Hierarchical Computational Anatomy: Unifying the Molecular to Tissue Continuum Via Measure Representations of the Brain

Michael Miller[1*], Daniel Tward[2], and Alain Trouvé[3]

[1]Department of Biomedical Engineering & Kavli Neuroscience Discovery Institute & Center for Imaging Science, Johns Hopkins University, Baltimore, USA

[2]Departments of Computational Medicine & Neurology, University of California Los Angeles, Los Angeles, USA

[3]Centre Giovanni Borelli (UMR 9010), Ecole Normale Supérieure Paris-Saclay, Université Paris-Saclay

[*]Corresponding author. Email: mim@jhu.edu

September 16, 2021

Abstract

Objective: The objective of this research is to unify the molecular representations of spatial transcriptomics and cellular scale histology with the tissue scales of Computational Anatomy for brain mapping. Impact statement: We present a unified representation theory for brain mapping of the micro-scale phenotypes of molecular disease simultaneously with the connectomic scales of complex interacting brain circuits. Introduction: Mapping across coordinate systems in computational anatomy allows us to understand structural and functional properties of the brain at the millimeter scale. New measurement technologies such as spatial transcriptomics allow us to measure the brain cell by cell based on transcriptomic identity. We currently have no mathematical representations for integrating consistently the tissue limits with the molecular particle descriptions. The formalism derived here demonstrates the methodology for transitioning consistently from the molecular scale of quantized particles – as first introduced by Dirac as a generalized function – to the continuum and fluid mechanics scales appropriate for tissue. Methods: We introduce two methods based on notions of generalized functions and statistical mechanics. We use generalized functions expanded to include functional descriptions - electrophysiology, transcriptomic, molecular histology – to represent the molecular biology scale integrated with a Boltzman like procedure to pass from the sparse particles to empirical probability laws on the functional state of the tissue. Results: We demonstrate a unified mapping methodology for transferring molecular information in the transcriptome and histological scales to the human atlas scales for understanding Alzheimer's disease. Conclusions: We demonstrate a unified brain mapping theory for molecular and tissue scales.

1 Introduction

One of the striking aspects of the study of the brain in modern neurobiology is the fact that the distributions of discrete structures that make up physical tissue, from neural cells to synapses to genes and molecules, exists across nearly ten orders of magnitude in spatial scale. This paper focusses on the challenge of building multi-scale representations that simultaneously connect the quantum nano-scales of modern molecular biology for characterizing neural circuits architecture in the brain with the classical continuum representations at the anatomical gross and meso scales.

We have been highly motivated by the Cell Census Network project (BICCN [1]) which brings the nano and micron scales of single cell measures of RNA via spatial transcriptomics [2–4] coupled to the tissue scales of mouse atlases. The recent review on bridging scales from cells to physiology [5] motivates the mathematical framework presented herein. The recent emergence of spatial transcriptomics as Nature method of the year highlights the importance and ascendence of such approaches for understanding the dense metric structure of the brain which represent the coarse physiological atlas scales built up from dense imaging measurements at the cellular scales.

Specifically, in our work in Alzheimer's disease the BIOCARD study [6] we are examining pathological Tau at both the micro histological and macro atlas scales of Tau particle detections, from 10-100 μ m [7, 8] and to human magnetic resonance millimeter scales for examining entire circuits in the medial temporal lobe. In the mouse cell counting project we are examining single-cell spatial transcriptomics using modern RNA sequencing in dense tissue at the micron scale and its representations in the Allen atlas coordinates [9].

Most noteworthy for any representation is that at the finest micro scales nothing is smooth; the distributions of cells and molecules are more well described as random quantum counting processes in space [10]. In contrast, information associated to atlasing methods at the gross anatomical tissue and organ scales of Computational Anatomy extend smoothly [11–16]. Cross-sectionally and even cross-species, gross anatomical labelling is largely repeatable, implying information transfers and changes from one coordinate system to another smoothly. This is built into the representation theory of diffeomorphisms and soft matter tissue models for which advection and transport hold [17–21], principles upon which continuum mechanics and its analogues are based.

The focus of this paper is to build a coherent representation theory across scales. For this we view the micron to millimeter scales via the same representation theory called *mathematical measures*, building the finest micron scales from discrete units termed particle measures which represent molecules, synapses and cells. As they aggregate they form tissues. This measure representation allows us to understand subsets of tissues that contain discretely positioned and placed functional objects at the finest quantized scales and simultaneously pass smoothly to the classical continuum scales at which stable functional and anatomical representations exist. Since the study of the function of the brain on its geometric submanifolds -the gyri, sulci, subnuclei and laminae of cortex- are so important, we extend our general framework to exploit varifold measures [22] arising in the modern discipline of geometric measure theory. To be able to compare the brains we use diffeomorphisms as the comparator tool, with their action representing 3D varifold action which we formulate as "copy and paste" so that basic particle quantities that are conserved biologically are combined with greater multiplicity and not geometrically distorted as would be the case for measure transport.

The functional features are represented via singular delta-diracs at the finest micro structure scales. The functional feature is abstracted into a function space rich enough to accomodate the molecular machinery as represented by RNA or Tau particles, as well as electrophysiology associated to spiking neurons, or at the tissue scales of medical imaging dense contrasts of magnetic resonance images (MRIs). We pass to the classical function continuum via introduction of a scale-space that extends the descriptions of cortical micro-circuits to the meso and anatomical scales. This passage from the quantized features to the stochastic laws is in fact akin to the Boltzman program transferring the view from the Newtonian particles to the stable distributions describing them. For this we introduce a scale-space of kernel density transformations which allows us to retrieve the empirical averages represented by the determinism of the stochastic law consistent with our views of the macro tissue scales.

The representation provides a recipe for scale traversal in terms of a cascade of linear space scaling composed with non-linear functional feature mapping. Following the cascade implies every scale is a measure so that a universal family of measure norms can be introduced which simultaneously measure the disparety between brains in the orbit independent of the probing technology, RNA identities, Tau or amyloid histology, spike trains, or dense MR imagery.

Our brain measure model implies the existence of a sequence . This scale-space of pairs, the measure representation of the brain and the associated probing measurement technologies we call Brainspace. To formulate a consistent measurement and comparison

technology on Brainspace we construct a natural metric upon it allowing us to study its geometry and connectedness. The metric between brains is constructed via a Hamiltonian which defines the geodesic connections throughout scale space, providing for the first time a hierarchical representation that unifies micro to millimeter representation in the brain and makes Brainspace into a metric space. Examples of representation and comparision are given for Alzheimer's histology integrated to magnetic resonance imaging scales, and spatial transcriptomics.

2 Results

2.1 Measure Model of Brain Structures

To build a coherent theory we view the micron to anatomical scales via the same representation theory building upon discrete units termed particles or atoms. As they aggregate they form tissues. This is depicted in Figure 1 in which the top left panel shows mouse imaging of CUX1 labelling of the inner layers of mouse cortex (white) and CTP2 imaging of the outer layers (green) at 2.5 micron in plane resolution. Notice the discrete nature of the cells clearly resolved which form the layers of tissue which are the global macro scale features of layer 2,3,4 which stain more prolificaly in white and the outer layers 5,6 which stain more prolifically in green.

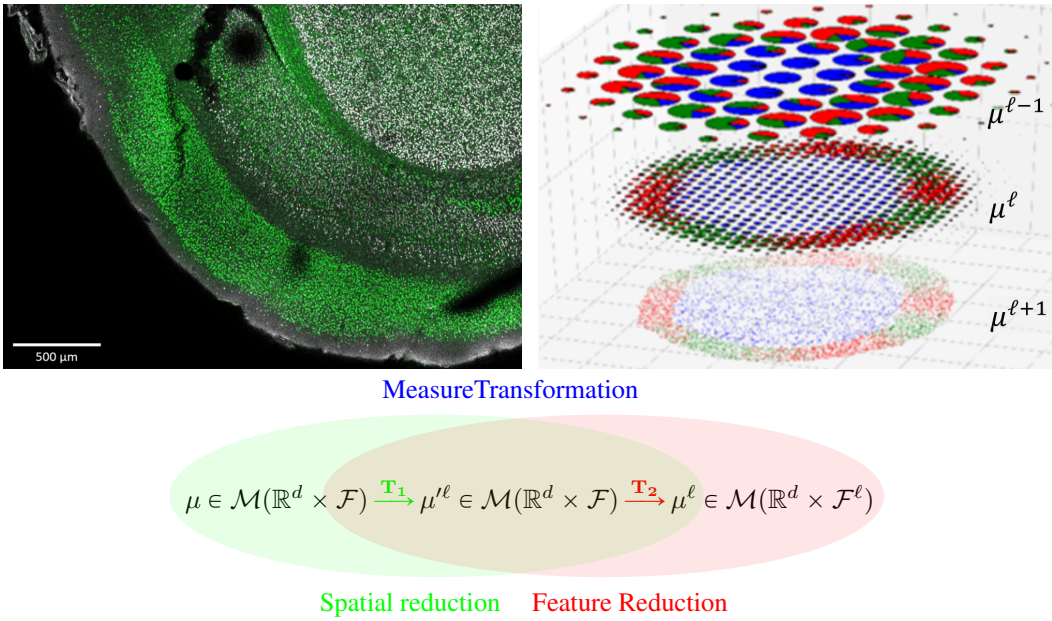

Figure 1: Top Left: Tissue from a NexCre+/-;Brn2fl/+ adult mouse mimicing a wild-type mouse with CUX1 labelling of layer ii/iii,iv and CTIP2 in layers v,vi in green. Shows sections at $2.5^2 \times 50\,\mu\text{m}^3$ 6 tile images, 1433×1973 pixels; taken from Uli Mueller. Top Right: Showing the abstraction of a coarse-to-fine hierarchy $\mu^{\ell-1}, \mu^{\ell}, \mu^{\ell+1}$ with fine molecular scales shown at the bottom with colors depicting $\mathcal{F}$ function ascending scales. Bottom: Space and function transformation shown as a composition $\mu \xrightarrow{T_1} \mu'^{\ell} \xrightarrow{T_2} \mu^{\ell}$.

Our representation exists simultaneously at both the micro and tissue millimeter scales. A key aspect of anatomy is that at a micro or nano scale, information is encoded as a huge collection of pairs (x_i, f_i) where $x_i \in \mathbb{R}^d$ $(d = 2, 3)$ describes the position of a "particle" and f_i is a functional state in a given set $\mathcal{F}$ attached to it (protein or RNA signature or Tau tangle, or for single

cell Neurophysiology the dynamics of neural spiking). Basically everything is deterministic, with every particle attached to its own functional state among possible functional state in $\mathcal{F}$. But zooming out, the tissue level, say mm scale, appears through the statistical distribution of its constituents with two key quantities, *the local density* of particles ρ and the *conditional probability distribution* of the functional features $\mu_x(df)$ at any location x. At position x, we no longer have a deterministic functional state but a probability distribution μ_x on functional states.

The integration of both descriptions into a common mathematical framework can be done quite naturally in the setting of mathematical measures which are mathematical constructs that are able to represent both the discrete and continuous worlds as well as natural level of approximation between both. Indeed the set $\mathcal{M}(\mathbb{R}^d \times \mathcal{F})$ of finite positive measures on $\mathbb{R}^d \times \mathcal{F}$ contains discrete measures

$$\mu = \sum_{i \in I} w_i \delta_{x_i} \otimes \delta_{f_i} \ , \tag{1}$$

where w_i is a positive weight that can encode the collection (x_i, f_i) at micro scale.

