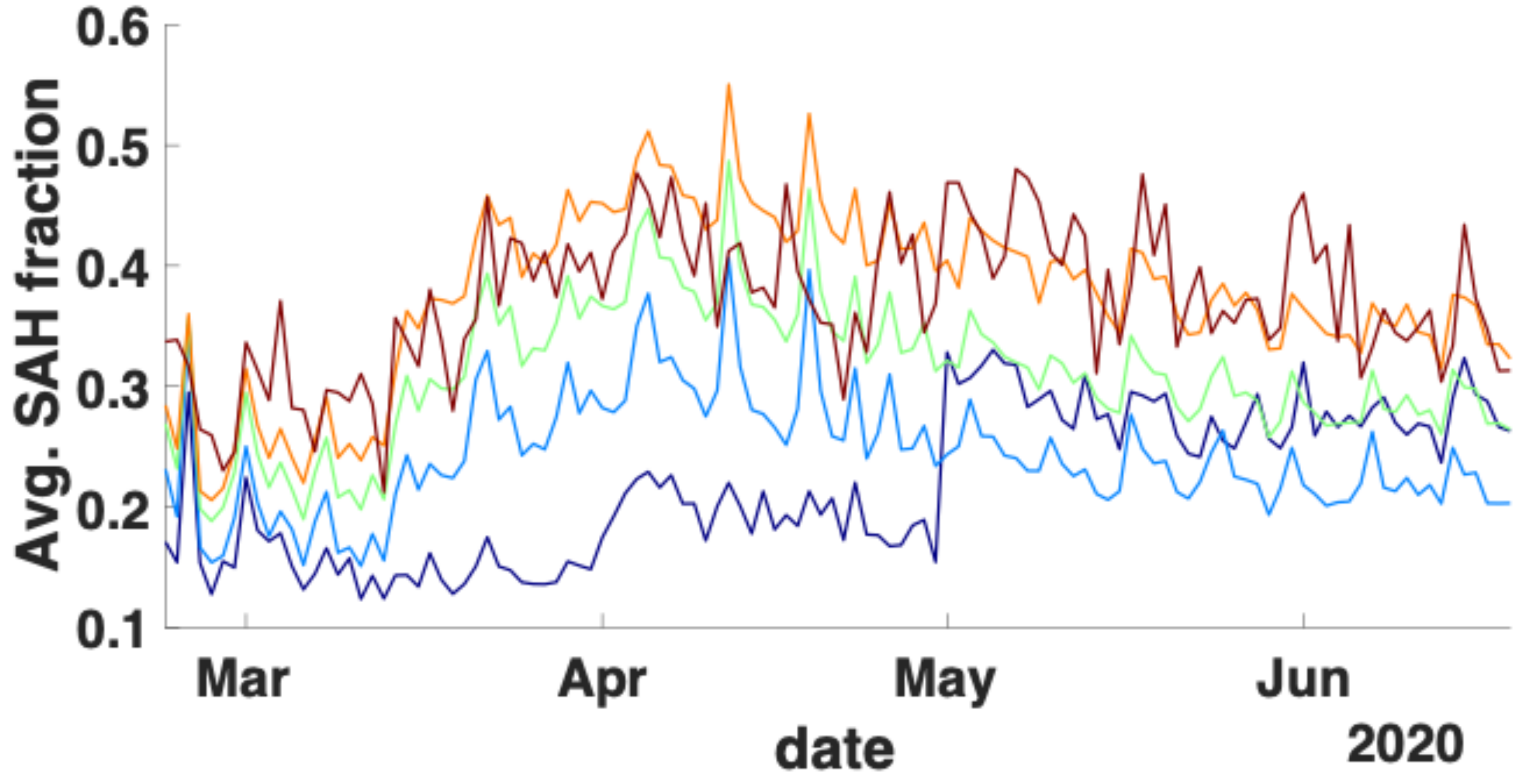


(c)

Outlier detection

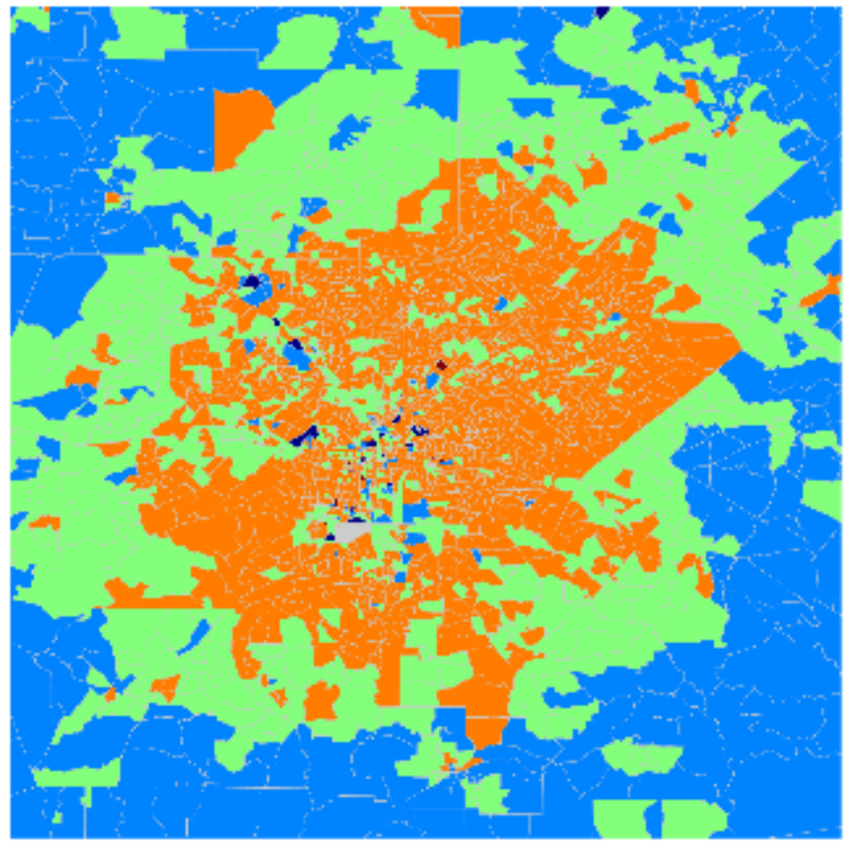


(d)



(e)

Outlier (1.2% of data)

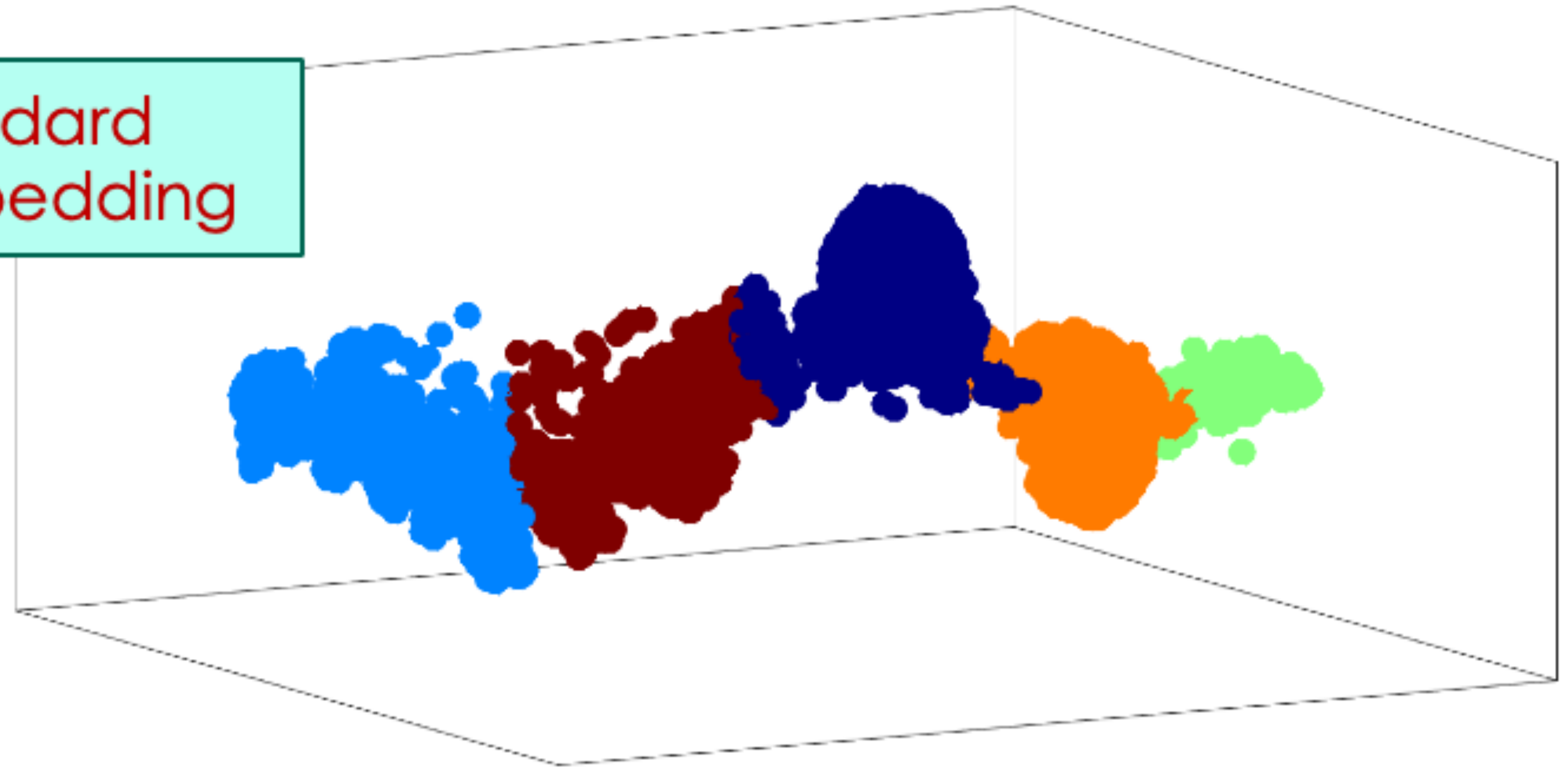


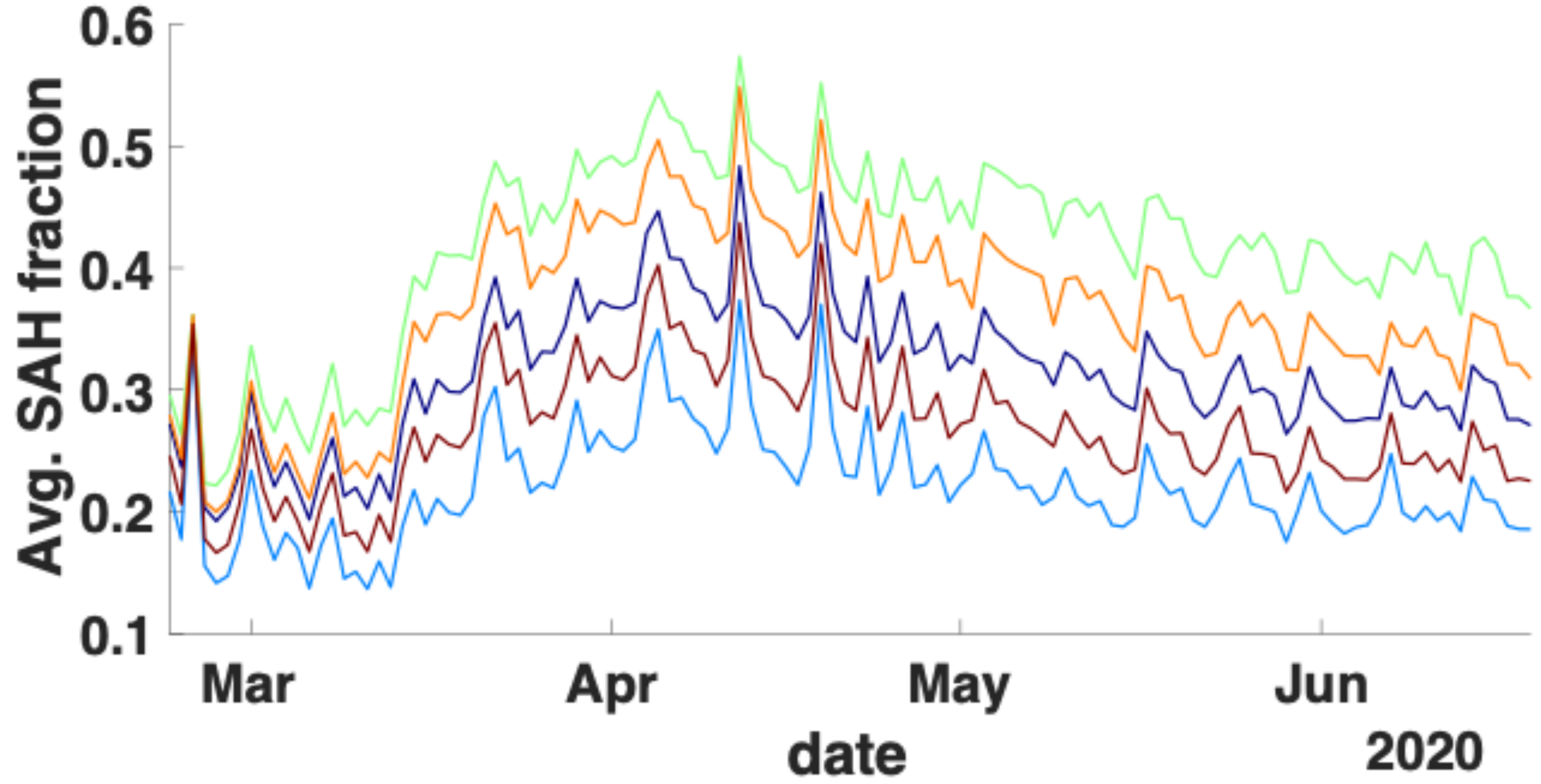
(f)

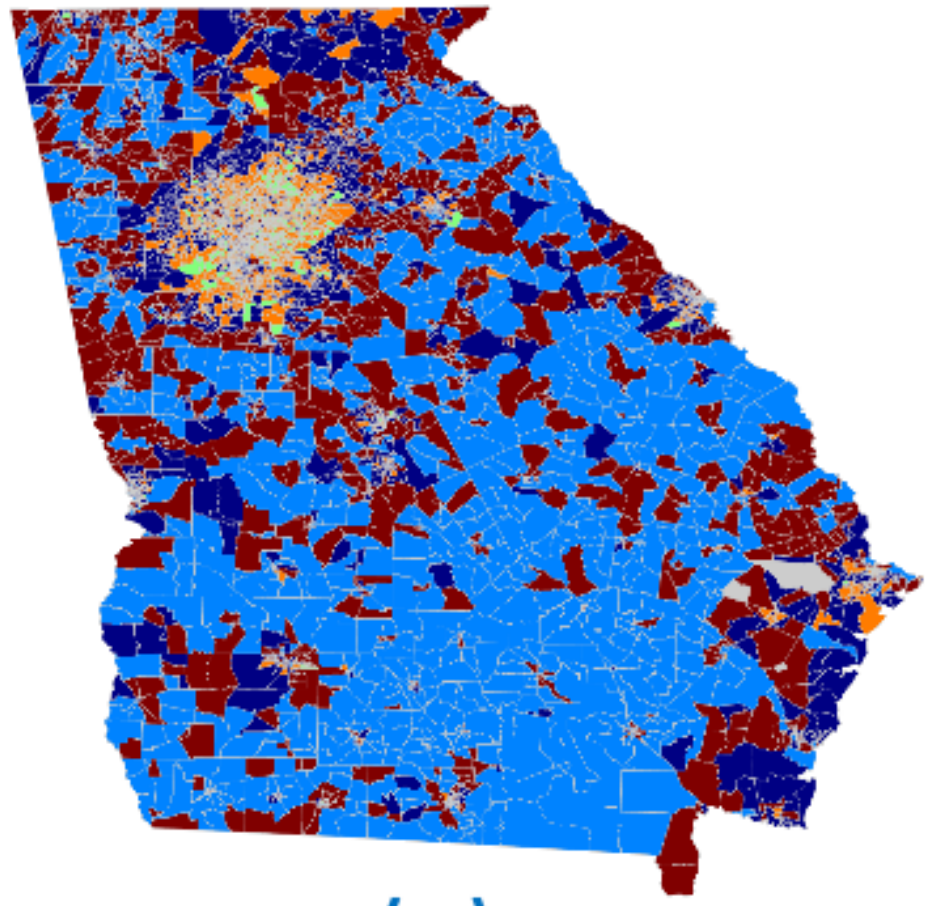
Example: COVID-19 mobility data

Social Distancing Metric dataset from SafeGraph Inc.
https://docs.safegraph.com/docs/social-distancing-metrics

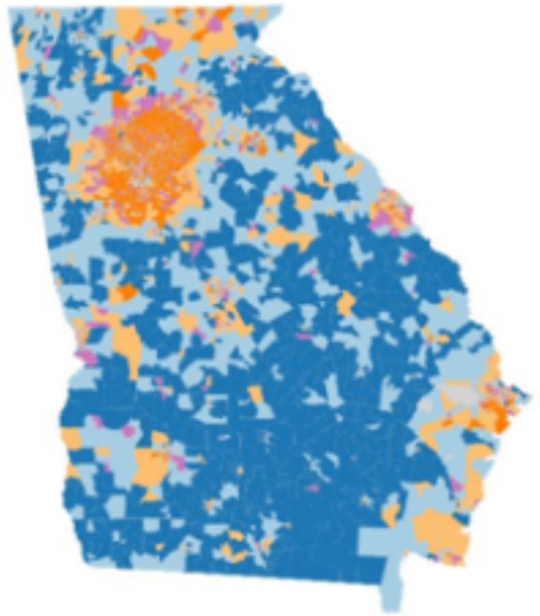
Standard embedding







c.f. Levin et al., "Cell Phone Mobility Data and Manifold Learning." https://doi.org/10.1101/2020.10.31.20223776



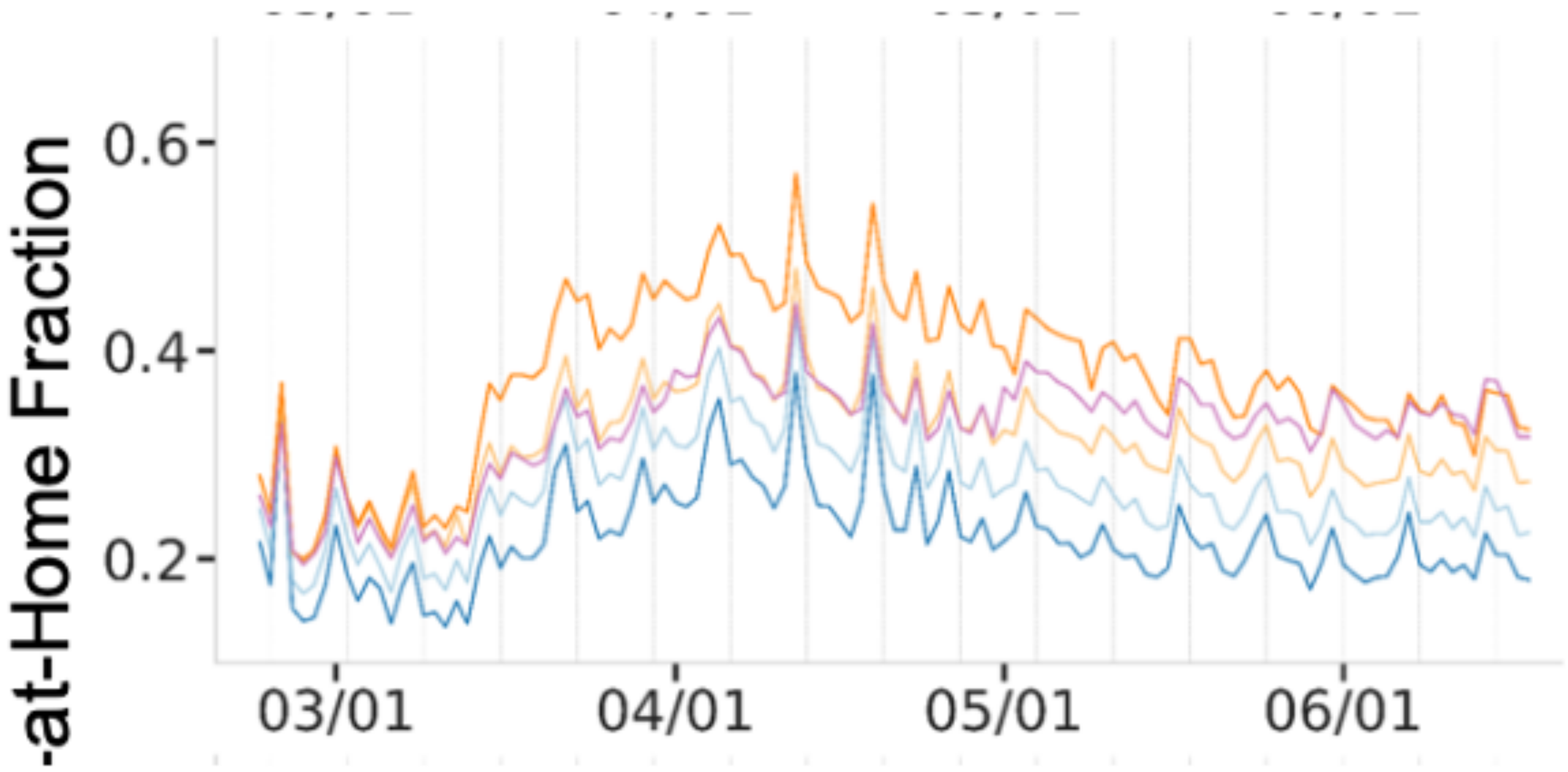
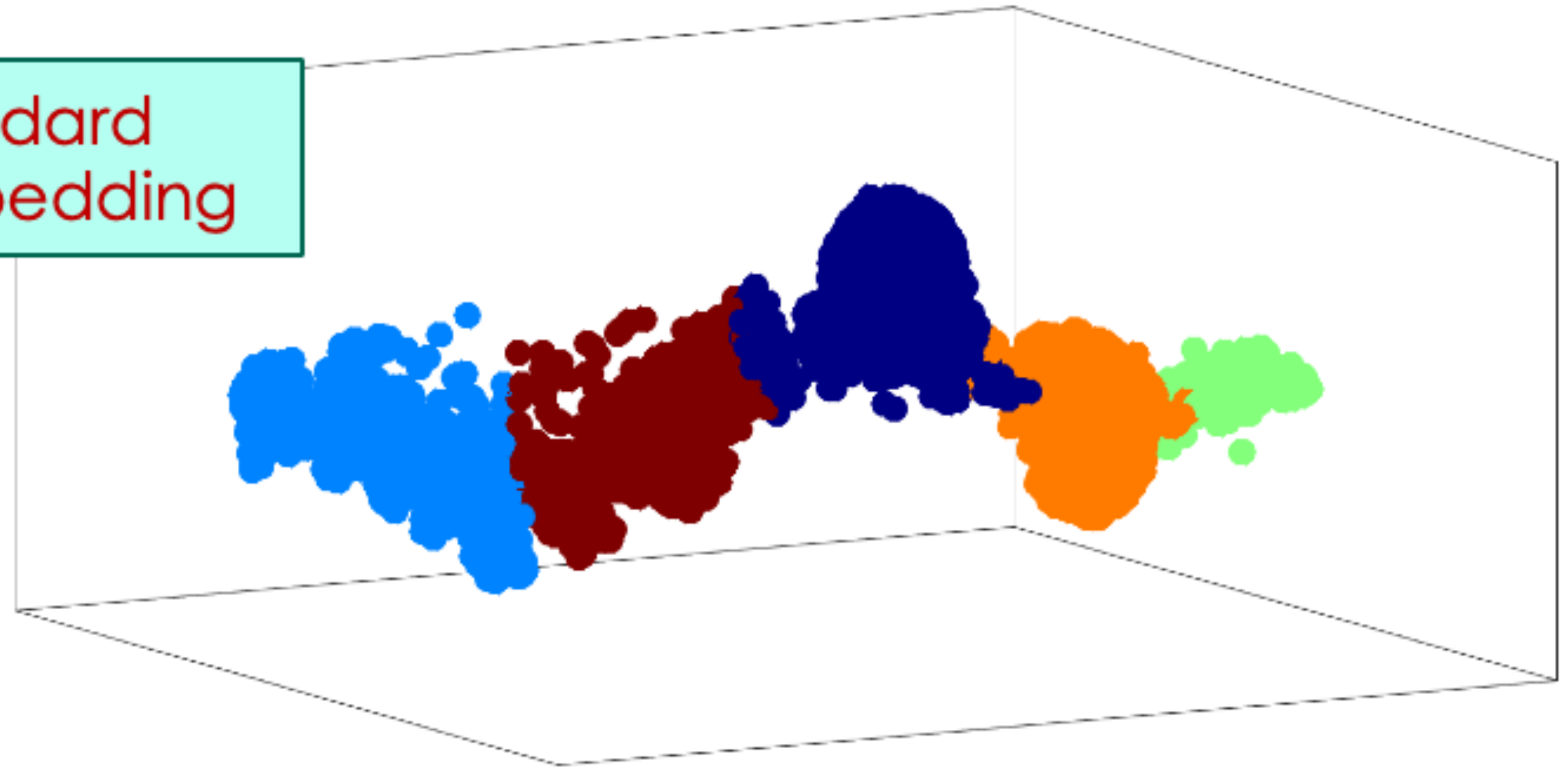


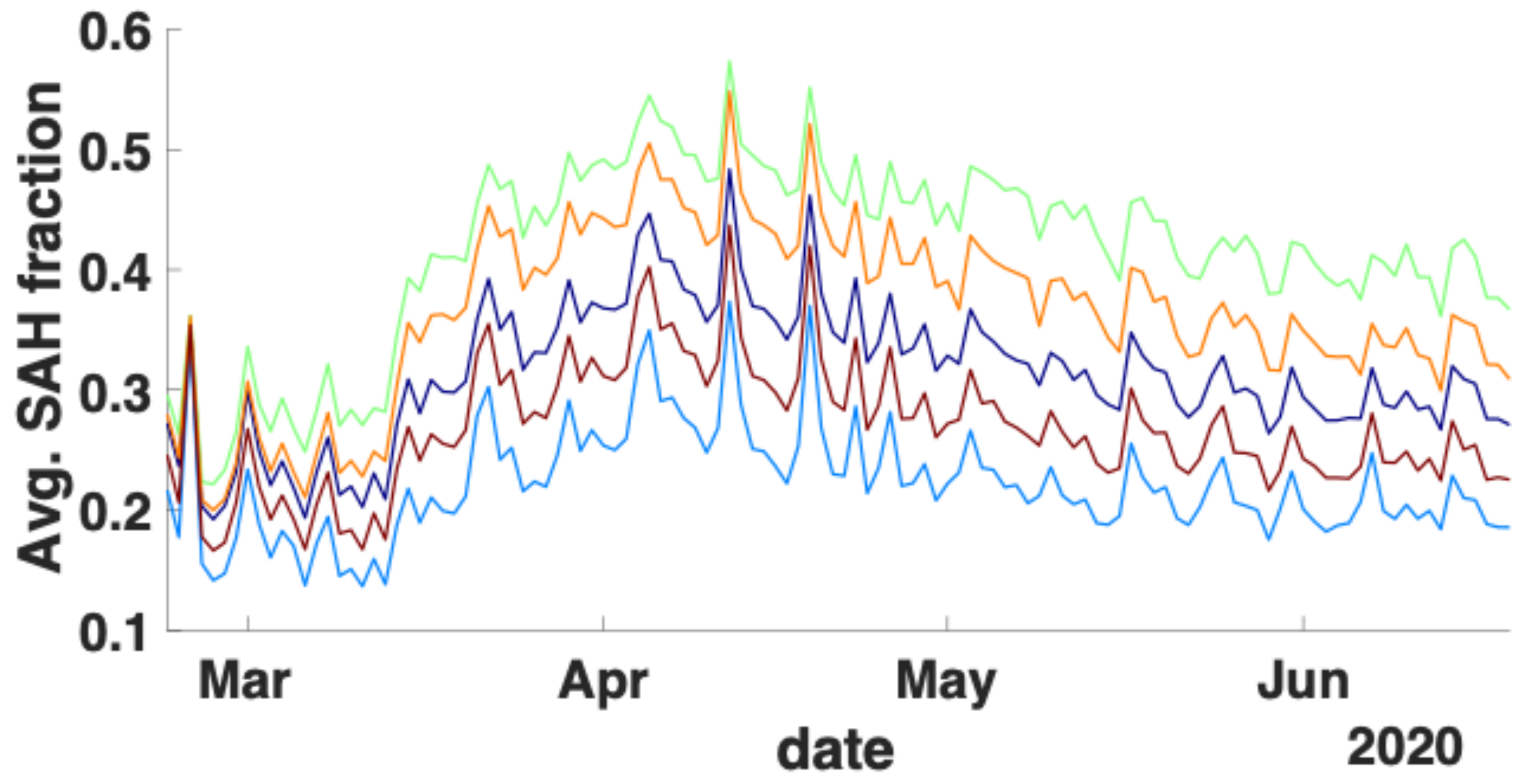
Figure 2: Laplacian eigenmaps, clustering done after embedding in **14 dimensions**