As in Boltzmann modelling we describe the features statistically at a fixed spatial scale transferring our attention to their stochastic laws modelled as conditional probabilities in $\mathcal{M}_P(\mathcal{F})$ with integral 1. For this we factor the measures into the marginal ρ on $\mathbb{R}^d$ with $\rho(dx) = \int_{\mathcal{F}} \mu(dx, df)$, and the field of probability distributions on $\mathcal{F}$ conditioned on x :

$$\mu(dx, df) = \rho(dx)\, \mu_x(df) \ , \text{ on } \mathbb{R}^d \times \mathcal{F} \ , \tag{2a}$$

with field of conditionals

$$x \mapsto \mu_x \in \mathcal{M}_P(\mathcal{F}) \ . \tag{2b}$$

Continuous tissues we abstract as brain measures μ with marginal ρ having a tissue density $\rho(dx) = \rho_c(x)dx$ with respect to the Lebesgue measure on $\mathbb{R}^d$. A fundamental link between the molecular and continuum tissue can be addressed through the law of large numbers since if $(x_i, f_i)_{i \geqslant 0}$ is an independent and identically distributed sample drawn from law μ/M of $\mathbb{R}^d \times \mathcal{F}$ where $M = \int_{\mathbb{R}^d \times \mathcal{F}} \mu(dx, df)$ is the total mass of such μ, then we have almost surely the weak convergence

$$\mu_N := \frac{M}{N} \sum_{i=1}^{N} \delta_{x_i} \otimes \delta_{f_i} \to \mu \ .$$

Passing from the tissue scales to the finest molecular and cellular scales behooves us to introduce a scale-space so that empirical averages which govern it are repeatable. As depicted in the top center panel of Figure 1, our multi-scale model of a brain is as a sequence of measures:

$$\boldsymbol{\mu} = (\mu^\ell)_{\ell = 0, 1, \ldots} \ . \tag{3}$$

Our idealization of Brainspace as a sequence of measures as depicted in Figure 1 descends from the the coarse tissue scale (top) to the finest particle representation (bottom), with color representing function $f \in \mathcal{F}$, and radius space-scale.Throughout the range of scales is denoted shorthand $\ell < \ell_{max}$ to mean $0 \leqslant \ell < \ell_{max}$ with lowest scale $\ell = 0$ and upper ℓ_{max} not attained.

2.2 Model for Crossing Scales

The brain being a multi-scale collection of measures requires us to be able to transform from one scale to another. We do this by associating a scale-space to each particle feature by pairing to each measure a kernel function transforming it from a generalized

function to a classical function $\delta_z : h \mapsto h(z)$. The kernels carry a resolution scale σ or reciprocally a bandwidth, analogous to Planck's scale.

To define this we introduce the abstract representation of our system as a collection of descriptive elements $z \in \mathcal{Z}$ made from spatial and functional features. We transform our mathematical measure $\mu(\cdot)$ on $\mathcal{Z}$ generating new measures $\mu'(\cdot)$ on $\mathcal{Z}'$ by defining correspondences via kernels $z \mapsto k(z, dz')$, with the kernel acting on the particles $K[\delta_{z_i}](dz') = k(z_i, dz')$; the measures transform as

$$K[\mu](dz') \;=\; \int_{\mathcal{Z}} k(z, dz')\mu(dz)\;. \tag{4}$$

Depicted in the bottom row of Figure **??** is the construction of our formalism for transformations of scale, linear spatial smoothing followed by non-linear transformation of the feature laws. The first operator transforms via a linear kernel $k_1((x, f), \cdot)$ leaving the feature space unchanged $\mathcal{Z} = \mathbb{R}^d \times \mathcal{F}$. The second transforms nonlinearly to new features $\mathcal{F}^\ell$ via the kernel $k_2((x, \alpha), \cdot)$ transforming any feature probability $\alpha \in \mathcal{M}_P$; decomposing $\mu'^\ell(dx, df) = \rho(dx)\mu'^\ell_x(df)$ it smooths the conditionals μ'^ℓ_x:

$$T_1 :\; \mu \mapsto \mu'^\ell = \int_{\mathbb{R}^d \times \mathcal{F}} k_1((x, f), \cdot)\, \mu(dx, df)\,; \tag{5a}$$

$$T_2 : \mu'^\ell \mapsto \mu^\ell \;\; = \int_{\mathbb{R}^d} k_2((x, \mu'^\ell_x), \cdot)\, \rho(dx)\;. \tag{5b}$$

At every scale μ^ℓ remains a measure allowing the renormalization (5a), (5b) to be executed recursively.

We use computational lattices to interpolate $\mu = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$ between the multi-scale continuum. The core spatial resampling kernel projects the particles $(x_i \in \mathbb{R}^d)_{i\in I}$ to the rescaling lattice $(y_j \in Y_j \subset \mathbb{R}^d)_{j\in J}$ defined via π apportioning the fraction that particle x_i shares with site Y_j. The second kernel uses maps ϕ from machine learning transforming features:

$$k_1((x, f), \cdot) = \sum\nolimits_{j\in J} \pi(x, Y_j)\, \delta_{y_j} \otimes \delta_f(\cdot) \tag{6a}$$

$$k_2((x, \mu'^\ell_x), \cdot) = \delta_x \otimes \delta_{\phi(\mu'^\ell_x)}(\cdot) \tag{6b}$$

$$\text{with} \quad \begin{cases} \pi(x_i, \cdot) : Y_j \mapsto \pi(x_i, Y_j),\; \pi(x_i, \mathbb{R}^d) = 1 \\ \phi : \alpha \in \mathcal{M}_P(\mathcal{F}) \mapsto \phi(\alpha),\; \phi(\alpha) \in \mathcal{F}^\ell \end{cases} .$$

The renormalized measures become

$$\mu'^\ell = \sum_{j\in J} w_j \delta_{y_j} \otimes \mu'^\ell_{y_j} \tag{6c}$$

$$\mu^\ell = \sum\nolimits_{j\in J} w_j \delta_{y_j} \otimes \delta_{\phi(\mu'^\ell_{y_j})} \tag{6d}$$

$$\text{with} \quad \begin{cases} w_j = \sum_{i\in I} w_i \pi(x_i, Y_j) \\ \mu'^\ell_{y_j} = \frac{1}{w_j} \sum_{i\in I}\; w_i \pi(x_i, Y_j)\, \delta_{f_i} \\ \rho = \sum_{j\in J} w_j \delta_{y_j} \end{cases} .$$

The indexing across scales is defined disjointly $I \cap J = \varnothing$. Methods 4.2 examines these models.

2.3 The Dynamical Systems Model via Varifold Action of Multi-scale Diffeomorphisms

We want to measure and cluster brains by building a metric space structure. We do this by following the original program

of D'Arcy Thompson building bijective correspondence. In this setting this must be done at every scale with each scale having different numbers of particles and resolutions.

We build correspondence between sample brains via dense connections of the discrete particles to the continuum at all scales using the diffeomorphism group and diffeomorphic transport. For this define the group of k-times continuously differentiable diffeomorphisms $\varphi \in G_k$ with group operation function composition $\varphi \circ \varphi'$. For any brain $\mu = \sum_{i \in I} w_i \delta_{x_i} \otimes \delta_{f_i} \in \mathcal{M}$, the diffeomorphisms act

$$(\varphi, \mu) \mapsto \varphi \cdot \mu = \sum\nolimits_{i \in I} w_i |d\varphi(x_i)| \delta_{\varphi(x_i)} \otimes \delta_{f_i} \,. \tag{7a}$$

Space scales are represented as the group product, $\boldsymbol{\varphi} := (\varphi^\ell)_{\ell < \ell_{max}}$, acting component-wise with action

$$\boldsymbol{\varphi} \cdot \boldsymbol{\mu} := (\varphi^\ell \cdot \mu^\ell)_{\ell < \ell_{max}} \,. \tag{7b}$$

The $|d\varphi(x)|$ term in the action enables the crucial property that when a tissue is extended to a larger area, the total number of its basic constituents increase accordingly and are not conserved, in contrast to classic measure or probability transport. We call this the "copy and paste" varifold action.

Dynamics occurs through the group action generated as a dynamical system in which we introduce time-indexed vector fields $t \mapsto u_t^\ell$ which act as controls on the diffeomorphisms:

$$\dot{\varphi}_t^\ell = u_t^\ell \circ \varphi_t^\ell \,; \tag{8a}$$

the multi-scale flow is defined

$$\dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t \circ \boldsymbol{\varphi}_t := (u_t^\ell \circ \varphi_t^\ell)_{\ell < \ell_{max}}. \tag{8b}$$

Geodesic mapping flows under a control process along paths of minimum energy respecting the boundary conditions. The controls are coupled across scales by successive refinements $v^\ell, \ell < \ell_{max}$, with $u^{-1} = 0$:

$$u_t^\ell = u_t^{\ell-1} + v_t^\ell \,. \tag{9}$$

Figure 2 right shows the multi-scale control hierarchy.

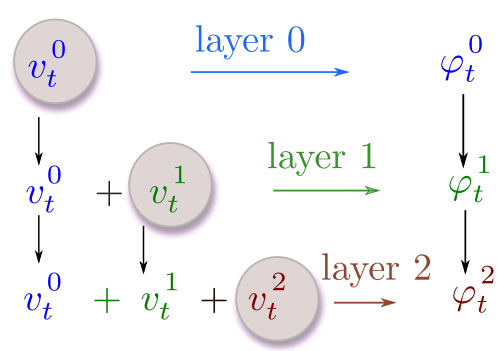

Figure 2: Hierarchical system, controls $u^\ell = u^{\ell-1} + v^\ell, u^0 = v^0$ and flows $\varphi^\ell, \ell = 0, 1, \ldots$.

The dynamical system is an observer and dynamics equation:

$$\begin{cases} \boldsymbol{\varphi}_t \cdot \boldsymbol{\mu},\ t \in [0,1] & \text{Observer} \\ \dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t \circ \boldsymbol{\varphi}_t,\ \boldsymbol{\varphi}_0 = \mathbf{Id} & \text{Dynamics} \end{cases} . \tag{10}$$

Dynamics translates into a navigation in the orbit of brains and provides a metric distance between brains. Paths of minimum energy connecting the identity $\boldsymbol{\varphi}_0 = \mathbf{Id}$ to any fixed boundary condition (BC) $\boldsymbol{\varphi}_1$ where $\boldsymbol{\varphi}_1$ is accessible defines the distance extending LDDMM [23] to a hierarchy of diffeomorphisms, and is a geodesic for an associated Riemannian metric [24] ; see Eqn. (22) in Methods 4.4. The metric from $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_1$ in the orbit accesible from $\boldsymbol{\mu}_0$ via diffeomorphisms is the shortest length geodesic paths with BCs $\boldsymbol{\varphi}_0 \cdot \boldsymbol{\mu}_0 = \boldsymbol{\mu}_0$ and $\boldsymbol{\varphi}_1 \cdot \boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$.

The Methods 4.4 discusses the spaces and the smoothness required for the geodesics to define a metric.

2.4 Model for Optimal Control for Brain Mapping via the Varifold Measure Norm

The BC for matching two brains is defined using measure norms. Brains with 0 norm difference are equal; brains with small normed difference are similar. Every brain has a variable number of particles, without correspondence between particles. Measure norms accomodate these variabilities. Our norm is modelled as a varifold norm constructed by integrating against a smooth kernel:

$$\|\mu\|^2_{W*} := \int_{\mathbb{R}^d \times \mathcal{F}} K_W[\mu](x,f)\mu(dx,df) \ ; \tag{11}$$

the multi-scale norm is given as $\|\boldsymbol{\mu}\|^2_{\boldsymbol{W}*} := \sum_{\ell<\ell_{max}} \|\mu^\ell\|^2_{W^*_\ell}$. In the Methods section 4.3 we define separable Gaussian kernels in space and function each with a scale.

Geodesic mapping solves for the control minimizing the energy with the boundary condition the measure norm. We model the controls to be smooth via reproducing kernel Hilbert spaces (RKHS's) V_ℓ, norms $\|\cdot\|_{V_\ell}$, with multi-scale $\boldsymbol{u} \in \boldsymbol{V} := \prod_{\ell<\ell_{max}} V_\ell$.