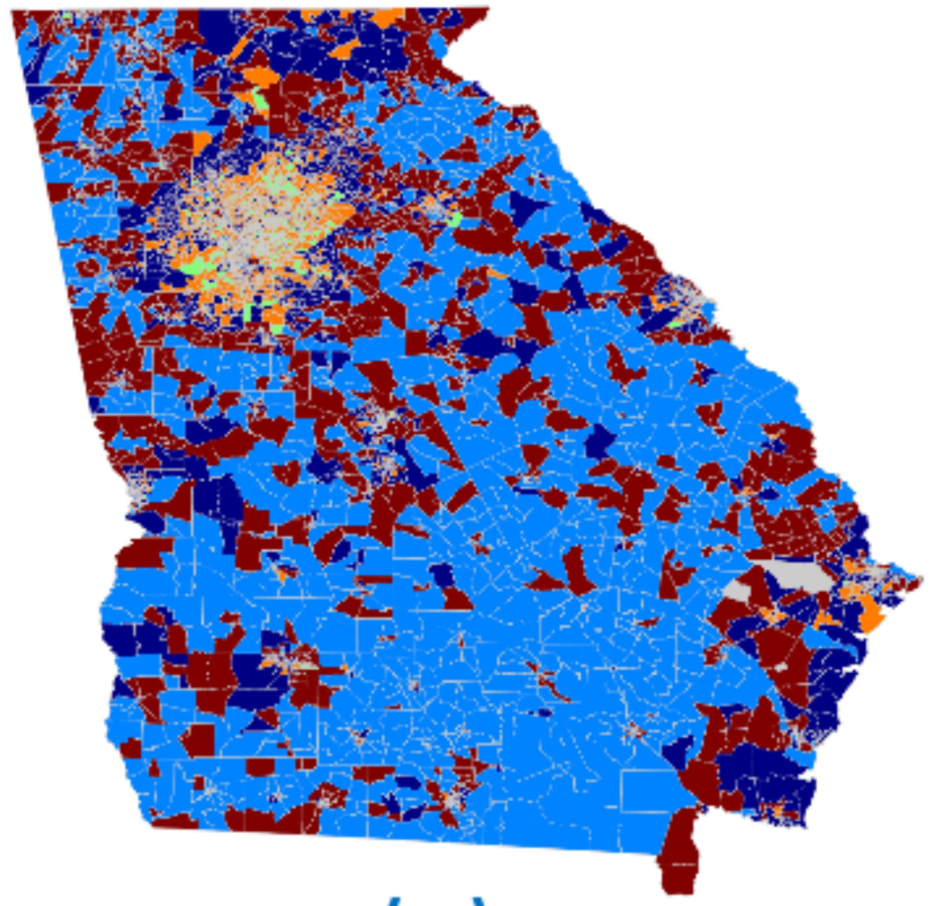
Example: COVID-19 mobility data

Social Distancing Metric dataset from SafeGraph Inc.
https://docs.safegraph.com/docs/social-distancing-metrics

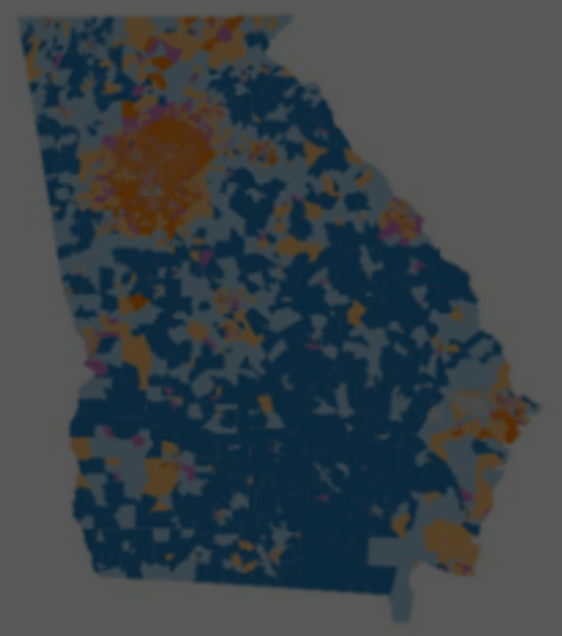
Standard embedding







c.f. Levin et al., "Cell Phone Mobility Data and Manifold Learning." https://doi.org/10.1101/2020.10.31.20223776



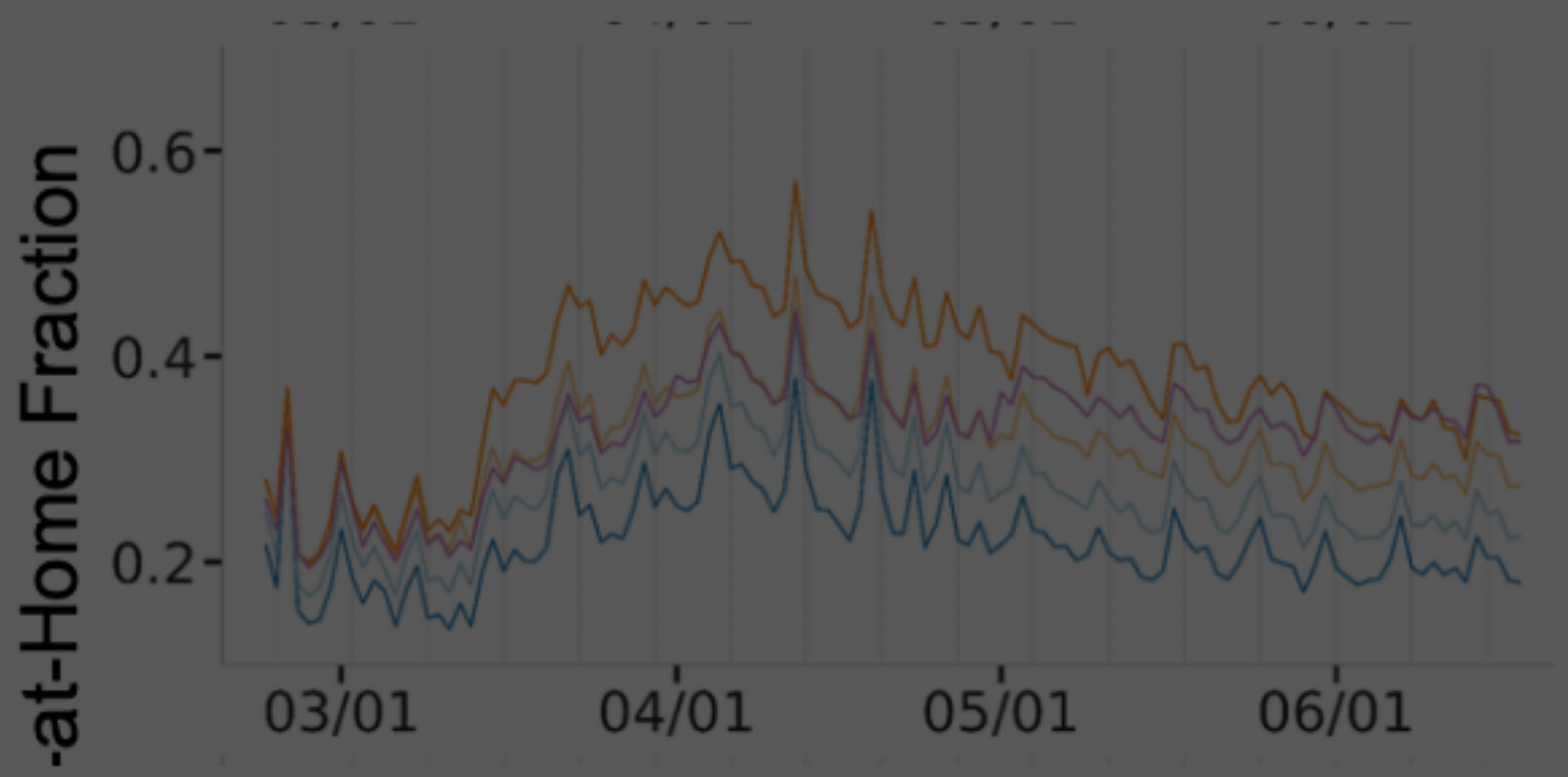


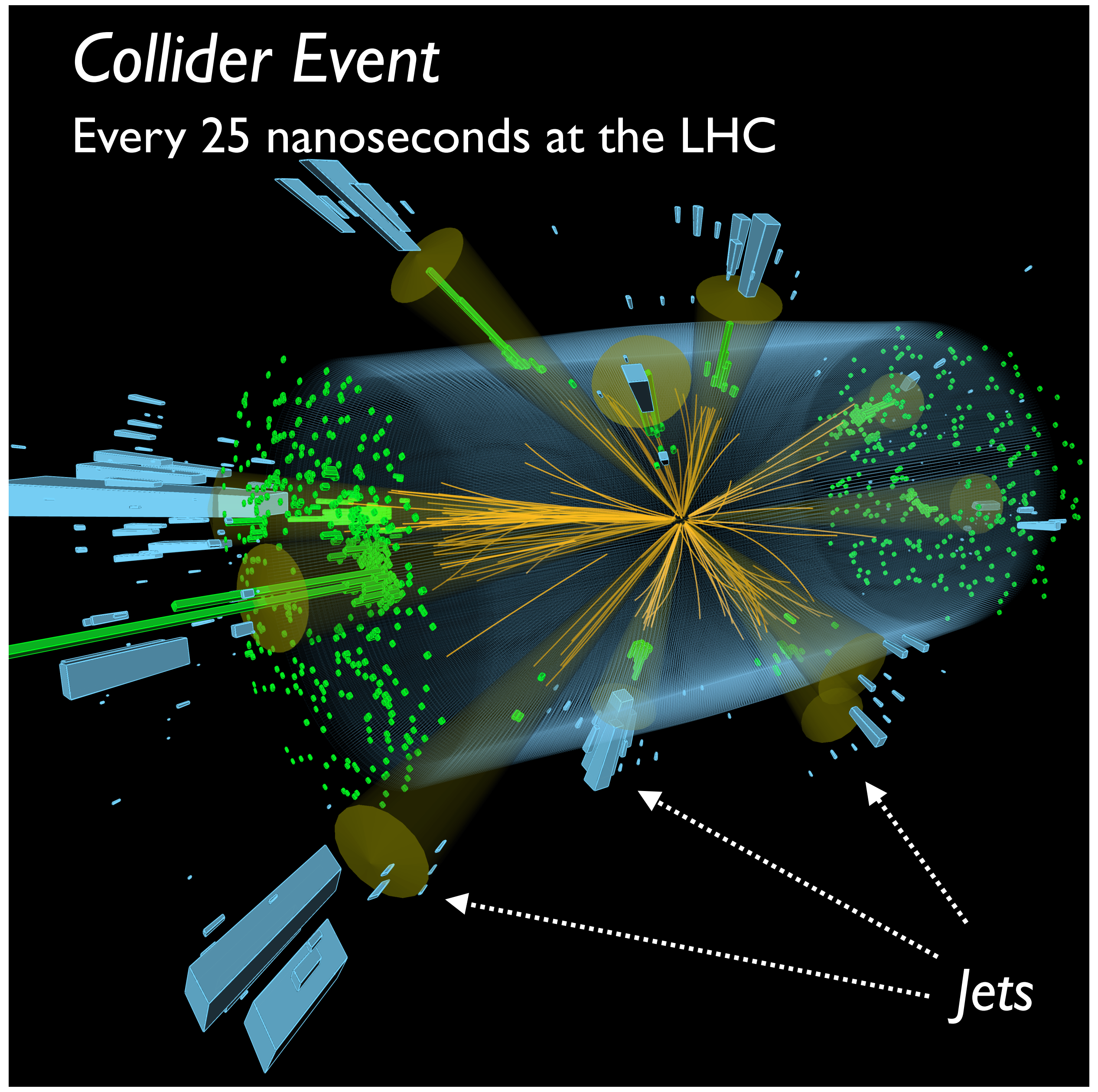
Figure 2: Laplacian eigenmaps, clustering done after embedding in **14 dimensions**

We achieve better quality clustering and are able to identify outliers, even with an embedding into 3 dimensions (**reduction from 117 dimensions**)

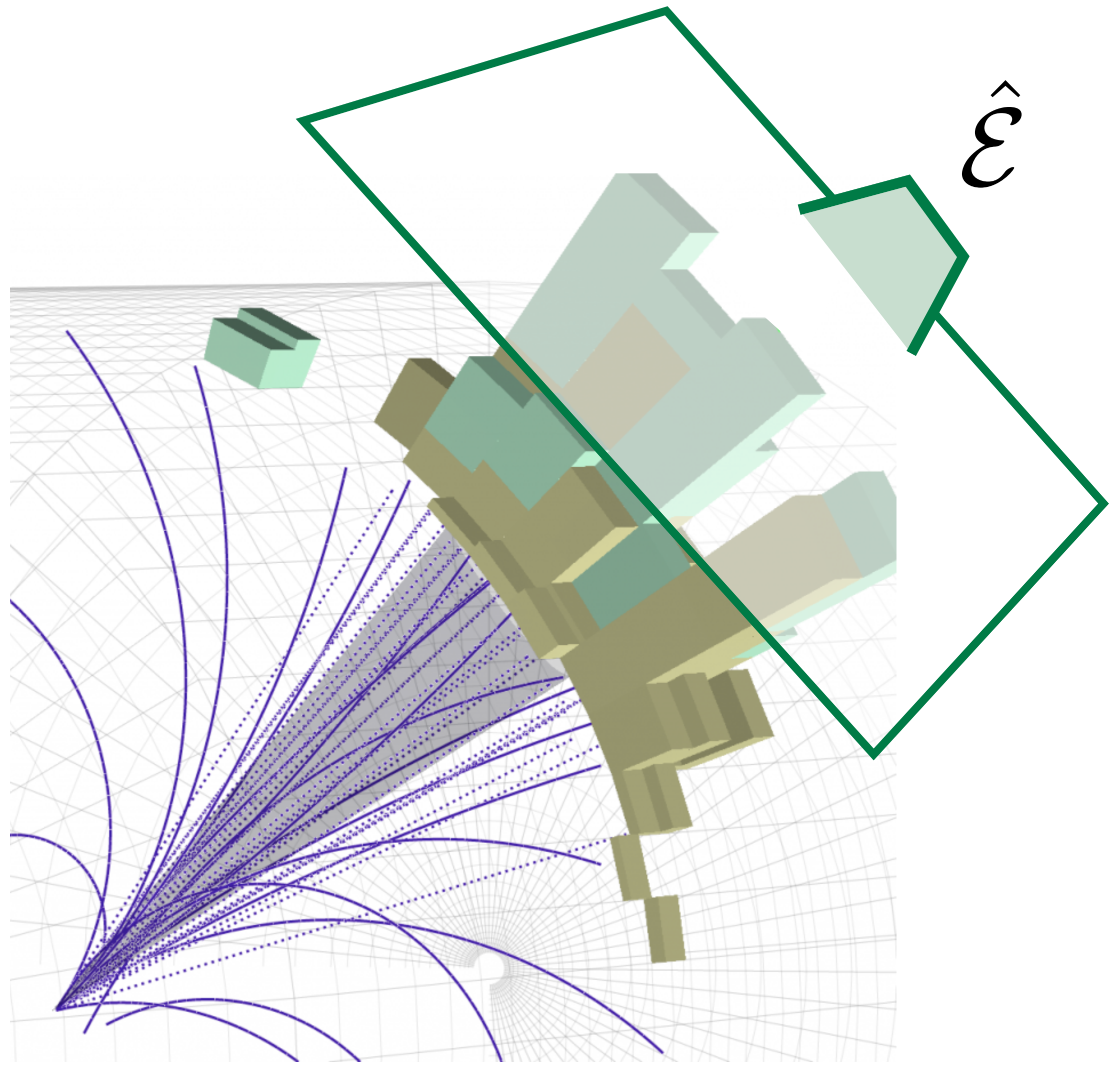
Applications: HEP?

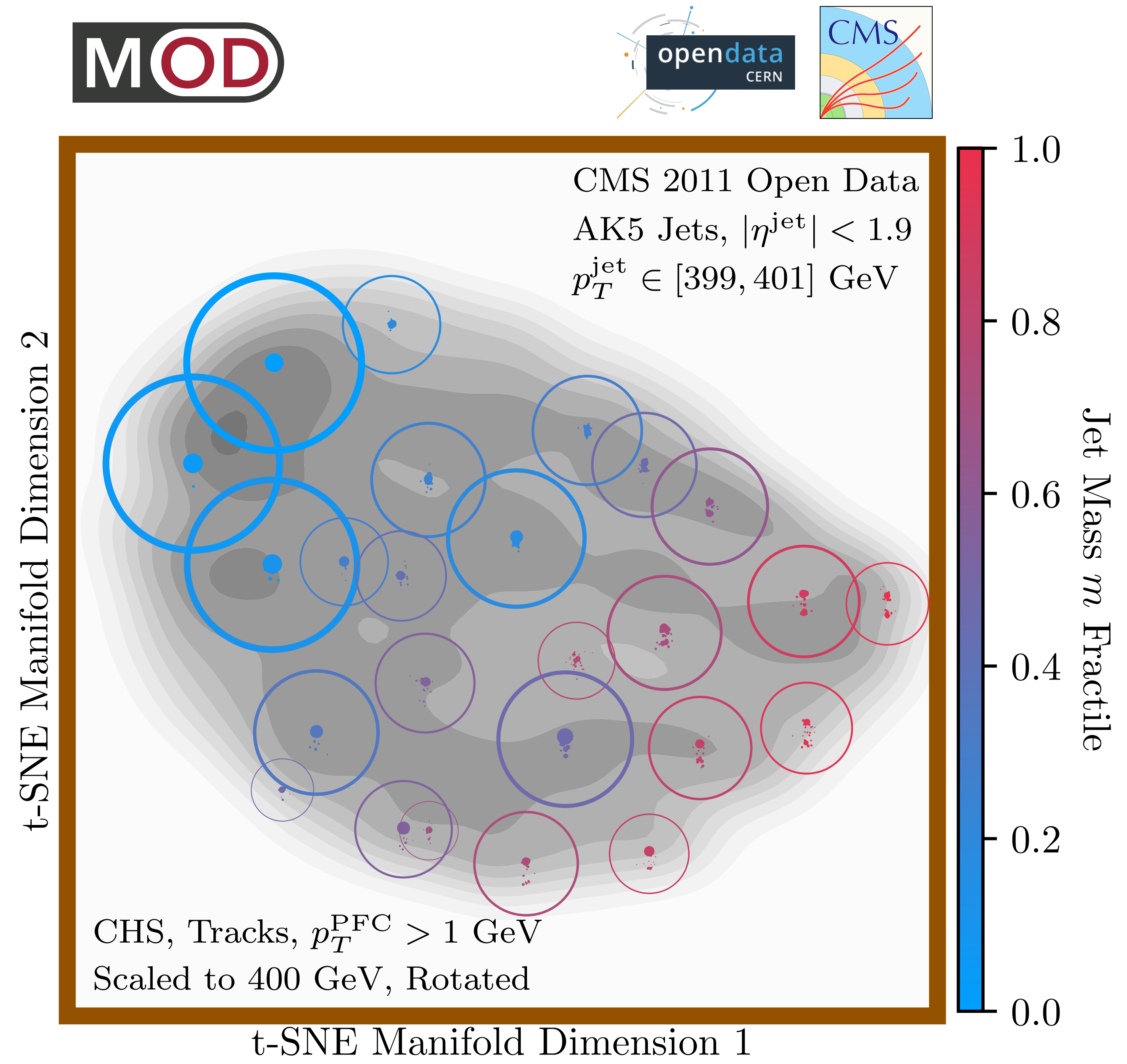
Future directions for HEP applications

Recall: Collider events have a geometry that can be used to organize, categorize and interpret collisions, e.g.,



Events summarized by "energy flow"



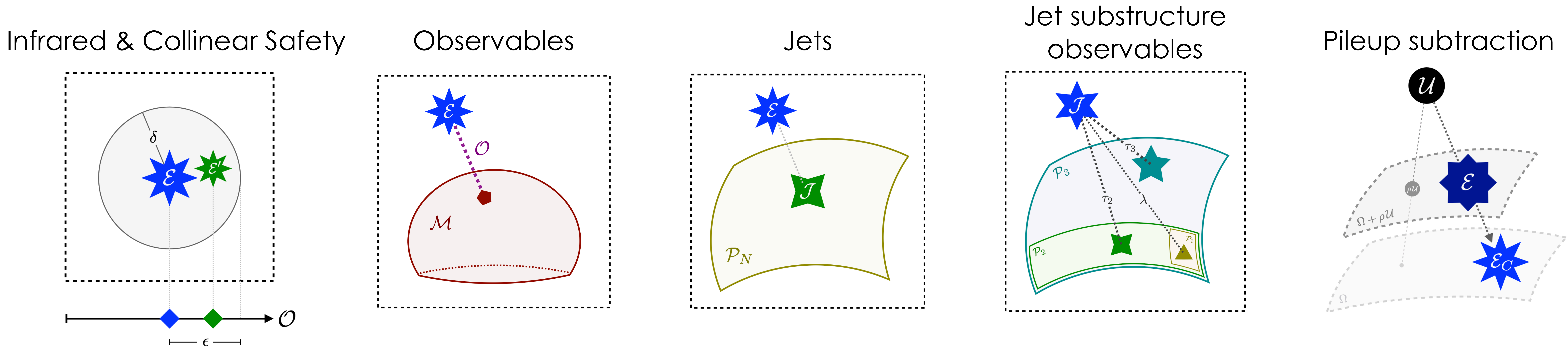


Komiske, Metodiev, Thaler,
The hidden geometry of particle collisions
J. HEP, 6 (2020)
The Metric Space of Collider Events
Phys. Rev. Lett. 123, 041801 (2019)

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

Recall: Collider events have a geometry that can be used to organize, categorize and interpret collisions, e.g.,



Komiske, Metodiev, Thaler,
The hidden geometry of particle collisions
J. HEP, 6 (2020)
The Metric Space of Collider Events
Phys. Rev. Lett. 123, 041801 (2019)

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science:
https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

Does quantum dynamics make sense on this metric space?

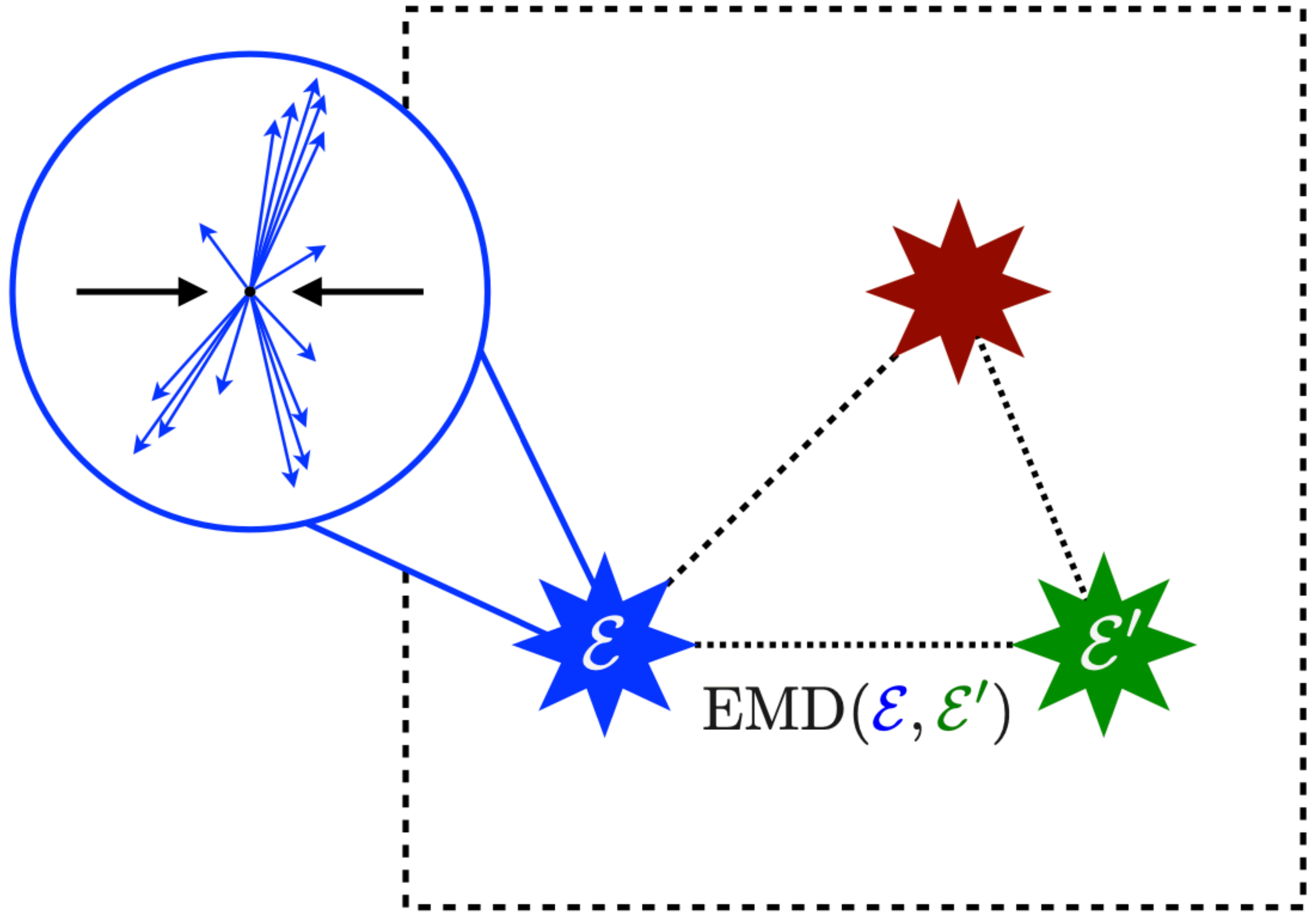
energy (GeV)

cost to move energy

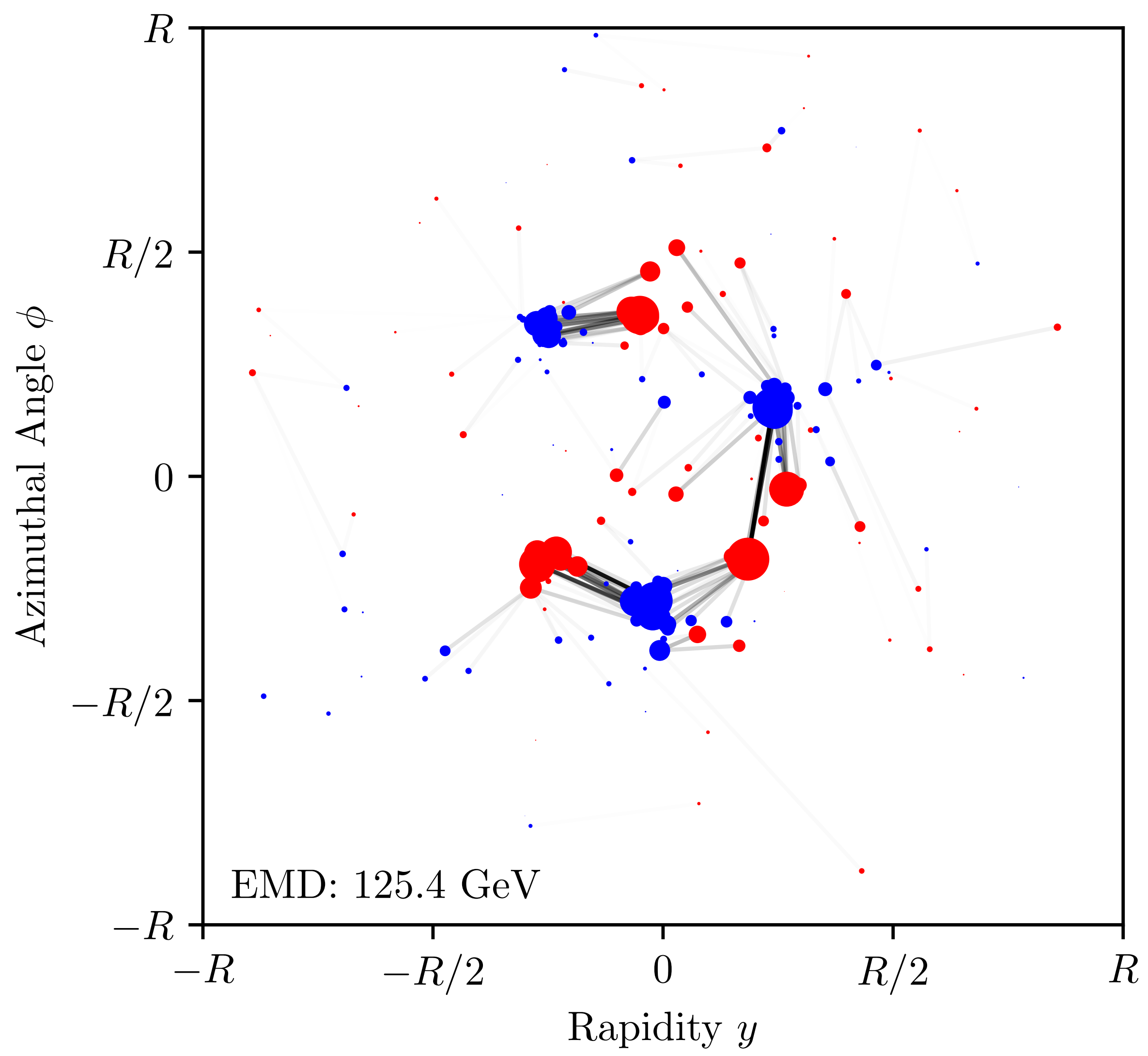
cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\,\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events