The optimal control $\boldsymbol{u}_\cdot := (\boldsymbol{u}_t, t \in [0,1])$ is square-integrable under the V-norms satisfying for $\alpha > 0$:

$$\begin{cases} \min\limits_{\boldsymbol{u}_\cdot \in L^2([0,1],\boldsymbol{V})} \dfrac{1}{2} \sum\limits_{\ell<\ell_{max}} \displaystyle\int_0^1 \|u^\ell_t - u^{\ell-1}_t\|^2_{V_\ell} dt + \dfrac{\alpha}{2}\|\boldsymbol{\varphi}_1 \cdot \boldsymbol{\mu} - \boldsymbol{\mu}_{obs}\|^2_{\boldsymbol{W}*} \\ \dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t \circ \boldsymbol{\varphi}_t,\ \boldsymbol{\varphi}_0 = \mathbf{Id} \end{cases} . \tag{12}$$

For modelling, each RKHS is taken to have diagonal kernels $K^\ell(\cdot,\cdot) = g^\ell(\cdot,\cdot)\mathrm{id}_d$, g^ℓ the Green's functions with id_d the $d \times d$ identity (see [24] for non-diagonal kernels). Easing notation we remove explicit indexing of the optimal controls by scales ℓ when implied. Optimal control reparameterizes the flows by introducing the state process indexing the measures and endpoint condition:

$$q_t := (x_{i,t} = \varphi_t(x_i),\ w_{i,t} = w_i|d\varphi_t|(x_i))_{i\in I} \tag{13a}$$

$$\mu(q_t) := \sum_{i\in I} w_{i,t}\delta_{x_{i,t}} \otimes \delta_{f_i} \ , \tag{13b}$$

$$\text{with } U(q_1) := \frac{\alpha}{2}\|\mu(q_1) - \mu_{obs}\|^2_{W*} = \frac{\alpha}{2}\|\varphi_1 \cdot \mu - \mu_{obs}\|^2_{W*} \ . \tag{13c}$$

Hamiltonian control reparameterizes (12) in "co-states" $t \to p_{i,t} = (p^x_{i,t}, p^w_{i,t})$ and Greens functions, for $\ell < \ell_{max}$,

$$u^\ell_t(\cdot) = \sum_{j<\ell_{max}} \sum_{i \in I_j} \left(\bar{g}^{\ell \wedge j}(x^j_{i,t}, \cdot) p^{x,j}_{i,t} + p^{w,j}_{i,t} w^j_{i,t} \nabla_1 \bar{g}^{\ell \wedge j}(x^j_{i,t}, \cdot) \right) \; ; \tag{14a}$$

then $\boldsymbol{q}_1 \mapsto \boldsymbol{U}(\boldsymbol{q}_1) = \sum_{\ell<\ell_{max}} U(q^\ell_1)$ is continuously differentiable implying each co-state is absolutely integrable:

$$\begin{cases} p^x_{i,t} = (d\varphi_{t,1})^T(x_{i,t}) p^x_{i,1} + p^w_{i,1} w_{i,1} \displaystyle\int_t^1 (d\varphi_{t,s})^T(x_{i,t}) \nabla(\mathrm{div} u_s)(x_{i,s}) ds \; , \quad p^x_{i,1} = -\nabla_{x_i} U(q_1) \\ p^w_{i,t} = p^w_{i,1} |d\varphi_{t,1}|(x_{i,t}) \; , \; p^w_{i,1} = -\nabla_{w_i} U(q_1) \end{cases} , \tag{14b}$$

for $\varphi_{t,s} = \varphi_s \circ (\varphi_t)^{-1}$, with $\bar{g}^\ell := \sum_{j=0}^\ell g^j$ and $\nabla_1 \bar{g}(x,y)$ the x-gradient. The gradients are given by the smooth field h_q:

$$h_{q_1} := \alpha K_W[\mu(q_1) - \mu_{obs}] \; , \text{ on } \mathbb{R}^d \times \mathcal{F} \; , \tag{15a}$$

$$\text{with } \begin{cases} \nabla_{x_i} U(q_1) = w_{i,1} \nabla_{x_i} h_{q_1}(x_{i,1}, f_i) \\ \nabla_{w_i} U(q_1) = h_{q_1}(x_{i,1}, f_i) \end{cases} . \tag{15b}$$

Methods 4.5 establishes the smoothness conditions for the Hamiltonian equations, methods 4.6 the gradients of the norms.

2.4.1 Gradients of the norm endpoints unifying the molecular and tissue models.

Imaging at the tissue continuum scales has the measures as dense limits $\mu = \int_{\mathbb{R}^d} \rho_c(x) \delta_x \otimes \mu_x \, dx$. Calculating the variations on dense voxel images unifies the tissue scales with the sparse molecular scales. The state $q_t := (\varphi_t, w_t = w_0 |d\varphi_t|)$ with action gives

$$\mu(q_t) := \varphi_t \cdot \mu = \int_{\mathbb{R}^d} w_t(x) \rho_c(x) \delta_{\varphi_t(x)} \otimes \mu_x dx \; . \tag{16}$$

The average of h_q over the feature space determines the boundary gradient:

$$\begin{cases} \nabla_\varphi U(\varphi_1, w_1) = w_1(x) \rho_c(x) \displaystyle\int_{\mathcal{F}} \nabla_\varphi h_{q_1}(\varphi_1(x), f) \mu_x(df) \\ \nabla_w U(\varphi_1, w_1) = \rho_c(x) \displaystyle\int_{\mathcal{F}} h_{q_1}(\varphi_1(x), f) \mu_x(df) \end{cases} \tag{17}$$

The continuum unifies with LDDMM [23]; taking $I(x) \in \mathbb{R}^+$ with $\mu_x = \delta_{I(x)}$, $\mu^I := \int_{\mathbb{R}^d} \delta_x \otimes \delta_{I(x)} dx$ gives

$$\varphi \cdot \mu^I = \int_{\mathbb{R}^d} \delta_x \otimes \delta_{I \circ \varphi^{-1}(x)} dx \; . \tag{18}$$

2.5 MRI and Tau Histology Scales for Alzheimer's

2.5.1 Bayes Segmentation of MRI

Figure 3 shows the multi-scale data from the clinical BIOCARD study [6] of Alzheimer's disease within the medial temporal lobe [7, 8, 25]. The top (left) panel shows clinical magnetic resonance imaging (MRI) with the high-field 200 μm MRI scale (right) shown depicting the medial temporal lobe including the collateral sulcus and lateral bank of the entorhinal cortex. Bayes classifiers

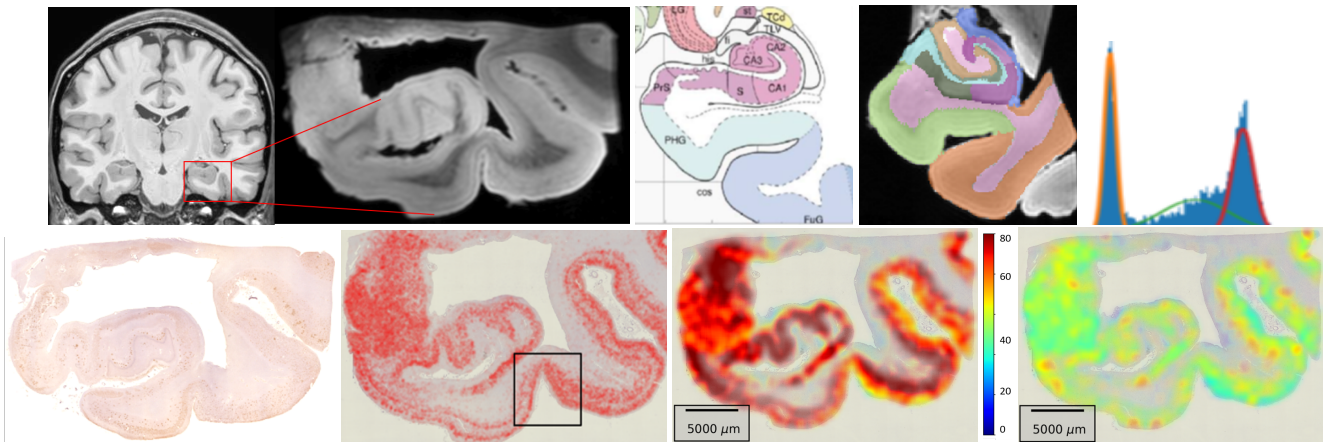

Figure 3: Top: Medial lobe at 1 mm and high-field 200 μm MRI; Mai-Paxino atlas section of the MRI with hippocampus and entorhinal cortex; right shows the Bayes compartments. Bottom: Alzheimer 4 μm Tau histology (red) from section depicted via high-field MRI Fig. 3; right shows detected Tau particles in simlar section. Box depicts trans-entorhinal region from top row. Right two panels shows mean particle size and standard deviation at μm and tissue scales; deep red color denotes 80 μm^2 Tau area.

for brain parcellation performs feature reduction as a key step for segmentation at tissue scales [26]. Feature reduction maps the distribution on gray levels $\mathcal{F} = [0, 255]$ to probabilities on N tissue types, defined by the integration over the decision regions $\theta_n \subset [0, 255]$:

$$\phi_n(\mu_x) = \int_{\theta_n} \mu_x(df) = p_n, \; n = 1, \ldots, N \; .$$

Figure 3 (top row, right) depicts a Bayes classifier for gray, white and cerebrospinal fluid compartments generated from the temporal lobe high-field MRI section corresponding to the Mai-Paxinos section (panel 3).

2.5.2 Gaussian Scale-Space Resampling of Tau Histology:

For histology at the molecular scales the measure encodes the detected Tau and amyloid particles $\mu = \sum_i w_i \delta_{x_i} \otimes \delta_{f_i}$ for fine scale particles with function the geometric features $\mathcal{F} = \mathbb{R}^+$. Crossing from the histology accumulates the $(x_j \in \mathbb{R}^d)_{i \in I}$ Tau to the tissue lattice $(y_j \in Y_j \subset \mathbb{R}^d)_{j \in J}$ using Gaussian scale-space with normal reweighting in $\mathbb{R}^2$. Figure 3 (bottom row) shows the detected Tau particles as red dots at 4μ m with feature reduction to $\mathcal{F}^\ell$ done via moments expansion on tau distribution:

$$\begin{aligned}
&\pi_\sigma(x_i, Y_j) = 1/(2\pi\sigma^2) \int_{Y_j} e^{-|y - x_i|^2/2\sigma^2} dy \; . \\
&\mu'^\ell = \sum_{j \in J} w_j \delta_{y_j} \otimes \mu'^\ell_{y_j} \\
&\mu^\ell = \textstyle\sum_{j \in J} w_j \delta_{y_j} \otimes \delta_{\phi(\mu'^\ell_{y_j})} \\
&\quad \text{with} \begin{cases} w_j = \sum_{i \in I} w_i \pi_\sigma(x_i, Y_j) \\ \mu'^\ell_{y_j} = \frac{1}{w_j} \sum_{i \in I} w_i \pi_\sigma(x_i, Y_j) \, \delta_{f_i} \\ \phi(\mu'^\ell_{y_j}) = \left(\int_{\mathcal{F}} f \mu'^\ell_{y_j}(df), \int_{\mathcal{F}} (f)^2 \mu'^\ell_{y_j}(df) \right) \end{cases} .
\end{aligned} \tag{19}$$

The bottom row (right two panels) shows the two moments at the tissue scale depicting the mean and variance of the particle size reconstructed from the 4μm scale using Gaussian resampling onto the tissue lattice. The mm scale depicts the global folding property of the tissue. The color codes the mean tissue Tau area as a function of position at the tissue scales with deep red color denoting $80\,\mu\text{m}^2$ maximum Tau area for the detected particles.

2.6 Cellular Neurophysiology: Neural Network Temporal Models

Single unit neurophysiology uses temporal models of spiking neurons with a "neural network" $\mu = \sum_i \delta_{x_i} \otimes \delta_{f_i}$, $x_i \in \mathbb{R}^d$, $f_i \in \mathcal{F}$ taking each neuron x_i modelled as a counting measure in time $N_i(t), t \geqslant 0$ with the spike times the feature $f_i = (t_k)_{1 \leqslant k \leqslant n_{f_i}}$:

$$\mathcal{F} = \{f = (t_k)_{1 \leqslant k \leqslant n_f} : n_f \geqslant 1, t_1 < t_2 \cdots < t_{n_f}\} .$$

The Poisson model with intensity $\lambda(t), t \geqslant 0$ [10] has probabilities $\Pr(N_i(t) = n) = \frac{(\int_{t_0}^t \lambda(s)ds)^n}{n!} e^{-\int_{t_0}^t \lambda(s)ds}$.