$\mathcal{E} \longrightarrow \mathcal{E}'$



* for R sufficiently large

i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

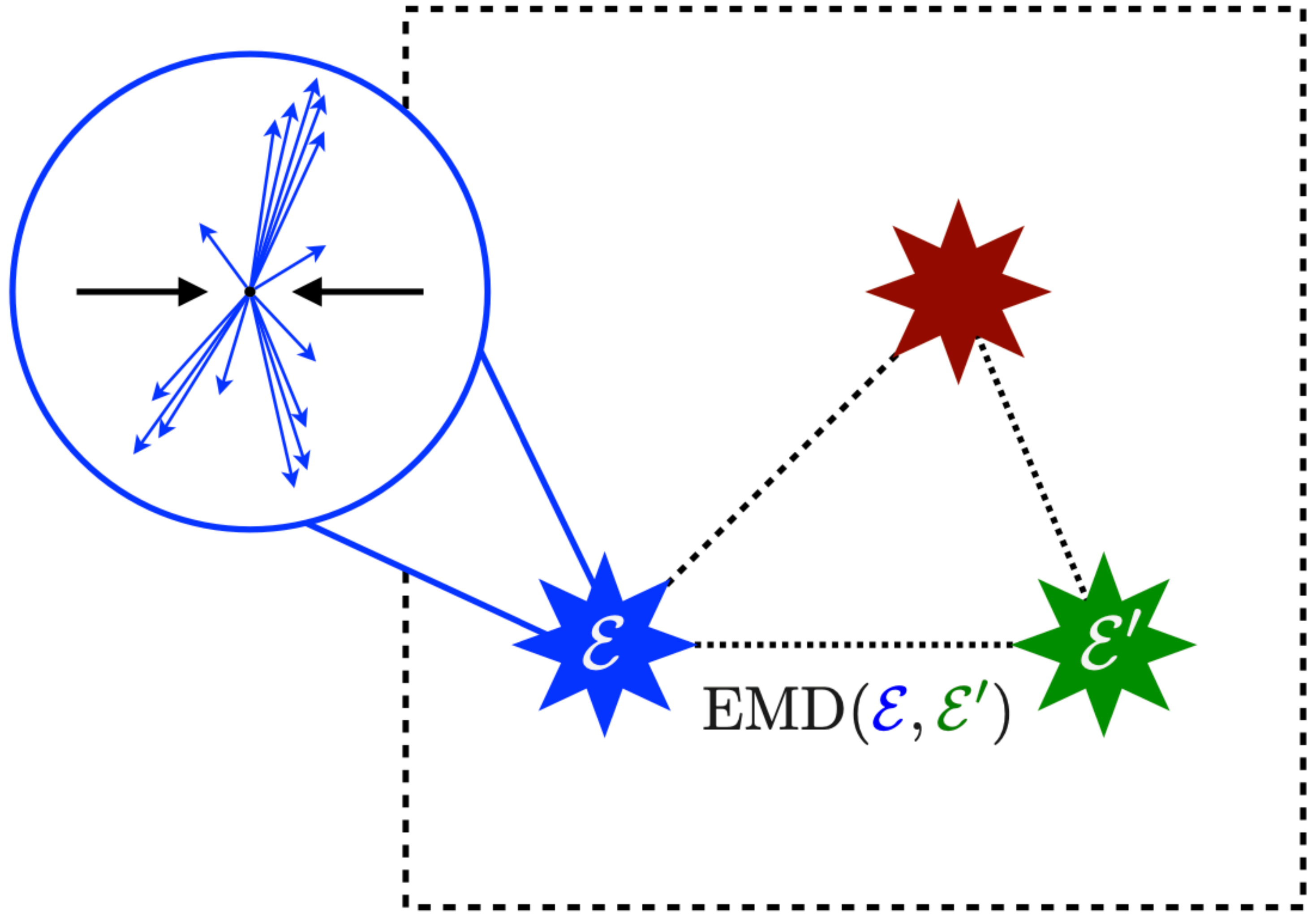
Does quantum dynamics make sense on this metric space?

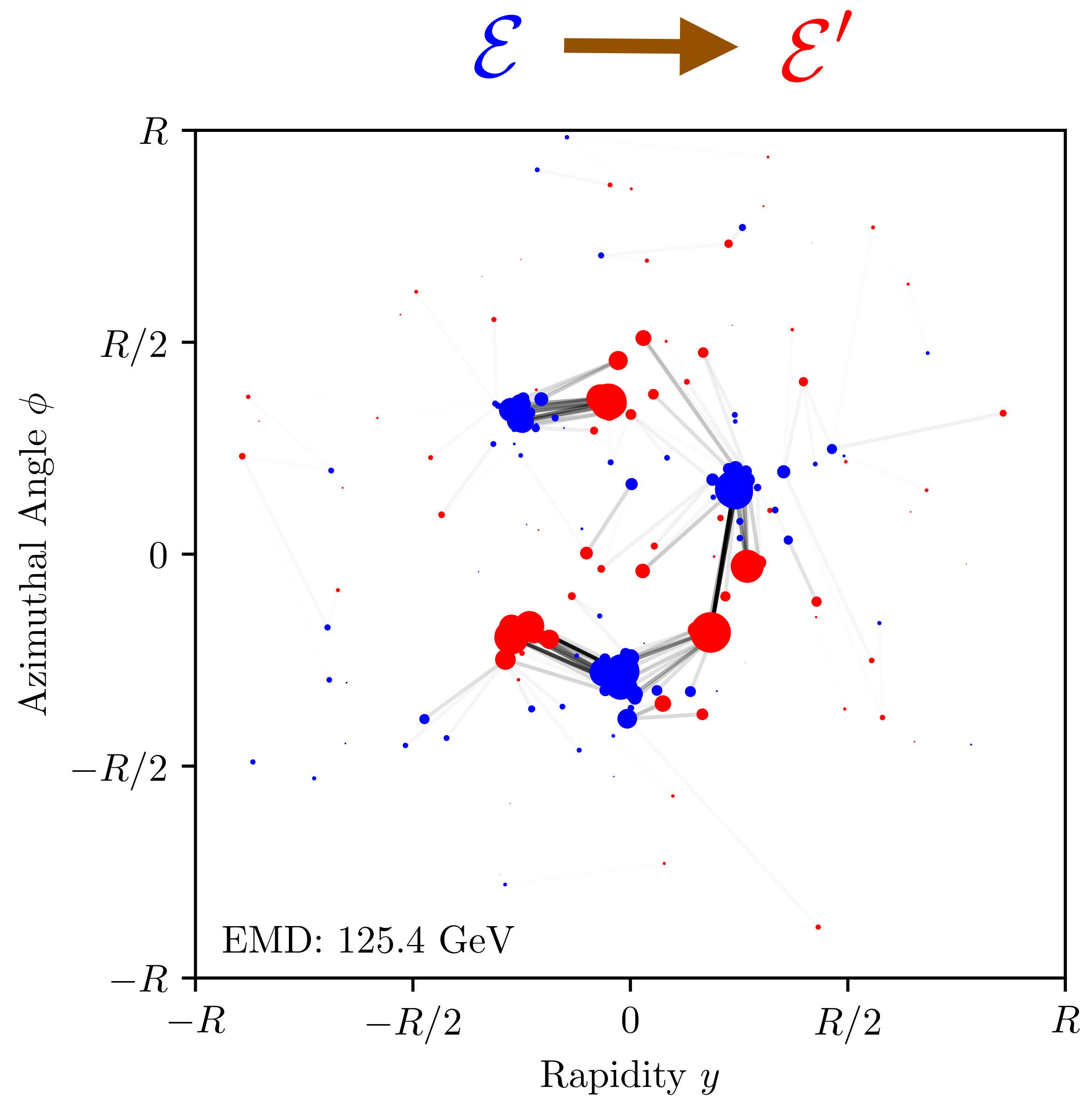
- *Not a Riemannian manifold!*

energy (GeV) — cost to move energy — cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events





* for R sufficiently large
i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

Does quantum dynamics make sense on this metric space?

- *Not a Riemannian manifold!*
    - but maybe *formally* it is?

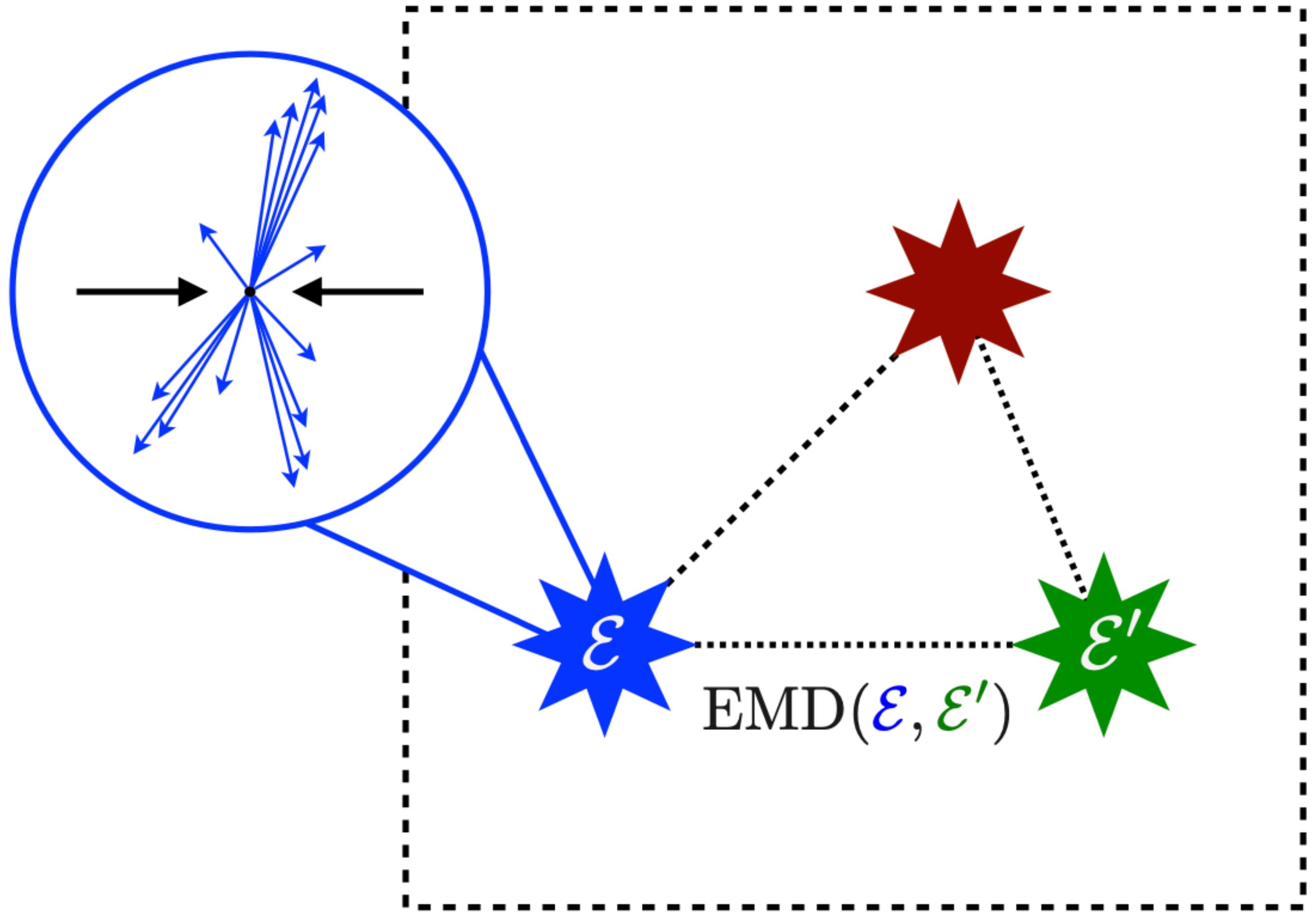
energy (GeV)