Post-stimulus time (PST) [27] and interval histograms are used to examine the instantaneous discharge rates and inter-spike interval statistics [28]. The interval histogram abandons the requirement of maintaining the absolute phase of the signal for measuring temporal periodicity and phase locking. Synchrony in the PST is measured using binning $[b_i, b_{i+1})$, $i = 1, \cdots, B$ and Fourier transforms, j=$\sqrt{-1}$:

$$\phi_n^{PST}(\mu_x) = \sum_i e^{\mathrm{j}\,\omega_n i} \int_{\mathcal{F}} \frac{1}{n_f} \sum_{k=1}^{n_f} \mathbf{1}_{[b_i, b_{i+1})}(t_k) \mu_x(df) \; , n = 0, 1 \ldots \; .$$

The $n = 0$ frequency computes integrated rate; each phase-locked feature is complex $\phi_n \in \mathbb{C}$.

2.7 Scale Space Resampling of RNA to Cell and Tissue Scales

Methods in spatial-transcriptomics which have emerged for localizing and identifying cell-types via marker genes and across different cellular resolutions [4, 29–32] presents the opportunity of localizing in spatial coordinates the transcriptionally distinct cell-types. Depicted in Figure 4 are the molecular measurements at the micron scales with MERFISH [33].

The molecular measures represent RNA locations with sparse RNA features, $\mu^{\ell+1} = \sum_{i \in I} \delta_{x_i} \otimes \delta_{f_i}$, $\mathcal{F}^{\ell+1} = \mathbb{R}^{167}$. Crossing to cells $(Y_j \subset \mathbb{R}^2)_{j \in J}$ partitions into closest particle subsets defined by the distance $d(x_i, Y_j)$ of particle x_i to cell Y_j. The RNA particles $(x_i \in \mathbb{R}^d)_{i \in I}$ are resampled to the cell centers $(\delta_{y_j}, y_j \in Y_j \subset \mathbb{R}^2)_{j \in J}$ via indicator functions accumulating to nonsparse mixtures of RNA within the closest cell. The new reduced feature vector becomes the conditional probability on 17 cell-types in $\mathcal{F}^\ell \subset [0,1]^{17}$. The conditional probabilities on the RNA vectors are calculated via principle components $E_n, n = 1, 2, \ldots$,

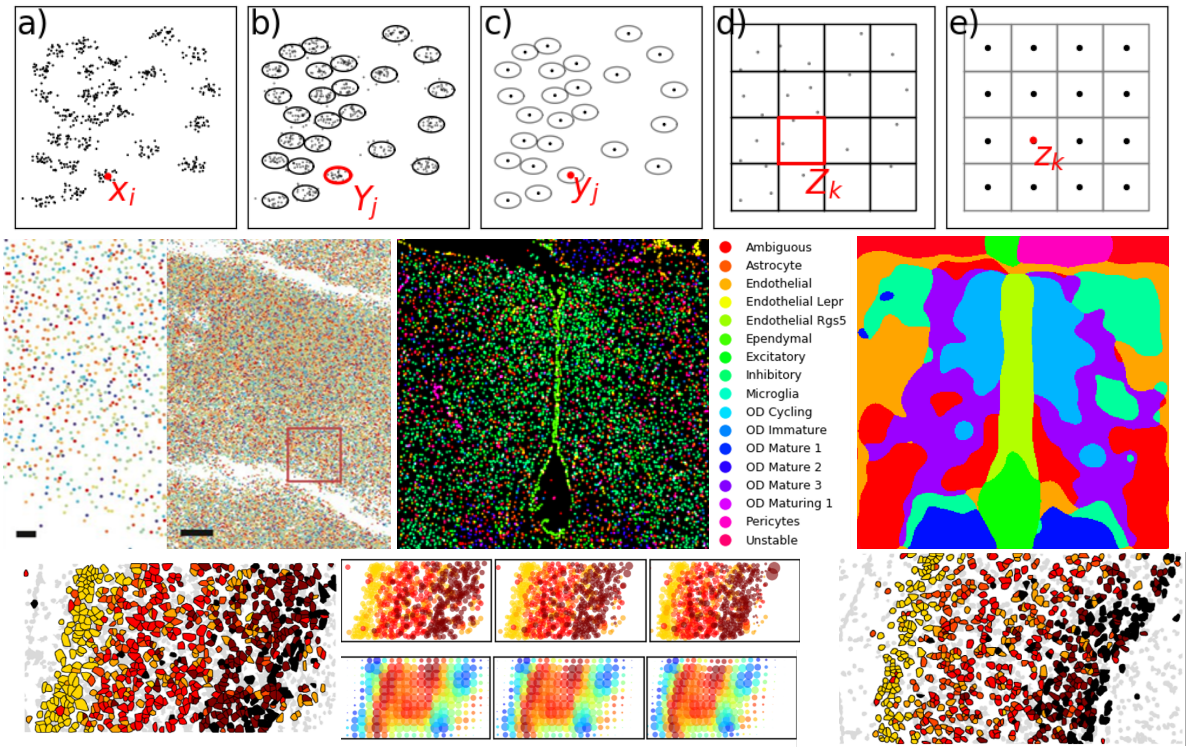

Figure 4: Row 1 shows cartoon of multi-scale renormalization of particles to cells and cell-centers to regular lattice representing the tissue. Row 2 shows RNA markers[33] (left) of the 167 RNA gene species (bar scales $1, 10\,\mu$m); middle 17 cell types clustered on cell centers; right shows $K = 10$ means clustering to tissue with Gaussian resampling with $\sigma = 25$pix on the $450 \times 200\,\text{pix}^2$ grid. Bottom row shows neuronal cell types [4] for two sections: far left a template, and far right a target. Intermediate panels show shows $\varphi_t \cdot \mu$ for $t = 0, 0.5, 1.0$ of neuronal cells mapping cell types (top) and entropy (bottom). The varifold norm $\sigma_x^\ell = \sigma_f^\ell = 10\,\text{pix}, 50\,\text{pix}$ with the feature a one hot encoding of cell type. The vector fields have RKHS norm induced by the differential operator $\|L^\ell v^\ell\|^2$ with $L^\ell := ((1-(\alpha^\ell)^2\nabla^2)\text{id}_d)^2$, $\alpha^\ell = 5\,\text{pix}, 50\,\text{pix}$ with full width at half maximum (FWHM) of the Green's kernel $FWHM = 20\,\text{pix}, 145\,\text{pix}$, for $\ell = 1, 2$, respectively.

modelling $\varepsilon_{y_j}^\ell(n) = \int E_n(f)\mu_{y_j}'^\ell(df)$ features as independent, Gaussian distributed:

$$\pi(x_i, Y_j) = \begin{cases} 1 & \text{for } d(x_i, Y_j) < d(x_i, Y_{j' \neq j}) \\ 0 & \text{otherwise} \end{cases} ,$$

$$\mu^\ell = \sum\nolimits_{j \in J} w_j\, \delta_{y_j} \otimes \delta_{f_j^\ell}$$

$$\text{with} \begin{cases} w_j = \sum_{i \in I} \pi(x_i, Y_j) \\ \mu_{y_j}'^\ell = \frac{1}{w_j} \sum_{i \in I} \pi(x_i, Y_j)\, \delta_{f_i} \\ f_j^\ell(c) = \Pr\{C_{y_j} = c | \varepsilon_{y_j}^\ell(1), \varepsilon_{y_j}^\ell(2), \dots\}\,,\ c = 1, \dots, 17\,. \end{cases} ,$$

Resampling to tissue $(\delta_{z_k}, z_k \in Z_k \subset \mathbb{R}^2)_{k \in K}$ uses normal rescaling. The new feature vector becomes the probability of the cell at any position being one of 10 tissue types $\mathcal{F}^{\ell-1} \subset [0,1]^{10}$. The probability of tissue type is calculated using 10-means clustering on the cell type probabilities. The distance for 10-means clustering is computed using the Fisher-Rao metric [34] between the

empirical feature laws $\mu'^{\ell-1}_{z_k}$. The output of 10-means are a partition of feature space $\cup_t \mathscr{F}_t = \mathcal{F}^\ell$ giving new features:

$$\pi_\sigma(y_j, Z_k) = \frac{1}{2\pi\sigma^2}\int_{Z_k} e^{-|z-y_j|^2/2\sigma^2} dz \; .$$

$$\mu^{\ell-1} = \sum\nolimits_{k\in K} w_k \delta_{z_k} \otimes \delta_{f_k^{\ell-1}}$$

$$\text{with} \begin{cases} w_k = \sum_{j\in J} w_j \, \pi_\sigma(y_j, Z_k), \\ \mu'^{\ell-1}_{z_k} = \frac{1}{w_k}\sum_{j\in J} w_j \, \pi_\sigma(y_j, Z_k)\, \delta_{f_j^\ell} \\ f_k^{\ell-1}(t) = \int_{\mathscr{F}_t} \mu'^{\ell-1}_{z_k}(df) \; , \; t = 1, \ldots, 10 \; . \end{cases} .$$

Figure 4 (bottom left panels) shows the RNA forming $\mu^{\ell+1}$ depicted as colored markers corresponding to the different gene species (bar scale $1, 10$ microns). Panel 2 shows the cell type feature space of 17 cell types making up μ^ℓ associated to the maximal probability in the PCA projection from a classifier on the PCA dimensions based on the mixtures of RNA at each cell location. Bottom right shows the 10 tissue features associated to the 10-means procedure. In both scales the probabilities are concentrated on one class with probability 1 and the others 0.

2.8 Geodesic Mapping for Spatial Transcriptomics, Histology and MRI

2.8.1 Navigation Between Sections of Cell Identity in Spatial Transcriptomics.

Figure 4 (top panels 1,2) shows sections from [4] depicting neuronal cell types via colors including excitatory cells eL2/3 (yellow), eL4 (orange), red eL5 (red), inhibitory cells ST (green), VIP (light blue), each classified via high dimensional gene expression features vectors via spatial transcriptomics.

The measure $\mu = \sum_i w_i \delta_{x_i} \otimes \delta_{f_i}$, $\mathcal{F} = \{\text{cell types}\}$ crosses to atlas scales using π_σ in $\mathbb{R}^2$ as above Eqn. (19) with feature reduction expectations of moments, $\mathcal{F}^\ell = \{\text{size, variance,entropy}\}$:

$$\mu^\ell = \sum\nolimits_{j\in J} w_j \delta_{y_j} \otimes \delta_{\phi(\mu'^\ell_{y_j})}$$

$$\text{with } \phi(\mu'^\ell_{y_j}) = \left(\sum_{\mathcal{F}} f \mu'^\ell_{y_j}(f), \sum_{\mathcal{F}} (f)^2 \mu'^\ell_{y_j}(f), -\sum_{\mathcal{F}} \mu'^\ell_{y_j}(f) \log \mu'^\ell_{y_j}(f) \right) \; .$$

The right panel shows the tissue scale features associated to the cell identity given by the entropy a measure of dispersion across the cell identities given by the expectation of the log probability function with zero entropy meaning the space location feature distribution μ'^ℓ_x has all its mass on 1 cell type.

Figure 4 shows results of transforming the neuronal cell types (row 2) and the entropy (row 3) at the scales of the cell and tissue. For all of our geodesic mapping examples shown we enforce vector field smoothness via differential operators specifying the norms in the RKHS $\|v^\ell\|^2_{V_\ell} := \|L^\ell v^\ell\|^2$ with $L^\ell := ((1-(\alpha^\ell)^2\nabla^2)\mathrm{id}_d)^2$ for $\ell < \ell_{max}$.

2.8.2 Navigation Between Sections of Histology.

Figure 5 (rows 1 and 2) shows navigation between the cortical folds of the $4\,\mu$m histology. Shown in panel 1 is a section showing the machine learning detection of the Tau particles. Columns 2,3, and 4 depict the template, mapped template and target showing the mathematical measure representation of the perirhinal cortex constructed from the positions and sizes at the 4μm scale (top row) and reconstruction using Gaussian resampling onto the tissue scale (bottom row). The color codes the mean of μ_x representing Tau area as a function of position at the tissue scales with deep red of $80\,\mu\mathrm{m}^2$ of Tau area the maximum value. The gradients in tau

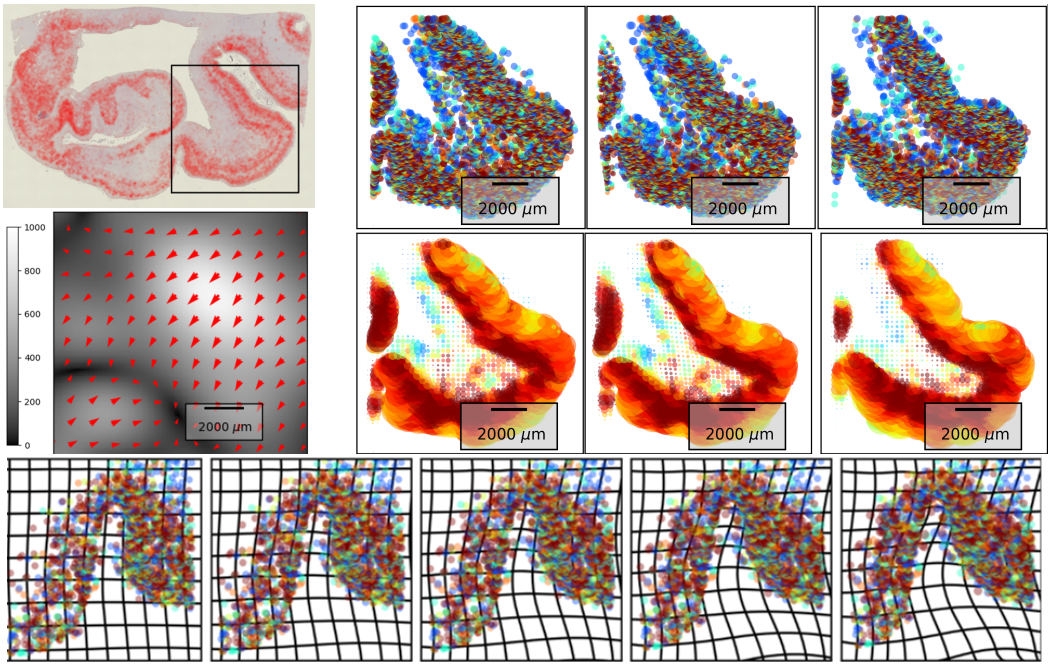

Figure 5: Rows 1,2: Second section 4μ m histology (similar to Figure 3) with box depicting perirhinal cortex. Columns 2, 3 and 4 show template, mapped template and target at the molecular (top) and tissue scales (bottom) showing the first moments of Tau size on the perirhinal cortex; saturated red color denotes $80\,\mu\text{m}^2$ Tau area. Mapped template shows narrowing $1000\,\mu$m of perirhinal sulcus. Second row left panel depicts vector field encoding of the geodesic mapping with associated scale bar. Row 3: Geodesic navigation $\varphi_t \cdot \mu$ of collateral sulcus (Figure 3, box) showing $1000\ \mu$m widening of folds for molecular and tissue scales depicting the mean transforming. The vector field mappings have RKHS norm induced by the differential operator $\|L^\ell v^\ell\|^2$ with $L^\ell := ((1-(\alpha^\ell)^2\nabla^2)\text{id}_d)^2$, with α^ℓ with $FWHM = 580\mu\text{m}, 3300\mu\text{m}$, for $\ell = 1, 2$ respectively. The varifold norm has $\sigma_x^\ell = 800\,\mu\text{m}, 160\,\mu\text{m}$ for the mappings.

tangle area between superficial and deep layers is apparent with the deep red as high as $80\,\mu\text{m}^2$ for the innermost tissue fold. The bottom left panel shows the vector field encoding of the geodesic transformation of the of the perirhinal cortex mapping between the two sections in column 1. The narrowing of the banks of the perirhinal cortex is exhibited at the tissue scale for motions order $1000\,\mu$m (brightness on scale bar).

Figure 5 (row 3), shows the collateral sulcus fold at the boundary of the trans-entorhinal cortex region transforming based on the normed distances between sections with deformation motions $1000\,\mu$m in size. Shown is the micron scale depicting the transformation of the gyrus with the color representing the entropy of the particle identity distribution.

2.8.3 Mapping between MRI and Histology Simultaneously

All of the examples thus far have created the multi-scale data generated using the resampling kernels from the finest scales. As illustrated in our early figures much of the data is inherently multi-scale, with the measurement technologies generating the coarse scale representations. Shown in Figure 6 is data illustrating our Alzheimer's study of post mortem MR images that are simultaneously collected with amyloid and Tau pathology sections. MR images have a resolution of approximately 100 μm, while pathology images have a resolution of approximately 1 μm. For computational purposes the MRI template and target images were downsampled to 759 and 693 particles, respectively with the tau tangles downsampled to 1038 and 1028 particles, respectively. We treated every pixel in the MR image as a coarse scale particle with image intensity as its feature value Eqn. (18), and every

detected tau tangle as a fine scale particle with a constant feature value, and performed varifold matching to align to neighboring sections. The endpoint representing the two scales is $\boldsymbol{U} = \frac{1}{2}\sum_{\ell=1,2}\|\mu^{\ell}_{temp} - \mu^{\ell}_{obs}\|^2_{W^*_{\ell}}$. For each scale norm we use a varifold kernel given by the products of Gaussian distributions with the varifold measure norm Eqn. (21) at each scale. For the MRI scale, the weights are identically $w = 1$ with the function component given by the MRI image value; for the tau particles there is no function component making $f = g$ with the kernel 1 for all values.

The top two rows of Figure 6 shows the imaging data for both sections. The bottom row shows the transformed template image at the fine scale. The high resolution mapping carries the kernels across all the scales as indicated by the geodesic equation (14a). Notice the global motions of the high resolution of the fine particles.

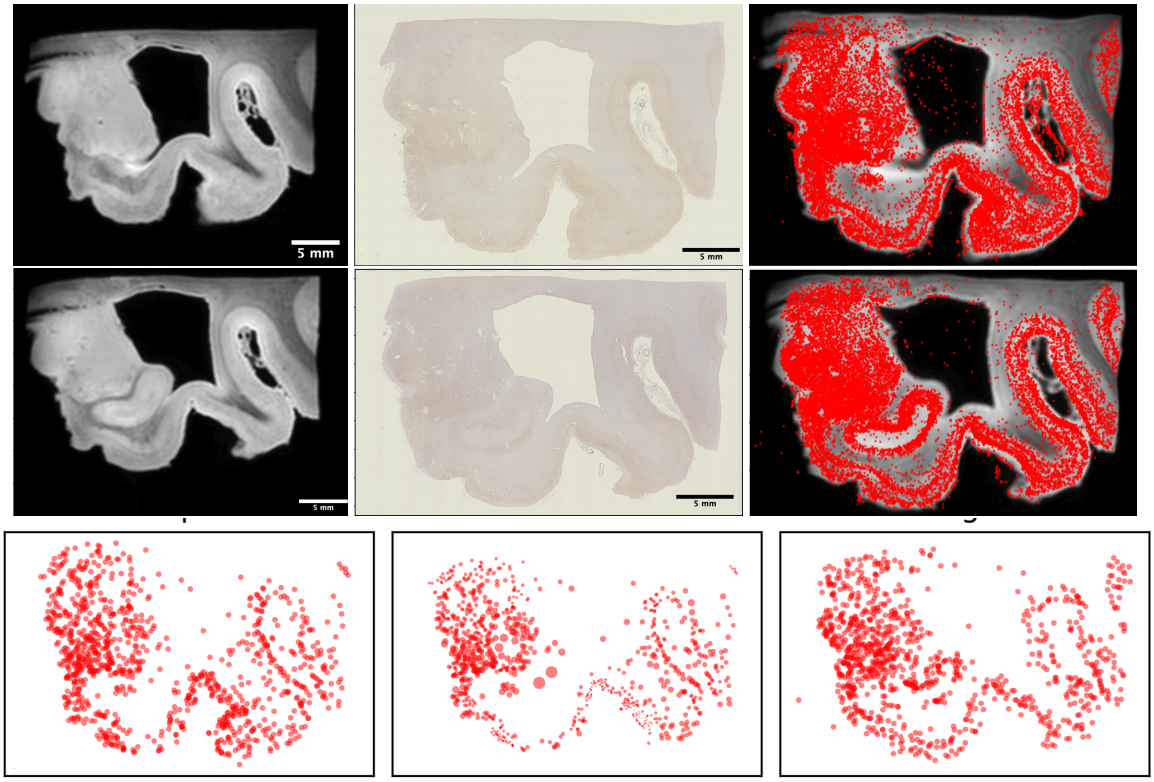

Figure 6: Whole brain section showing mapping MRI and histology at the multiple scales. Top two row shows the MRI and Tau histology for two sections with the detected Tau particle superimposed over the MRI (right); bottom row shows the finest scales for the template, middle the template mapped, with the right shows the target; the varifold norm has $\sigma^{\ell}_x = 4000\,\mu\text{m}, 100\,\mu\text{m}$. The vector field mappings have RKHS norm induced by the differential operator $\|L^{\ell}v^{\ell}\|^2$ with $L^{\ell} := ((1-(\alpha^{\ell})^2\nabla^2)\text{id}_d)^2$, α^{ℓ} giving $FWHM = 1500\,\mu\text{m}, 6400\,\mu\text{m}$, for $\ell = 1, 2$ respectively.

3 Discussion

Computational anatomy was originally formulated as a mathematical orbit model for representing medical images at the tissue scales. The model generalizes linear algebra to the group action on images by the diffeomorphism group, a non-linear algebra, but one that inherits a metric structure from the group of diffeomorphisms. The formulation relies on principles of continuity of medical images as classical functions, generalizating optical flow and advection of material to diffeomorphic flow of material, the

material represented by the contrast seen in the medical imaging modality such as fiber orientation for diffusion tensor imaging, and or bold contrast for gray matter content. Unifying this representation to images built at the particle and molecular biological scale has required us to move away from classical functions, to the more modern 20th century theory of non-classical generalized functions. Mathematical measures are the proper representation as they generally reflect the property that probes from molecular biology associated to disjoints sets are additive, the basic starting point of measure theory. Changing the model from a focus on groups acting on functions to groups acting on measures allows for a unified representation that has both a metric structure at the finest scales, as well as a unification with the tissue imaging scales.

The brain measure formulation, carries with it implicitly the notion of scale-space, i.e. the existence of a sequence of pairs across scales, the measure representation of the brain and the associated scale-space reproducing kernel Hilbert space of functions which correspond to the probing measurement technologies. As such part of the prescription of the theory is a method for crossing scales and carrying information from one scale to the other. Important to this approach is that at every scale we generate a new measure, therefore the recipe of introducing "measure norms" built from RKHS's for measuring brain disparity is universal across the hierarchy allowing us to work simultaneously with common data structures and a common formalism. Interestingly, the measure norms do not require identical particle numbers across brains in brain space at the molecular scales.

The key modelling element of brain function is that the conditional feature probability is manipulated from the quantized features to the stochastic laws. These are the analogues of the Boltzman distributions generalized to the complex feature spaces representing function. As they correspond to arbitary feature spaces not necessarily Newtonian particles, we represent them simply as empirical distributions on the feature space, with the empirical measure constructed from the collapse of the fine scale to the resampled coarse scale. To model rescaling through scale-space explicitly, the two kernel transformation are used allowing us to retrieve the empirical averages represented by the determinism of the stochastic law consistent with our views of the macro tissue scales. This solves the dilemna that for the quantized atomic and micro scales cell occurence will never repeat, i.e. there is zero probability of finding a particular cell at a particular location, and conditioned on finding it once it will never be found again in the exact same location in another preparation. The properties that are stable are the probability laws with associated statistics that may transfer across organisms and species.