cost to move energy

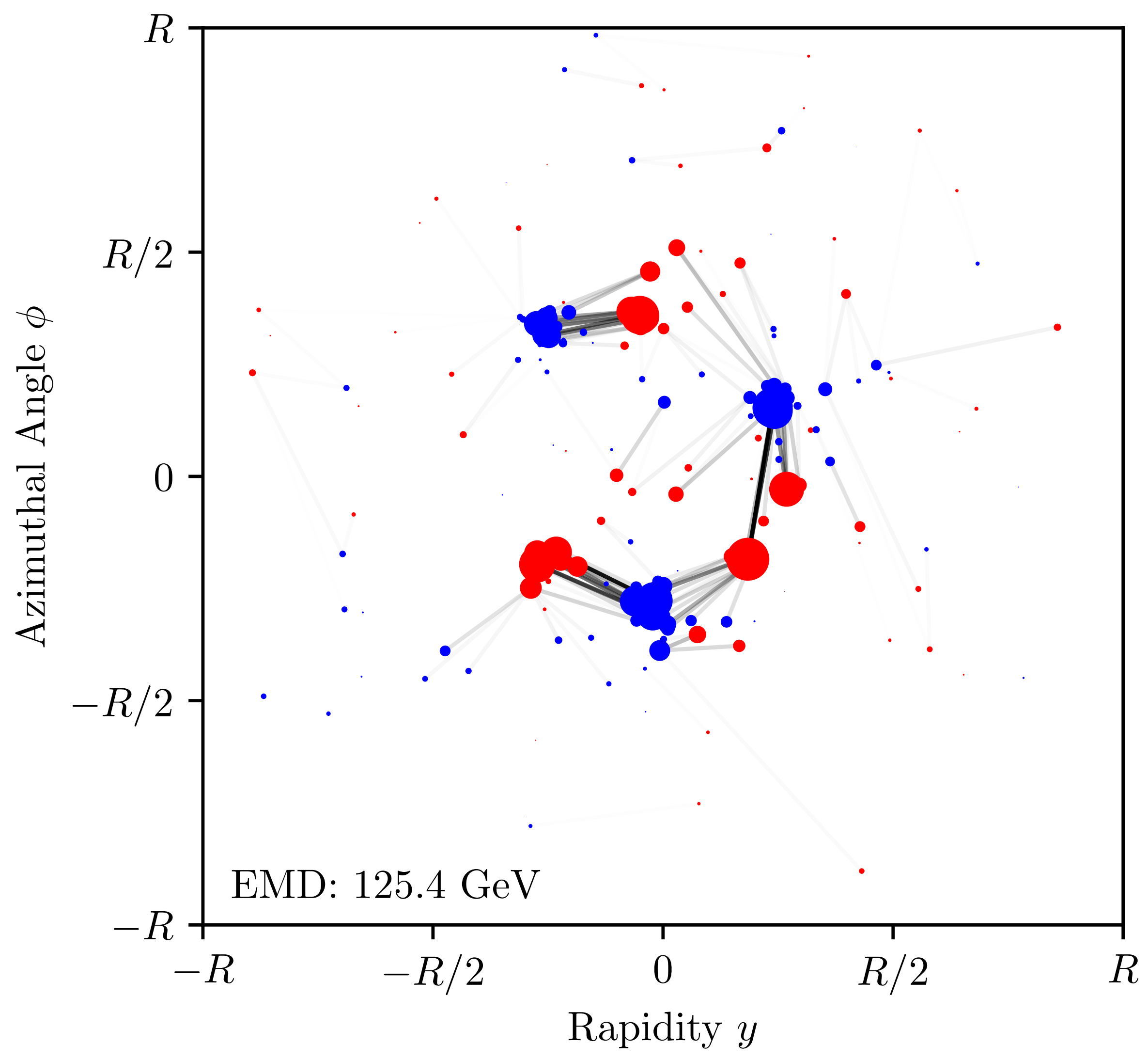
cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\,\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events



$\mathcal{E} \longrightarrow \mathcal{E}'$



* for *R* sufficiently large
i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

Does quantum dynamics make sense on this metric space?

- *Not a Riemannian manifold!*
  - but maybe *formally* it is?
- What is the analogous quantum-classical correspondence principle?

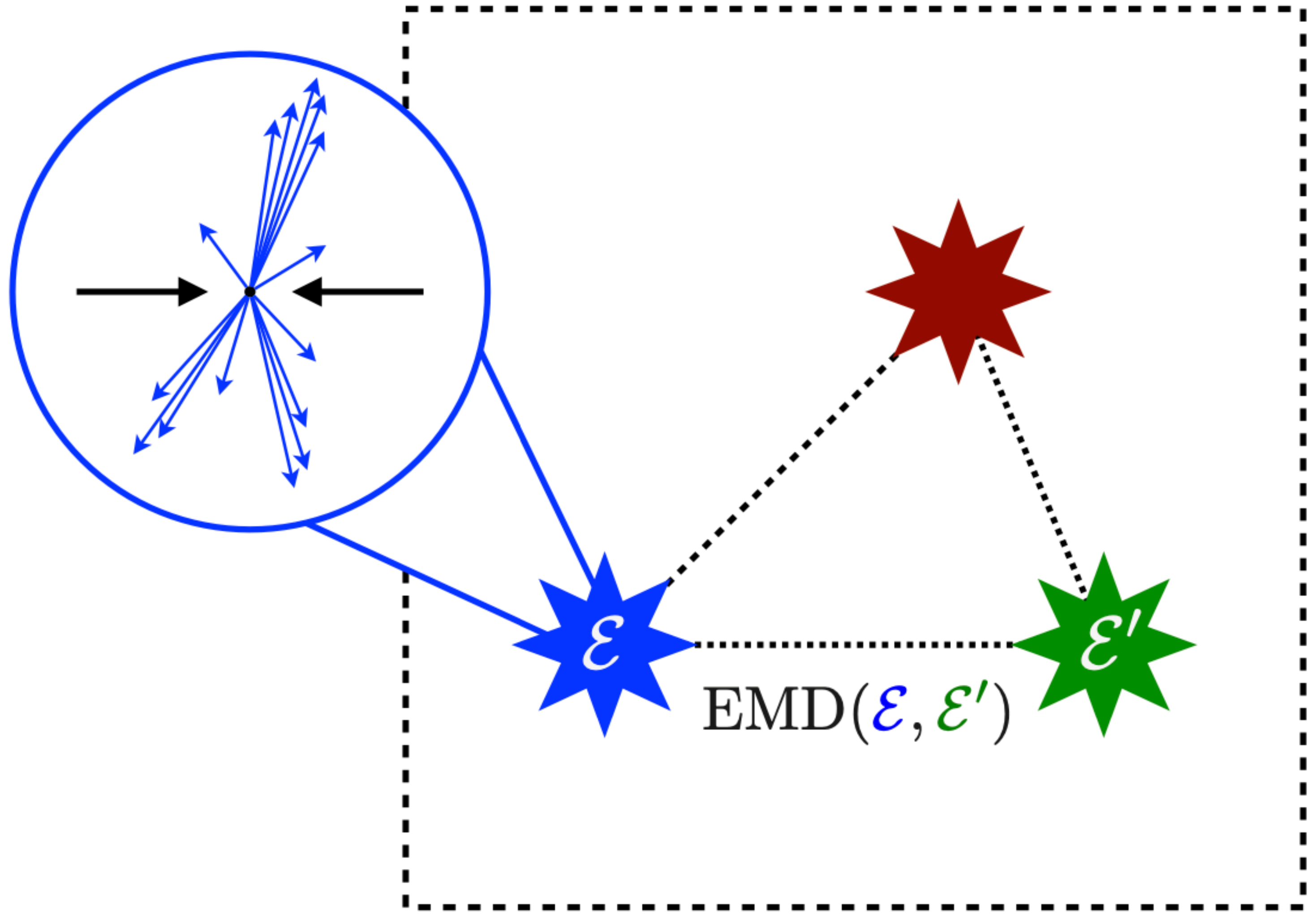
energy (GeV)

cost to move energy

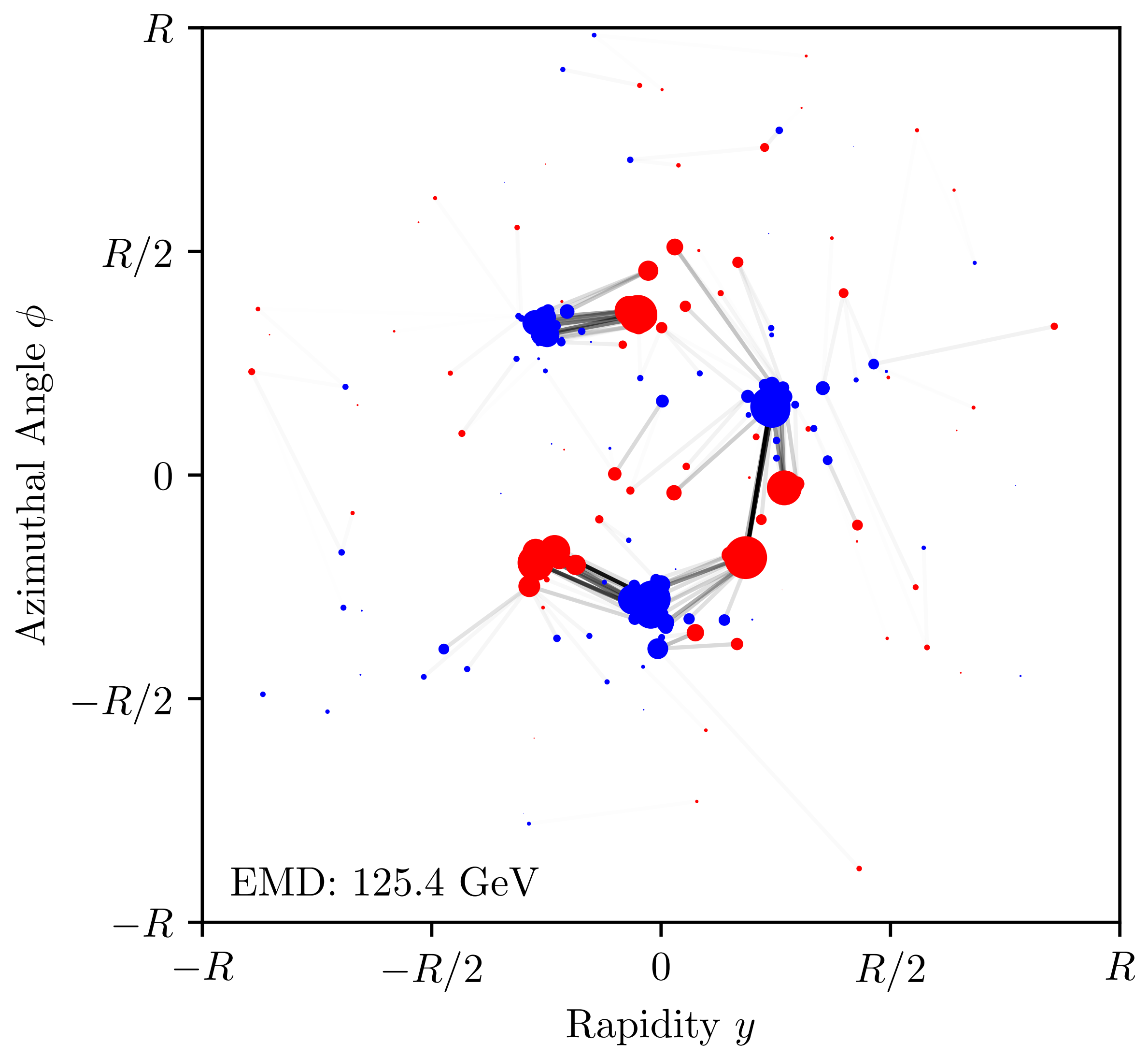
cost to create energy

$$\text{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\frac{\text{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events



$\mathcal{E} \rightarrow \mathcal{E}'$



* for R sufficiently large

i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

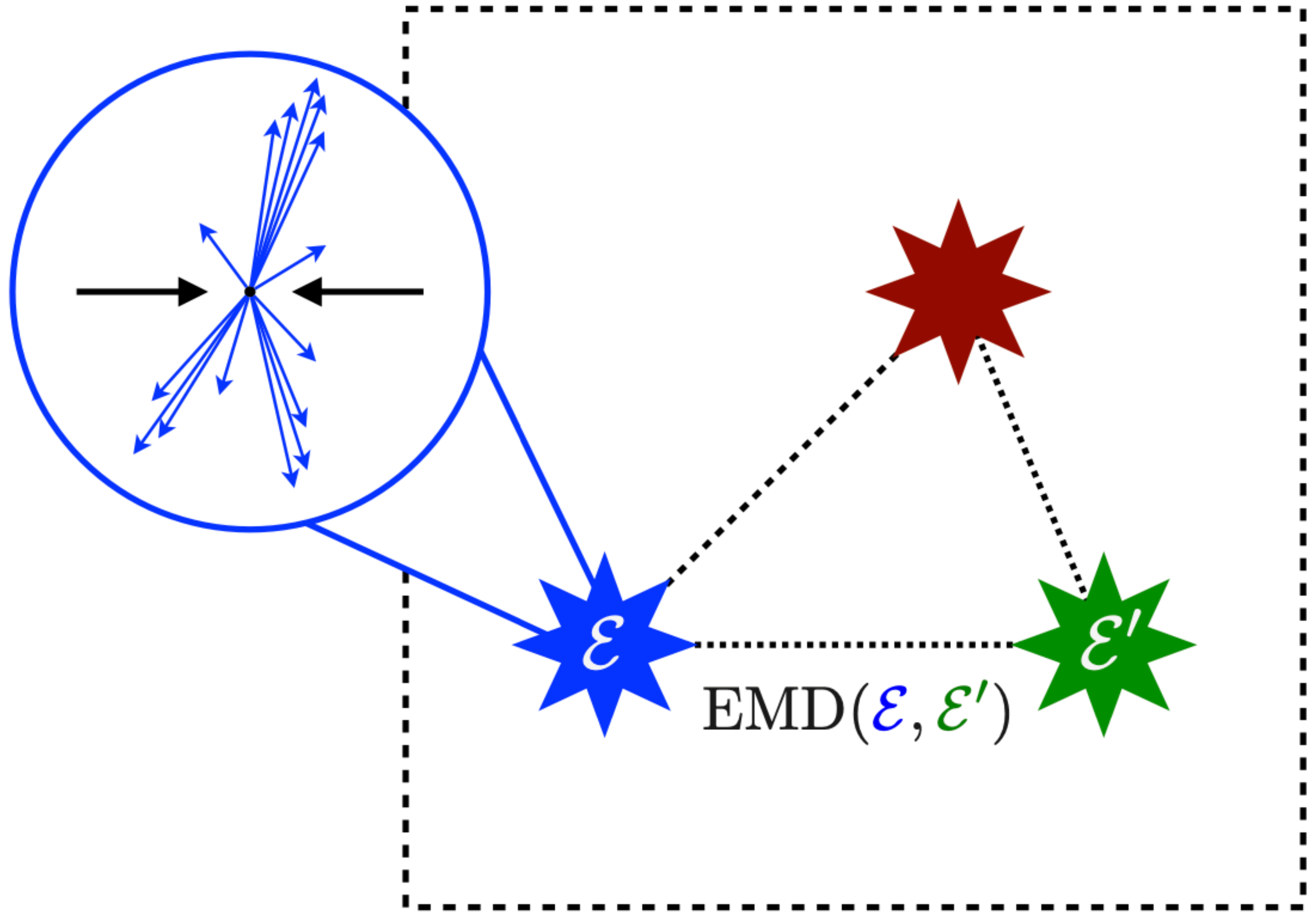
Does quantum dynamics make sense on this metric space?