Importantly, our introduction of the $|d\varphi(x)|$ term in the action enables the crucial property that when a tissue is extended to a larger area, the total number of its basic constituents should increase accordingly and not be conserved. This is not traditional measure transport which is mass preserving which is not a desirable feature for biological samples. Rather we have defined a new action on measures that is reminiscent of the action on d-dimensonal varifolds [35, 36]. We call this property "copy and paste", the notion being that the brain is built on basic structuring elements that are conserved.

Successive refinement for the small deformation setting has been introduced in many areas associated to multigrid and basis expansions. The notion of building multi-scale representation in the large deformation LDDMM setting was originally explored F. Riesser et al. [37] in which the kernels are represented as a sum of kernels and Sommer et al. [38] in which the kernel is represented as vector bundles. In their multi-scale setting there is a post-optimization decomposition in which the contribution of the velocity field into its different components can then each be integrated. In that multi-scale setting the basic Euler-Lagrange equation termed EPDIFF remains that of LDDMM [39]. In the setting proposed here we separate the scales before optimisation via the hierarchy of layered diffeomorphisms and use a multi-scale representation of the brain hierarchy itself which is directly associated to the differomorphism at that scale. This gives the fundamental setting of the product group of diffeomorphisms with the Euler-Lagrange equation corresponding to the sequence of layered diffeomorphisms [24].

The aggregation across scales from particle to tissue scales on lattices provides the essential link to inference on graphs. It is natural for these aggregated features with associated conditional probability laws to become the nodes in Markov random field modelling for spatial inference; see examples in spatial transcriptomics and tissue segmentation [40]. Building neighborhood relations as conditional probabilities between lattice sites from which global probabilites laws are constructed with the Hammersley-

Clifford theorem links us to Grenander's metric pattern theory formalisms with the atoms and conditional laws $(x_i, \mu_{x_i})_{i\in I}$ at any scale playing the roles of the generators.

4 Materials and Methods

4.1 Experimental and Technical Design

The objective of this research is to unify the molecular representations of spatial transcriptomics and cellular scale histology with the tissue scales of Computational Anatomy for brain mapping. To accomplish this we designed a mathematical framework for representing data at multiple scales using generalized functions, and mapping data using geodesic flows of multiple diffeomorphisms. We illustrate the method using several examples from human MRI and digital pathology, as well as mouse spatial transcriptomics.

4.2 Method of Resampling Across Lattice Scales

Taking $\mu = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$ on $\mathbb{R}^d \times \mathcal{F}$ with kernels k_1, k_2 of (6a), (6b) gives our two-step transformation. The transformation (5a) of $T_1 : \mu \mapsto \ \mu'^{\ell}$ gives with $w_j = \sum_{i\in I} w_i \pi(x_i, Y_j)$:

$$\begin{aligned} \mu'^{\ell} &= \int_{\mathbb{R}^d\times\mathcal{F}} \sum\nolimits_{j\in J} \pi(x, Y_j)\, \delta_{y_j} \otimes \delta_f \sum\nolimits_{i\in I} w_i\, \delta_{x_i} \otimes \delta_{f_i}\, (dx, df) \\ &= \sum\nolimits_{j\in J} w_j \delta_{y_j} \otimes \frac{1}{w_j} \left(\sum\nolimits_{i\in I} w_i \pi(x_i, Y_j) \delta_{f_i} \right) = \sum\nolimits_{j\in J} w_j \delta_{y_j} \otimes \mu'^{\ell}_{y_j} \ , \end{aligned} \tag{20}$$

with $\mu'^{\ell}_{y_j} = \sum_{i\in I} w_i \pi(x_i, Y_j) \delta_{f_i} / w_j$ which is (6c).

The space density $\rho = \int_{\mathcal{F}} \sum_{j\in J} w_j \delta_{y_j} \otimes \mu'^{\ell}_{y_j}(df) = \sum_{j\in J} w_j \delta_{y_j}$. The feature transformation (5b) of $T_2 : \mu'^{\ell} \ \mapsto \ \mu^{\ell}$ gives (6d):

$$\mu^{\ell} \ = \ \int_{\mathbb{R}^d} \delta_x \otimes \delta_{\phi(\mu'^{\ell}_x)} \left(\sum\nolimits_{j\in J} w_j \delta_{y_j}(dx) \right) \ = \sum\nolimits_{j\in J} w_j \delta_{y_j} \otimes \delta_{\phi(\mu'^{\ell}_{y_j})} \ .$$

4.3 Gaussian Kernel Varifold Norm

Our varifold norm construction models the measures as elements of a Hilbert space W^* which is dual to an RKHS W with a kernel K_W. We introduce the dual bracket notation for $h \in W, \mu \in W^*$, $(\mu \mid h) := \int_{\mathbb{R}^d\times\mathcal{F}} \mu(dx, df) h(x, f)$. The norms are generated by integrating against the kernel according to (11) written with the dual bracket $\|\mu\|^2_{W^*} := (\mu \mid K_W[\mu])$; the multi-scale norm is given by $\|\boldsymbol{\mu}\|^2_{\boldsymbol{W}^*} := \sum_{\ell<\ell_{max}} \|\mu^{\ell}\|^2_{W^*_{\ell}}$.

To ensure the brain measures are elements of W^* dual to the RKHS W, the kernel K_W is chosen to densely and continuously embed in bounded continuous functions $\mathscr{C}_b(\mathbb{R}^d \times \mathcal{F}, \mathbb{R})$ so that the signed measure spaces $\mathcal{M}_s(\mathbb{R}^d \times \mathcal{F})$ of brains are continuously embedded in the dual spaces $\mathscr{C}_b(\mathbb{R}^d \times \mathcal{F}, \mathbb{R})^*$. The Gaussian kernel (21) satisfies this condition, the kernel taken as separable Gaussians with $|\cdot|$ Euclidean distance:

$$k_W((x, f), (y, g)) := \exp\left(-\frac{1}{2\sigma_x^2} |x - y|^2 \right) \exp\left(-\frac{1}{2\sigma_f^2} |f - g|^2 \right) \ . \tag{21}$$

For measures μ_0, μ_1 with particle indexing i and j non-overlapping, gives

$$\begin{aligned}\|\mu_0-\mu_1\|^2_{W*} &= \sum_{i,i'} w_i w_{i'} \exp\left(-\frac{1}{2\sigma_x^2}|x_i-x_{i'}|^2\right)\exp\left(-\frac{1}{2\sigma_f^2}|f_i-f_{i'}|^2\right)\\ &-2\sum_{i,j} w_i w_j \exp\left(-\frac{1}{2\sigma_x^2}|x_i-x_j|^2\right)\exp\left(-\frac{1}{2\sigma_f^2}|f_i-f_j|^2\right)\\ &+\sum_{j,j'} w_j w_{j'} \exp\left(-\frac{1}{2\sigma_x^2}|x_j-x_{j'}|^2\right)\exp\left(-\frac{1}{2\sigma_f^2}|f_j-f_{j'}|^2\right) .\end{aligned}$$

For data carrying position information but no feature values (such as tau tangle locations), each f_i, f_j is constant and the resulting exponential terms are all 1.

4.4 Geodesics on the Group and the Riemannian Metric

The diffeomorphism group acts on the hierarchy $\boldsymbol{\varphi}\cdot\boldsymbol{\mu}$ component-wise Eqn. (7b) with the multi-scale group the product

$$\boldsymbol{G}_k = \underbrace{G_k \times \cdots \times G_k}_{\ell_{max}\text{ times}} ,$$

with elements $\boldsymbol{\varphi}\in\boldsymbol{G}_k$ satisfying the law of composition component-wise $\boldsymbol{\varphi}\circ\boldsymbol{\varphi}' = (\varphi^\ell\circ\varphi'^\ell)_{\ell<\ell_{max}}$. The group G_k supporting $k-$derivatives of the diffeomorphisms builds from $\mathscr{C}_0^k(\mathbb{R}^d,\mathbb{R}^d)$ a space of k-times continuously differentiable vector fields vanishing at infinity and its partial derivatives of order $p\leqslant k$ intersecting with diffeomorphisms with 1-derivative:

$$G_k = (\mathrm{id}+\mathscr{C}_0^k(\mathbb{R}^d,\mathbb{R}^d))\cap \mathrm{Diff}^1(\mathbb{R}^d,\mathbb{R}^d) .$$

Dynamics occurs via group action generated as a dynamical system in which the multi-scale control $t\mapsto \boldsymbol{u}_t := (u_t^\ell)_{\ell<\ell_{max}}$ flows the hierarchy $t\mapsto\boldsymbol{\varphi}_t$ satisfying $\dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t\circ\boldsymbol{\varphi}_t$ of (8a). The control is in the product $\boldsymbol{V}=\prod_{\ell<\ell_{max}} V_\ell$, each space an RKHS with norm-square $\|\cdot\|^2_{\boldsymbol{V}} = \sum_{\ell<\ell_{max}}\|\cdot\|^2_{V_\ell}$ selected to control the smoothness of the vector fields. The hierarchy of spaces are organized as a sequence of continuous embeddings:

$$V_0\hookrightarrow\cdots\hookrightarrow V_{\ell_{max}} ,$$

where $V_{\ell_{max}}$ is an additional layer containing the others with $V_{\ell_{max}} = \mathscr{C}_0^m(\mathbb{R}^d,\mathbb{R}^d)$ defined as a space of m-times continuously differentiable vector fields vanishing at infinity as well all its partial derivatives of order $p\leqslant m$.

The hierarchy is connected via successive refinements $u^\ell = u^{\ell-1}+v^\ell, u^0=v^0$ expressed via the continuous linear operator $\boldsymbol{A}:\boldsymbol{V}\to\boldsymbol{V}$ with $\boldsymbol{v}=\boldsymbol{A}\boldsymbol{u}$. The control process $\boldsymbol{u}_\cdot = (\boldsymbol{u}_t, t\in[0,1])\in L^2([0,1],\boldsymbol{V})$ has finite square-integral with total energy

$$\mathcal{E}_{\boldsymbol{A}}(\boldsymbol{u}_\cdot) = \frac{1}{2}\int_0^1 \|\boldsymbol{A}\boldsymbol{u}_t\|^2_{\boldsymbol{V}}dt .$$

Optimal curves which minimize the integrated energy $\mathcal{E}_{\boldsymbol{A}}$ between any two fixed boundary conditions (BC) $\boldsymbol{\varphi}_0=\mathbf{Id}$ and $\boldsymbol{\varphi}_1$ which is accessible with a path of finite energy extends the LDDMM setting [23] to a hierarchy of diffeomorphisms and describes a geodesic for an associated Riemannian metric [24] on $\boldsymbol{G}_k$:

$$d_{\boldsymbol{G}_k}(\mathbf{Id},\boldsymbol{\varphi}_1)^2 := \min_{\substack{\boldsymbol{u}.\in L^2([0,1],\boldsymbol{V}):\\ \dot{\boldsymbol{\varphi}}_t=\boldsymbol{u}_t\circ\boldsymbol{\varphi}_t,\text{with BC } \boldsymbol{\varphi}_0=\mathbf{Id},\boldsymbol{\varphi}_1}} \int_0^1 \|\boldsymbol{A}\boldsymbol{u}_t\|_{\boldsymbol{V}}^2 dt \ . \tag{22}$$

Existence of solutions for minimizers over $\boldsymbol{u}$ of (22) when $d_{\boldsymbol{G}_k}(\mathbf{Id},\boldsymbol{\varphi}_1)$ is finite can be established when $m \geqslant k \geqslant 1$.

4.5 Hamiltonian Control Generating the Geodesics

The Hamiltonian method reduces the parameterization of the vector field to the dynamics of the particles that encode the flow of states (13a). We write the dynamics explicitly as a linear function of the control:

$$\begin{aligned} &\dot{\boldsymbol{q}}_t = \boldsymbol{\xi}_{\boldsymbol{q}_t}(\boldsymbol{u}_t) := (\xi_{q_t^\ell}(u_t^\ell))_{\ell<\ell_{max}} \\ &\text{with } \ \xi_{q_t}(u_t) := (u_t(x_{i,t}), w_{i,t}\,\mathrm{div}u_t(x_{i,t}))_{i\in I} \ . \end{aligned} \tag{23}$$

We write the flow of the measure $\boldsymbol{\mu}(\boldsymbol{q}_t) := (\mu(q_t^\ell)_{\ell<\ell_{max}}$ with $\mu(q_t) := \sum_{i\in I} w_{i,t}\delta_{x_{i,t}} \otimes \delta_{f_i}$.

The control problem satisfying (12) reparameterized in the states becomes, for $\alpha > 0$:

$$\begin{cases} \min\limits_{\boldsymbol{u}.\in L^2([0,1],\boldsymbol{V})} \dfrac{1}{2}\displaystyle\int_0^1 \|\boldsymbol{A}\boldsymbol{u}_t\|_{\boldsymbol{V}}^2 dt + \dfrac{\alpha}{2}\|\boldsymbol{\mu}(\boldsymbol{q}_1) - \boldsymbol{\mu}_{obs}\|_{\boldsymbol{W}*}^2 \\ \quad \dot{\boldsymbol{q}}_t = \boldsymbol{\xi}_{\boldsymbol{q}_t}(\boldsymbol{u}_t),\ \boldsymbol{\mu}(\boldsymbol{q}_0) = \boldsymbol{\mu} \end{cases} . \tag{24}$$

Hamiltonian control introduces the co-states $\boldsymbol{p} = (p_i^{x,\ell}, p_i^{w,\ell})_{i\in I_\ell,\ell<\ell_{max}}$ with Hamiltonian

$$\begin{aligned} &\mathcal{H}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{u}) := (\ \boldsymbol{p} \mid \boldsymbol{\xi}_{\boldsymbol{q}}(\boldsymbol{u})) - \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u}\|_{\boldsymbol{V}}^2 \ , \\ &\quad \text{with } (\ \boldsymbol{p} \mid \boldsymbol{\xi}_{\boldsymbol{q}}(\boldsymbol{u})) = \textstyle\sum_{\ell<\ell_{max}} \left(\sum_{i\in I_\ell} \langle p_i^{x,\ell}, u^\ell(x_i^\ell)\rangle_{\mathbb{R}^d} + p_i^{w,\ell} w_i^\ell \mathrm{div} u^\ell(x_i^\ell)\right) \ . \end{aligned} \tag{25}$$

Under the assumption $V_{\ell_{max}} \hookrightarrow \mathscr{C}_0^{k+2}(\mathbb{R}^d,\mathbb{R}^d)$ the Pontryagin maximum [21] gives the optimum control satisfying for every ℓ:

$$\begin{cases} \dot{q}^\ell = \quad \partial_{p^\ell}\mathcal{H}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{u}) \\ \dot{p}^\ell = -\partial_{q^\ell}\mathcal{H}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{u}) \\ \partial_{u^\ell}\mathcal{H}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{u}) = 0 \end{cases} . \tag{26}$$

The first $\dot{q}^\ell$ specifies the state dynamics of Eqn. (23). The second $\dot{p}^\ell$ derives a differential equation which the co-state satisfies (see Eqn. (28a)). The third $\partial_{u^\ell}\mathcal{H}$ derives the optimal control equation (14a). See Appendix A for proof.

The co-state integral equation (14b) follows from the fact that it satisfies the differential equation for $\dot{p}^\ell$ of Eqn. (28a) (see Statement 2 of Appendix B). The fact that the endpoint added to the Hamiltonian $\boldsymbol{q} \mapsto \boldsymbol{U}(\boldsymbol{q})$ is continuously differentiable $\mathscr{C}_0^1$ implies that the solution to the differential equation is absolutely integrable.

4.6 Gradients of the Endpoint Varifold Matching Norm

The gradients of (15b) are efficiently rewritten using the state $q_t = (x_{i,t}, w_{i,t})_{i\in I}$ to define the norm-square in terms of h_q continuously differentiable in x and bounded $\mathscr{C}_b^{1,0}$,

$$h_{q_1} := \alpha K_W[\mu(q_1) - \mu_{obs}] \ \text{ on } \mathbb{R}^d \times \mathcal{F} \ . \tag{27}$$

We take the variation $\frac{d}{d\varepsilon}\boldsymbol{U}(\boldsymbol{q}(\varepsilon))|_{\varepsilon=0}$ varying each term $q_1(\varepsilon) = (x_{i,1}+\varepsilon\psi_i^x, w_{i,1}+\varepsilon\psi_i^w)_{i\in I}$ with dependence on scale implied.

$$\begin{aligned}
\frac{d}{d\varepsilon}\frac{\alpha}{2}\|\mu(q_1(\varepsilon)) - \mu_{obs}\|^2_{W*}|_{\varepsilon=0} &= \frac{d}{d\varepsilon}\frac{\alpha}{2}(\mu(q_1(\varepsilon)) - \mu_{obs} \mid K_W[\mu(q_1(\varepsilon)) - \mu_{obs}])|_{\varepsilon=0} \\
&= \frac{d}{d\varepsilon}(\mu(q_1(\varepsilon)) \mid \alpha\, K_W[\mu(q_1) - \mu_{obs}])|_{\varepsilon=0} \\
&= \sum_{i\in I}\frac{d}{d\varepsilon}\left((w_{i,1} + \varepsilon\psi_i^w)\, h_{q_1}(x_{i,1} + \varepsilon\psi_i^x, f_i)\right)|_{\varepsilon=0}\,. \\
&= \sum_{i\in I}(\langle w_{i,1}\nabla_{x_i} h_{q_1}(x_{i,1}, f_i), \psi_i^x\rangle_{R^d} + h_{q_1}(x_{i,1}, f_i)\psi_i^w)) = 0\,.
\end{aligned}$$

These represent the gradients of (15b). The matching condition has smooth gradients.

The tissue continuum of Computational Anatomy has $\mu = \int_{\mathbb{R}^d} \rho_c(x)\delta_x \otimes \mu_x\, dx$ with $q_t := (\varphi_t, w_t = w_0|d\varphi_t|)$, the measures parameterized by the state satisfy

$$\mu(q_t) := \varphi_t \cdot \mu = \int_{\mathbb{R}^d} w_t(x)\rho_c(x)\delta_{\varphi_t(x)} \otimes \mu_x dx\,.$$

The average of h_{q_1} over the feature space determines the boundary term variation.

$$\begin{array}{l}
\bar{h}^{\nabla}_{q_1}(x) := \int_{\mathcal{F}} \nabla_\varphi h_{q_1}(\varphi_1(x), f)\mu_x(df) \\
\bar{h}_{q_1}(x) := \int_{\mathcal{F}} h_{q_1}(\varphi_1(x), f)\mu_x(df)
\end{array}\,.$$

With $q_1 = (\varphi_1, w_1)$ take the variation $\varphi_1 \to \varphi_1(\varepsilon) = \varphi_1 + \varepsilon\psi^\varphi$, $w_1 \to w_1(\varepsilon) = w_1 + \varepsilon\psi^w$:

$$\begin{aligned}
\frac{d}{d\varepsilon}\frac{\alpha}{2}\|\mu(\varphi_1(\varepsilon), w_1) - \mu_{obs}\|^2_{W*}|_{\varepsilon=0} &= \frac{d}{d\varepsilon}\frac{\alpha}{2}(\mu(\varphi_1(\varepsilon), w_1) - \mu_{obs} \mid K_W[\mu(\varphi_1(\varepsilon), w_1) - \mu_{obs}])|_{\varepsilon=0}\,. \\
&= \frac{d}{d\varepsilon}(\mu(\varphi_1(\varepsilon), w_1) \mid h_{q_1})|_{\varepsilon=0} \\
&= \int_{\mathbb{R}^d\times\mathcal{F}} w_1(x)\rho_c(x)\frac{d}{d\varepsilon}h_{q_1}(\varphi_1(\varepsilon)(x), f)|_{\varepsilon=0}\mu_x(df)dx \\
&= \int_{\mathbb{R}^d}\langle w_1(x)\,\rho_c(x)\bar{h}^{\nabla}_{q_1}(x), \psi^\varphi(x)\rangle_{\mathbb{R}^d}dx = 0\,.
\end{aligned}$$

$$\begin{aligned}
\frac{d}{d\varepsilon}\frac{\alpha}{2}\|\mu(\varphi_1, w_1(\varepsilon)) - \mu_{obs}\|^2_{W*}|_{\varepsilon=0} &= \frac{d}{d\varepsilon}\frac{\alpha}{2}(\mu(\varphi_1, w_1(\varepsilon)) - \mu_{obs} \mid K_W[\mu(\varphi_1, w_1(\varepsilon)) - \mu_{obs}])|_{\varepsilon=0} \\
&= \frac{d}{d\varepsilon}(\mu(\varphi_1, w_1(\varepsilon)) \mid h_{q_1})|_{\varepsilon=0} \\
&= \int_{\mathbb{R}^d\times\mathcal{F}} \frac{d}{d\varepsilon}w_1(\varepsilon)(x)|_{\varepsilon=0}\rho_c(x)h_{q_1}(\varphi_1(x), f)\mu_x(df)dx \\
&= \int_{\mathbb{R}^d} \rho_c(x)\bar{h}_{q_1}(\varphi_1(x), f)\;\psi^w(x)dx\,.
\end{aligned}$$

This gives the gradients of the matching endpoint Eqn. (17). We note that computing the variation $\frac{d}{d\varepsilon}U(\varphi_1(\varepsilon), w_1(\varepsilon))|_{\varepsilon=0}$ requires U as a function of φ is $\mathscr{C}^1$ for $\varphi \in G_k \subset \mathrm{Diff}^k_{\mathrm{id}}(\mathbb{R}^d, \mathbb{R}^d)$ when $k \geqslant 2$.

Acknowledgements

Author Contributions

MM and AT designed the theoretical framework and DT built the computational implementation. All authors contributed to writing and editing the manuscript.

Funding

This work was supported by the National Institutes of Health (NIH) (www.nih.gov) grants R01EB020062 (MM), R01NS102670 (MM), U19AG033655 (MM), and R01MH105660 (MM), the National Science Foundation (NSF) (www.nsf.gov) 16-569 NeuroNex contract 1707298 (MM), the Computational Anatomy Science Gateway (DT and MM) as part of the Extreme Science and Engineering Discovery Environment (XSEDE Towns et al., 2014), which is supported by the NSF grant ACI1548562, Johns Hopkins University Alzheimer's Disease Research Center with NIH grant P50AG05146, the Dana Foundation's (www.dana.org) clinical neuroscience research program, and the Kavli Neuroscience Discovery Institute (kavlijhu.org) supported by the Kavli Foundation (www.kavlifoundation.org) (DT, MM, and JT).

Conflict of Interest

MM owns a founder share of Anatomy Works with the arrangement being managed by Johns Hopkins University in accordance with its conflict of interest policies. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The major contribution of this work is a mathematical and computational framework for modeling hierarchical neuroimaging data. Specific datasets shown in examples are for illustrative purposes, but can be made available upon request.