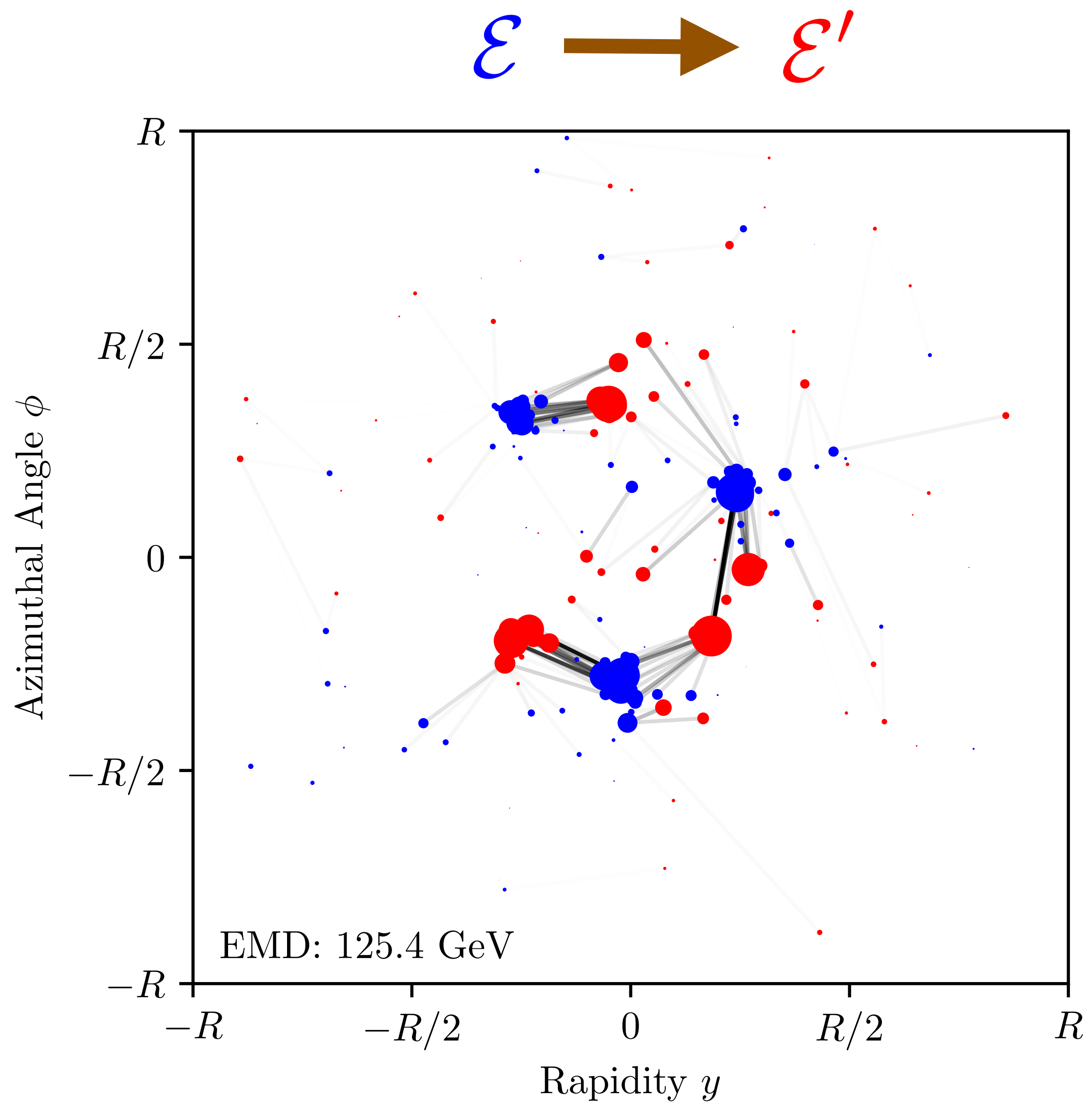
- *Not a Riemannian manifold!*
  - but maybe *formally* it is?
- What is the analogous quantum-classical correspondence principle?
- What is the quantization of the *energy flow* $\mathrm{EMD}(\,\cdot\,,\cdot\,\cdot\,)$?

energy (GeV) — cost to move energy — cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\,\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events





* for R sufficiently large
i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

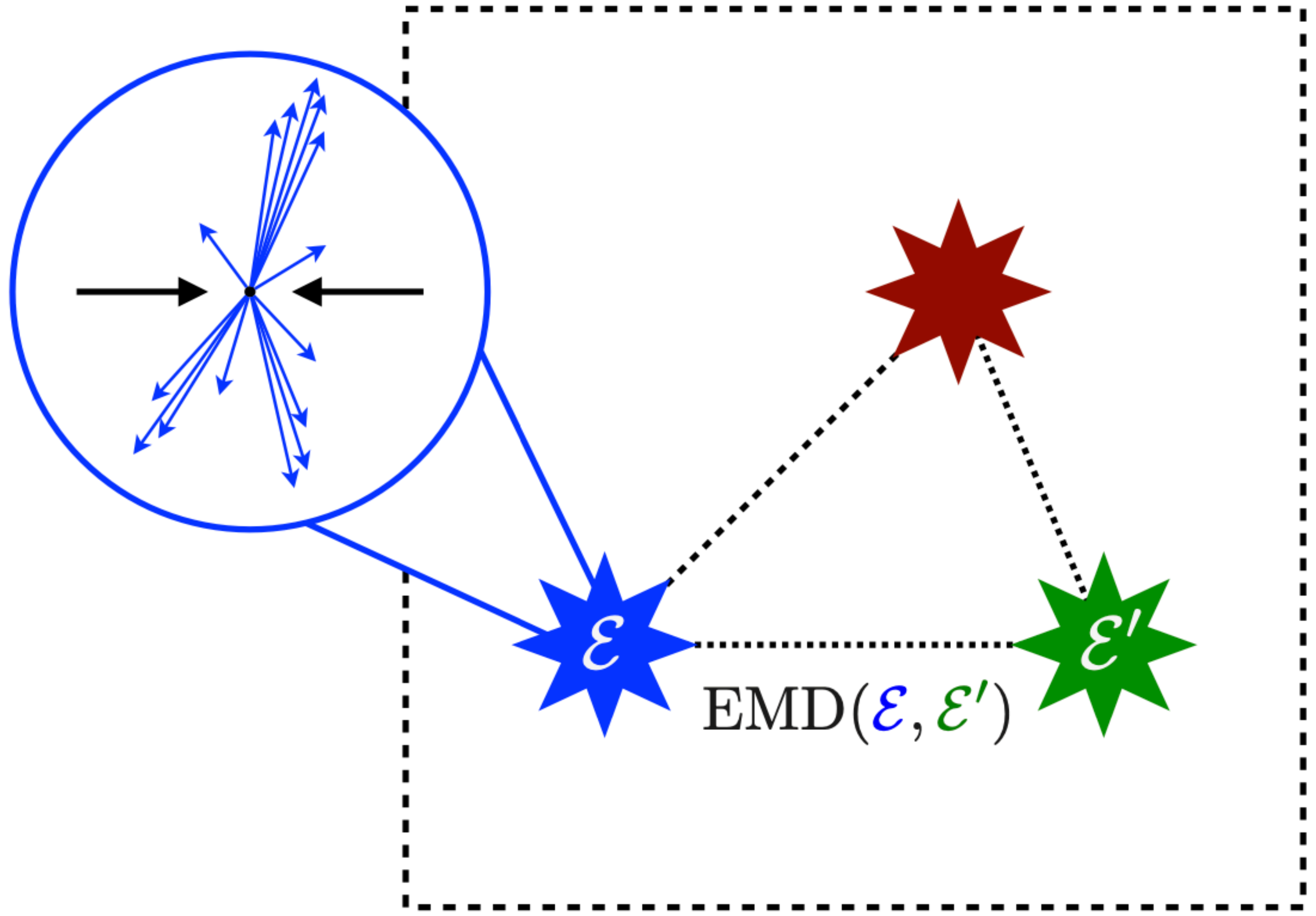
Does quantum dynamics make sense on this metric space?

- *Not a Riemannian manifold!*
  - but maybe *formally* it is?
- What is the analogous quantum-classical correspondence principle?
- What is the quantization of the *energy flow* $\mathrm{EMD}(\,\cdot\,,\cdot\cdot\,)$?
  - We quantized the geodesic flow. Optimal transport has dynamical (PDE) descriptions as well.

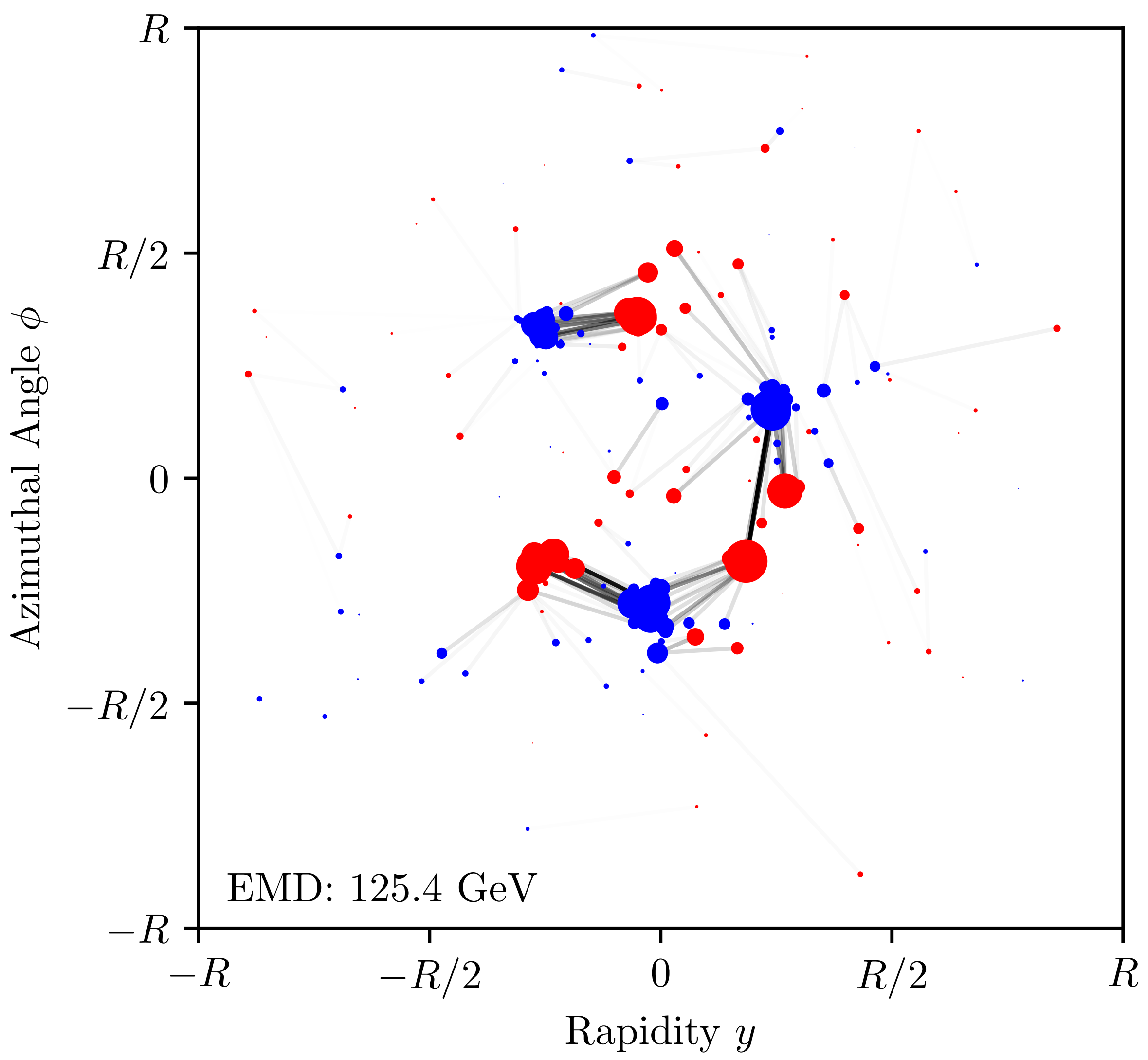
energy (GeV) — cost to move energy — cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\,\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events



$\mathcal{E} \longrightarrow \mathcal{E}'$



* for R sufficiently large
i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

Future directions for HEP applications

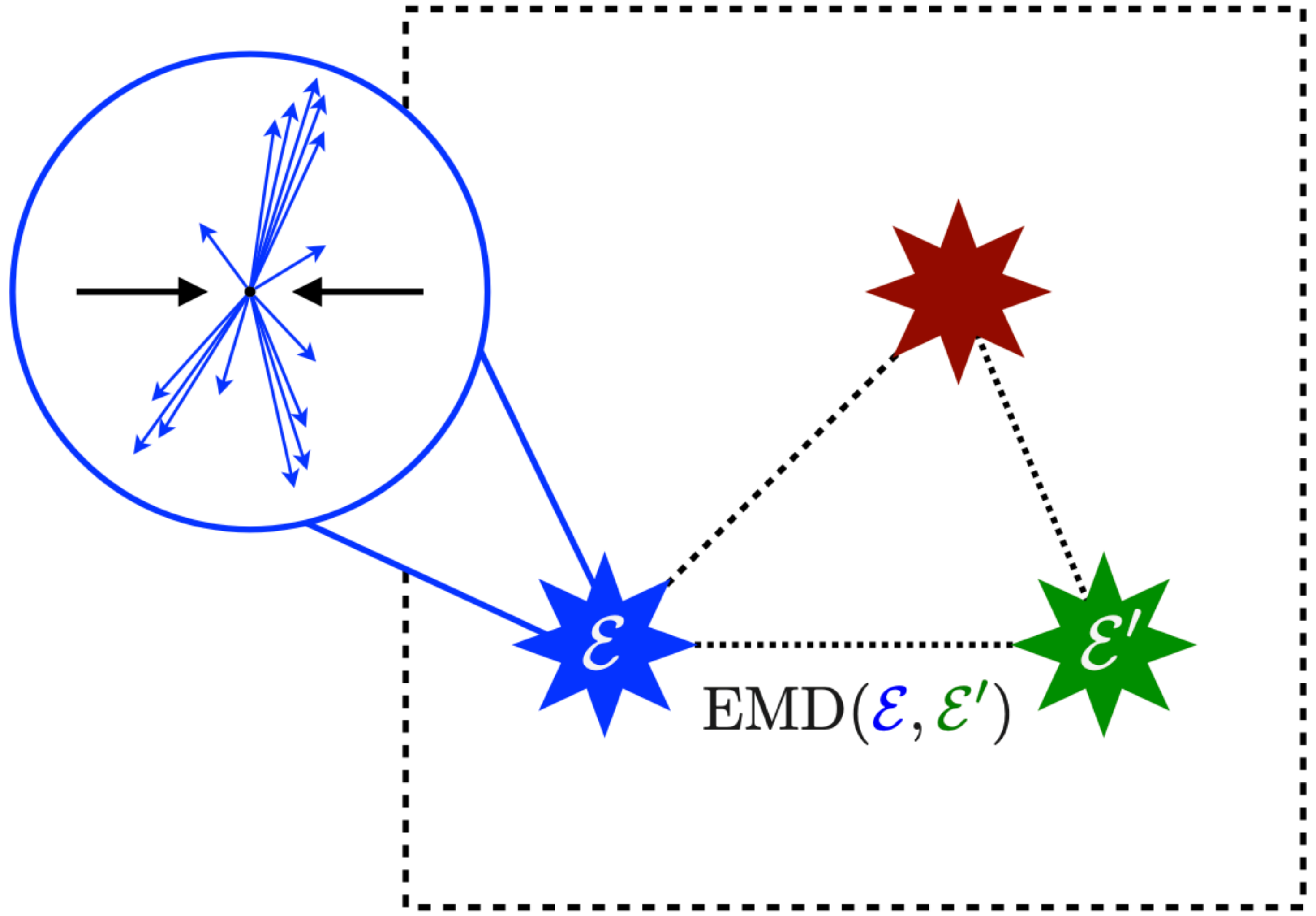
Does quantum dynamics make sense on this metric space?

- *Not a Riemannian manifold!*
    - but maybe *formally* it is?
- What is the analogous quantum-classical correspondence principle?
- What is the quantization of the *energy flow* $\mathrm{EMD}(\,\cdot\,,\cdot\cdot\,)$?
    - We quantized the geodesic flow. Optimal transport has dynamical (PDE) descriptions as well.
- The EMD is solvable quickly, but optimal transport plans are harder. Could quantization resolve optimal transport plans more rapidly?

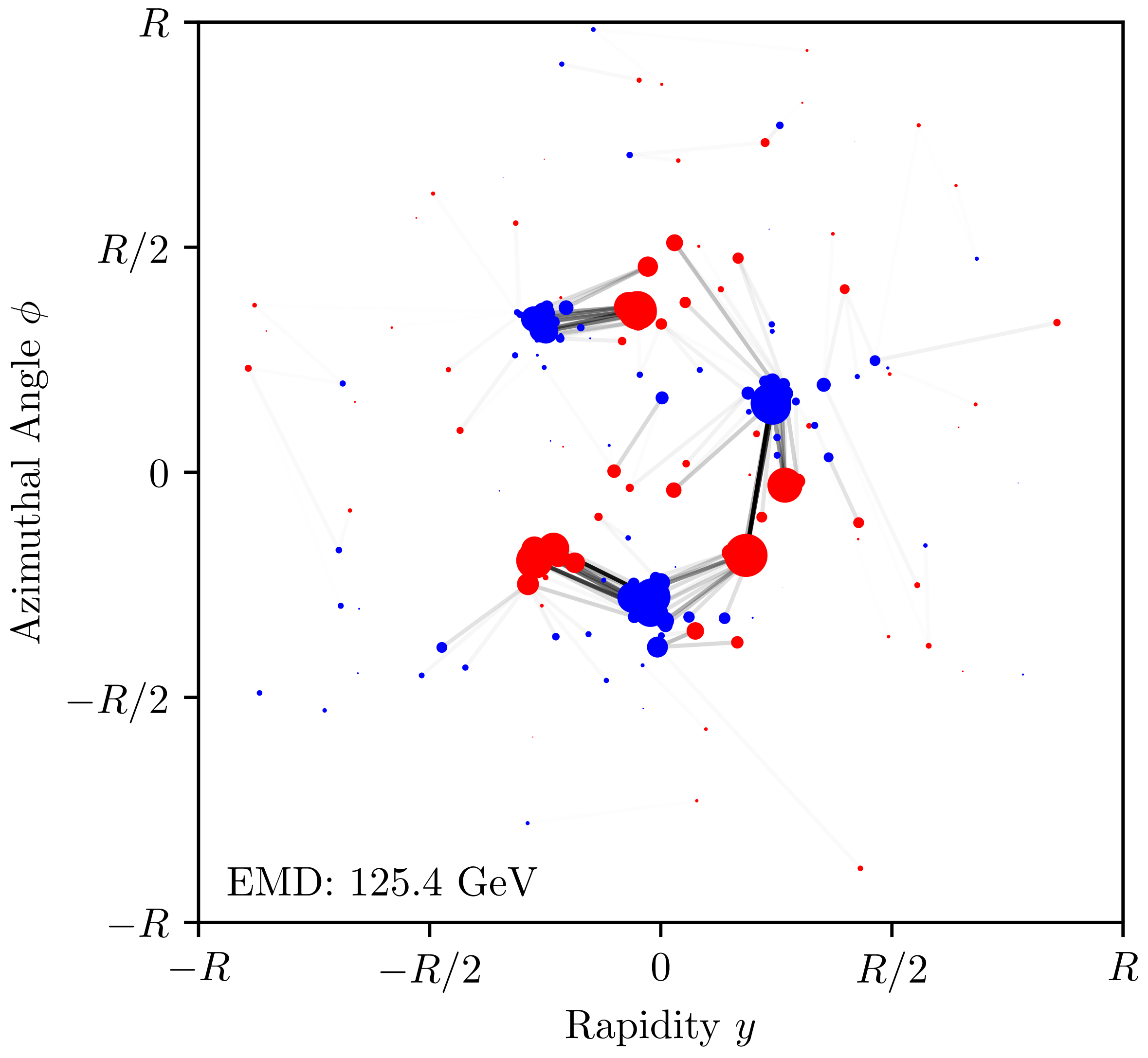
energy (GeV) — cost to move energy — cost to create energy

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') := \min_{\{f_{i,j}\geq 0\}} \sum_{i=1}^{N}\sum_{j=1}^{N'} f_{i,j}\,\frac{\mathrm{dist}(\hat{n}_i,\hat{n}_j)}{R} + \left|\sum_{i=1}^{N} E_i + \sum_{j=1}^{N'} E_j\right|$$

Metric* space of possible events



$\mathcal{E} \longrightarrow \mathcal{E}'$



* for R sufficiently large
i.e., $R \geq$ jet radius for conical jets

Figures from talk by Jesse Thaler, University of Chicago and Caltech AI+Science: https://www.youtube.com/watch?v=BMBSAWUxBn4

What about quantum algorithms?

We have a rigorous convergence theory and classical algorithm.

Inputs: $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$
Output: Propagated state $[\psi_h^\zeta](t)$
procedure PROPAGATE
    for $i, j = 1 : N$ **do** $[T_\epsilon]_{i,j} \leftarrow k(||v_i - v_j||^2/\epsilon)$
    $D_\epsilon \leftarrow$ diagonal matrix $\left(\sum_{j=1}^N [T_\epsilon]_{i,j}\right)_{1 \leq i \leq N}$
    $\Delta_{\epsilon,N} \leftarrow \frac{4(I_N - D_\epsilon^{-1} T_\epsilon)}{\epsilon}$
    $U_{\epsilon,N}^t \leftarrow \exp(-it\sqrt{\Delta_{\epsilon,N}})$
    $h \leftarrow \epsilon^{\frac{1}{(2+\alpha)}}$
    $p_0 \leftarrow v_j - v^*$ for v_j closest to point v^*
    while $1 \leq \ell \leq N$ **do**
        $[\psi_h^\zeta]_\ell \leftarrow e^{-||v_\ell - v^*||^2/2h} e^{\frac{i}{h}(v_\ell - v^*)^\intercal p_0/||p_0||}$
    return $[\psi_h^\zeta](t) = U_{\epsilon,N}^t [\psi_h^\zeta]$

What about quantum algorithms?

We have a rigorous convergence theory and classical algorithm.

- Uses only quantum primitives: unitary operators acting on localized wavepackets.

Inputs: $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$

Output: Propagated state $[\psi_h^\zeta](t)$

procedure PROPAGATE

  for $i, j = 1 : N$ **do** $[T_\epsilon]_{i,j} \leftarrow k(\|v_i - v_j\|^2/\epsilon)$

  $D_\epsilon \leftarrow$ diagonal matrix $\left(\sum_{j=1}^N [T_\epsilon]_{i,j}\right)_{1 \leq i \leq N}$

  $\Delta_{\epsilon,N} \leftarrow \frac{4(I_N - D_\epsilon^{-1} T_\epsilon)}{\epsilon}$

  $U_{\epsilon,N}^t \leftarrow \exp(-it\sqrt{\Delta_{\epsilon,N}})$

  $h \leftarrow \epsilon^{\frac{1}{(2+\alpha)}}$

  $p_0 \leftarrow v_j - v^*$ for v_j closest to point v^*

  while $1 \leq \ell \leq N$ **do**

    $[\psi_h^\zeta]_\ell \leftarrow e^{-\|v_\ell - v^*\|^2/2h} e^{\frac{i}{h}(v_\ell - v^*)^\top p_0/\|p_0\|}$

  return $[\psi_h^\zeta](t) = U_{\epsilon,N}^t [\psi_h^\zeta]$

What about quantum algorithms?

We have a rigorous convergence theory and classical algorithm.

- Uses only quantum primitives: unitary operators acting on localized wavepackets.
- Convergence theory tells a *negative* story for Hamiltonian simulation: *not enough sparsity*.

Inputs: $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$
Output: Propagated state $[\psi_h^\zeta](t)$
procedure PROPAGATE
    for $i, j = 1 : N$ **do** $[T_\epsilon]_{i,j} \leftarrow k(||v_i - v_j||^2/\epsilon)$
    $D_\epsilon \leftarrow$ diagonal matrix $\left(\sum_{j=1}^N [T_\epsilon]_{i,j}\right)_{1 \leq i \leq N}$
    $\Delta_{\epsilon,N} \leftarrow \frac{4(I_N - D_\epsilon^{-1} T_\epsilon)}{\epsilon}$
    $U_{\epsilon,N}^t \leftarrow \exp(-it\sqrt{\Delta_{\epsilon,N}})$
    $h \leftarrow \epsilon^{\frac{1}{(2+\alpha)}}$
    $p_0 \leftarrow v_j - v^*$ for v_j closest to point v^*
    while $1 \leq \ell \leq N$ **do**
        $[\psi_h^\zeta]_\ell \leftarrow e^{-||v_\ell - v^*||^2/2h} e^{\frac{i}{h}(v_\ell - v^*)^\top p_0/||p_0||}$
    return $[\psi_h^\zeta](t) = U_{\epsilon,N}^t [\psi_h^\zeta]$

What about quantum algorithms?

We have a rigorous convergence theory and classical algorithm.

- Uses only quantum primitives: unitary operators acting on localized wavepackets.
- Convergence theory tells a *negative* story for Hamiltonian simulation: *not enough sparsity*.
- … Also a *positive story*: With sufficient spectral truncation, we can evolve delta states.

Inputs: $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$

Output: Propagated state $[\psi_h^\zeta](t)$

procedure PROPAGATE

  for $i, j = 1 : N$ **do** $[T_\epsilon]_{i,j} \leftarrow k(||v_i - v_j||^2/\epsilon)$

  $D_\epsilon \leftarrow$ diagonal matrix $\left(\sum_{j=1}^N [T_\epsilon]_{i,j}\right)_{1 \leq i \leq N}$

  $\Delta_{\epsilon,N} \leftarrow \frac{4(I_N - D_\epsilon^{-1} T_\epsilon)}{\epsilon}$

  $U_{\epsilon,N}^t \leftarrow \exp(-it\sqrt{\Delta_{\epsilon,N}})$

  $h \leftarrow \epsilon^{\frac{1}{(2+\alpha)}}$

  $p_0 \leftarrow v_j - v^*$ for v_j closest to point v^*

  while $1 \leq \ell \leq N$ **do**

    $[\psi_h^\zeta]_\ell \leftarrow e^{-||v_\ell - v^*||^2/2h} e^{\frac{i}{h}(v_\ell - v^*)^\top p_0/||p_0||}$

  return $[\psi_h^\zeta](t) = U_{\epsilon,N}^t [\psi_h^\zeta]$

What about quantum algorithms?

We have a rigorous convergence theory and classical algorithm.

- Uses only quantum primitives: unitary operators acting on localized wavepackets.
- Convergence theory tells a *negative* story for Hamiltonian simulation: *not enough sparsity*.
- … Also a *positive story*: With sufficient spectral truncation, we can evolve delta states.
- Fast implementation of $U^t_{\epsilon,N}$ has widespread implications for quantum algorithms for manifold learning.

1: **Inputs:** $X_N = \{v_1, ..., v_N\}, v^*, \epsilon > 0, \alpha \geq 1, t > 0$
2: **Output:** Propagated state $[\psi^\zeta_h](t)$
3: **procedure** PROPAGATE
4: **for** $i, j = 1 : N$ **do** $[T_\epsilon]_{i,j} \leftarrow k(||v_i - v_j||^2/\epsilon)$
5: $D_\epsilon \leftarrow$ diagonal matrix $\left(\sum_{j=1}^N [T_\epsilon]_{i,j}\right)_{1\leq i \leq N}$
6: $\Delta_{\epsilon,N} \leftarrow \frac{4(I_N - D_\epsilon^{-1} T_\epsilon)}{\epsilon}$
7: $U^t_{\epsilon,N} \leftarrow \exp(-it\sqrt{\Delta_{\epsilon,N}})$
8: $h \leftarrow \epsilon^{\frac{1}{(2+\alpha)}}$
9: $p_0 \leftarrow v_j - v^*$ for v_j closest to point v^*
10: **while** $1 \leq \ell \leq N$ **do**
11: $[\psi^\zeta_h]_\ell \leftarrow e^{-||v_\ell - v^*||^2/2h} e^{\frac{i}{h}(v_\ell - v^*)^\intercal p_0/||p_0||}$
12: **return** $[\psi^\zeta_h](t) = U^t_{\epsilon,N}[\psi^\zeta_h]$

Thank you

Manifold learning via quantum dynamics

Akshat Kumar[1,2,a] and Mohan Sarovar[3,b]
[1] *Department of Mathematics, Clarkson University, Potsdam, NY 13699 USA*
[2] *Instituto de Telecomunicações, Lisbon, Portugal*
[3] *Sandia National Laboratories, Livermore, California 94550, USA*

arXiv:2112.11161

Shining light on data: Geometric data analysis through quantum dynamics

Akshat Kumar,[1*] Mohan Sarovar[2*]

arXiv:2212.00682

ON A QUANTUM-CLASSICAL CORRESPONDENCE: FROM GRAPHS TO MANIFOLDS

AKSHAT KUMAR

arXiv:2112.10748