A Hamiltonian Control Statement

We take for our RKHS $V_\ell, \ell < \ell_{max}$ with the operators $L^\ell : V_\ell \to V_\ell{}^*$ defining the isometries such that the inner products $\langle \cdot, \cdot \rangle_{V_\ell}$ satisfy for any $\tilde{u}^\ell,\ \tilde{v}^\ell \in V_\ell$:

$$\langle \tilde{u}^\ell, \tilde{v}^\ell \rangle_{V_\ell} = (L^\ell \tilde{u}^\ell \mid \tilde{v}^\ell)\ .$$

Appendix Statement 1. *We assume that $V_{\ell_{max}} = \mathscr{C}_0^m(\mathbb{R}^d, \mathbb{R}^d)$ with $m \geqslant k+2$. If $\boldsymbol{u}_\cdot$ is a solution of the optimal control problem* (24) *then there exists time-dependent co-state $(t \mapsto \boldsymbol{p}_t)$ such that for any $i \in I_\ell$ for all ℓ:*

$$\begin{cases} \dot{q}_{i,t}^\ell = (u_t^\ell(x_{i,t}^\ell), w_{i,t}^\ell \operatorname{div} u_t^\ell(x_{i,t}^\ell)) \\ \dot{p}_{i,t}^{x,\ell} = -(du_t^\ell)^T(x_{i,t}^\ell) p_{i,t}^{x,\ell} - p_{i,t}^{w,\ell} w_{i,t}^\ell \nabla(\operatorname{div} u_t^\ell)(x_{i,t}^\ell) \\ \dot{p}_{i,t}^{w,\ell} = -p_{i,t}^{w,\ell} \operatorname{div} u_t^\ell(x_{i,t}^\ell) \end{cases} \tag{28a}$$

The optimal control satisfies $\partial_{u^\ell}\mathcal{H}(\boldsymbol{p},\boldsymbol{q},\boldsymbol{u})=0$ and $\boldsymbol{v}=\boldsymbol{A}\boldsymbol{u}$, for any $\ell<\ell_{max}$ ($v^{\ell_{max}}=0$):

$$\begin{cases} v_t^\ell(\cdot)=u_t^\ell(\cdot)-u_t^{\ell-1}(\cdot)=\sum_{\ell\leqslant j<\ell_{max}}\sum_{i\in I_j}\Big(g^\ell(x_{i,t}^j,\cdot)p_{i,t}^{x,j}+p_{i,t}^{w,j}w_{i,t}^j\,\nabla_1 g^\ell(x_{i,t}^j,\cdot)\Big) \\ u_t^\ell(\cdot)=\sum_{j<\ell_{max}}\sum_{i\in I_j}\Big(\bar g^{\ell\wedge j}(x_{i,t}^j,\cdot)p_{i,t}^{x,j}+p_{i,t}^{w,j}w_{i,t}^j\,\nabla_1\bar g^{\ell\wedge j}(x_{i,t}^j,\cdot)\Big)\ , \end{cases} \tag{28b}$$

with $\bar g^\ell:=\sum_{j=0}^\ell g^j$ and $\nabla_1 g(a,b)$ denotes the gradient of g^ℓ with respect to the first variable a.

Proof. Under the assumption $V_{\ell_{max}}\hookrightarrow\mathscr{C}_0^{k+2}(\mathbb{R}^d,\mathbb{R}^d)$ then we have $(\boldsymbol{u},\boldsymbol{q})\mapsto\boldsymbol{\xi_q}(\boldsymbol{u})$ is $\mathscr{C}^2$ and standard results of optimal control theory applying the Poyntryagin maximum principle [21] gives

$$\begin{cases}\dot{\boldsymbol{q}}=\partial_{\boldsymbol{p}}\mathcal{H}\\ \dot{\boldsymbol{p}}=-\partial_{\boldsymbol{q}}\mathcal{H}\\ \partial_{\boldsymbol{u}}\mathcal{H}=0\end{cases}.$$

Taking the variation for $\partial_{p^\ell}\mathcal{H}$ of the Hamiltonian for scale ℓ varies $p_i^{x,\ell}(\varepsilon)=p_i^{x,\ell}+\varepsilon\psi_i^{p^x}$, $p_i^{w,\ell}(\varepsilon)=p_i^{w,\ell}+\varepsilon\psi_i^{p^w}$:

$$\frac{d}{d\varepsilon}\big(p^\ell(\varepsilon)\ \big|\,\xi_{q^\ell}(u^\ell)\big)_{|\varepsilon=0}=\frac{d}{d\varepsilon}\sum_{i\in I_\ell}\Big(\langle p_i^{x,\ell}+\varepsilon\psi_i^{p^x},u^\ell(x_i^\ell)\rangle+(p_i^{w,\ell}+\varepsilon\psi_i^{p^w})w_i^\ell\,\mathrm{div}u^\ell(x_i)\Big)_{|\varepsilon=0} \tag{29}$$

which gives the first equation for the state velocity (28a).

Taking the variation for $\partial_{q^\ell}\mathcal{H}$ of the Hamiltonian for scale ℓ varies $x_i^\ell(\varepsilon)=x_i^\ell+\varepsilon\psi_i^x$, $w_i^\ell(\varepsilon)=w_i^\ell+\varepsilon\psi_i^w$:

$$\frac{d}{d\varepsilon}\big(p^\ell\ \big|\,\xi_{q^\ell(\varepsilon)}(u^\ell)\big)_{|\varepsilon=0}=\sum_{i\in I_\ell}\Big(\langle p_i^{x,\ell},du^\ell(x_i^\ell)\psi_i^{x,\ell}\rangle+p_i^{w,\ell}w_i^\ell\langle\nabla(\mathrm{div}\,u^\ell)(x_i^\ell),\psi_i^x\rangle+p_i^{w,\ell}(\mathrm{div}\,u^\ell)(x_i^\ell)\psi_i^w\Big) \tag{30}$$

so that we get the second two equations of (28a). To calculate $\partial_{u^\ell}\mathcal{H}=0$, define $u^\ell(\varepsilon)=u^\ell+\varepsilon\psi^u$ implying $v^{\ell+1}(\varepsilon)=v^{\ell+1}-\varepsilon\psi^u$, $v^\ell(\varepsilon)=v^\ell+\varepsilon\psi^u$ for $\psi^u\in V_\ell$. We have

$$\begin{aligned}&\frac{d}{d\varepsilon}\left(\big(p^\ell\ \big|\,\xi_{q^\ell}(u^\ell(\varepsilon))\big)-\frac{1}{2}\sum_{j=\ell}^{\ell+1}\big(L^jv^j(\varepsilon)\ \big|\,v^j(\varepsilon)\big)\right)_{|\varepsilon=0} \\ =\ &\sum_{i\in I_\ell}\Big(\langle p_i^{x,\ell},\psi^u(x_i^\ell)\rangle+p_i^{w,\ell}w_i^\ell(\mathrm{div}\,\psi^u)(x_i^\ell)\Big)-\big(L^\ell v^\ell-L^{\ell+1}v^{\ell+1}\ \big|\,\psi^u\big)=0\ .\end{aligned} \tag{31}$$

After summation of (31) for $\ell\geqslant\ell_0$, we get for any $\psi^u\in V_{\ell_0}$ that

$$(L^{\ell_0}v^{\ell_0}\mid\psi^u)=\sum_{\ell\geqslant\ell_0}\sum_{i\in I_\ell}\Big(\langle p_i^{x,\ell},\psi^u(x_i^\ell)\rangle+p_i^{w,\ell}w_i^\ell(\mathrm{div}\,\psi^u)(x_i^\ell)\Big) \tag{32}$$

Now, for any $x,\alpha\in\mathbb{R}^d$, consider $\delta_x^\alpha\in V_{\ell_0}^*$ such that $(\delta_x^\alpha\mid\tilde v^\ell)=\langle\tilde v^\ell(x),\alpha\rangle_{\mathbb{R}^d}$ for any $\tilde v^{\ell_0}\in V_{\ell_0}$. The reproducing property on V_{ℓ_0} gives $\langle K^{\ell_0}\delta_x^\alpha,\ K^{\ell_0}\delta_y^\beta\rangle_{V_{\ell_0}}=g^{\ell_0}(x,y)\langle\alpha,\beta\rangle$. We get from (32) for $\psi^u=K^{\ell_0}\delta_x^\alpha=g^{\ell_0}(.,x)\alpha$ that

$$\langle v^{\ell_0}(x),\alpha\rangle=(L^{\ell_0}v^{\ell_0}\mid K^{\ell_0}\delta_x^\alpha)=\sum_{\ell\geqslant\ell_0}\sum_{i\in I_\ell}\Big(g^{\ell_0}(x_i^\ell,x)\langle p_i^{x,\ell},\alpha\rangle+p_i^{w,\ell}w_i^\ell\langle\nabla_1 g^{\ell_0}(x_i^\ell,x),\alpha\rangle\Big)$$

so that we get the first equality above of (28b) for v^{ℓ_0}.

Now since $u^{\ell_0} = \sum_{k \leqslant \ell_0} v^k$ we deduce the equality for u^{ℓ_0} given in (28b). □

B Hamiltonian co-state momentum integrable dynamics

We omit the superscripts ℓ below since co-states and states and flows are at the same scale.

Appendix Statement 2. *Assume $\boldsymbol{q} \to \boldsymbol{U}(\boldsymbol{q})$ is $\mathscr{C}^1$ in $\boldsymbol{q}$, then the co-state integral equations* (14b) *flowing from $t = 1$ solves the Hamiltonian differential equations $(\dot{p}^x, \dot{p}^w)$ of* (28a)*; the integral equations flowing from $t = 0$ satisfy:*

$$\begin{cases} p^x_{i,t} = (d\varphi_{t,0})^T(x_{i,t})p^x_{i,0} - p^w_{i,0}w_{i,0}\displaystyle\int_0^t (d\varphi_{t,s})^T(x_{i,t})\,\nabla(\mathrm{div} u_s)(x_{i,s})ds \\ p^w_{i,t} = p^w_{i,0}|d\varphi_{t,0}|(x_{i,t}) \end{cases} . \tag{33}$$

Proof. First take $p^w_{i,t} = p^w_{i,0}|d\varphi_{t,0}|(x_{i,t})$ and show it satisfies (28a). For $w_{i,t} = w_i|d\varphi_t|(x_i)$, $x_{i,t} = \varphi_t(x_i)$ then

$$\frac{d}{dt}\, w_{i,t}p^w_{i,t} \quad = \quad \frac{d}{dt}\, w_{i,0}|d\varphi_t|(x_i)p^w_{i,0}|d\varphi_t^{-1}|(x_{i,t}) = 0 \tag{34}$$

which implies $\dot{p}^w_{i,t} = -\frac{p^w_{i,t}\dot{w}_{i,t}}{w_{i,t}} = -p^w_{i,t}\mathrm{div}\, u_t(x_{i,t})$.

The $p^x_{i,t}$ satisfies (28a); rewrite the integral solution using $\varphi_{t,s} := \varphi_s \circ \varphi_t^{-1}$ and the identities:

$$p^x_{i,t} = (d\varphi_t^{-1})^T \circ \varphi_t(x_i)p^x_{i,0} - p^w_{i,0}w_{i,0}\int_0^t (d\varphi_t^{-1})^T \circ \varphi_t(x_i)d\varphi_s^T(x_i)\nabla(\mathrm{div}\, u_s)(x_{i,s})dx\ .$$

Differentiating requires the identity $\frac{d}{dt}\,(d\varphi_t^{-1})^T \circ \varphi_t = -(du_t)^T \circ \varphi_t\,(d\varphi_t^{-1})^T \circ \varphi_t$ (see below):

$$\begin{aligned} \dot{p}^x_{i,t} \quad &= \quad -(du_t)^T(x_{i,t})\left((d\varphi_t^{-1})^T(x_{i,t})p^x_{i,0} - p^w_{i,0}w_{i,0}\int_0^t (d\varphi_t^{-1})^T(x_{i,t})d\varphi_s^T(x_i)\nabla(\mathrm{div}\, u_s)(x_{i,s})dx\right) \\ &\quad -p^w_{i,0}w_{i,0}(d\varphi_t^{-1})^T(x_{i,t})d\varphi_t^T(x_i)\nabla(\mathrm{div}\, u_t)(x_{i,t}) \\ &= \quad -(du_t)^T(x_{i,t})p^x_{i,t} - p^w_{i,t}w_{i,t}\nabla(\mathrm{div}\, u_t)(x_{i,t})\ , \end{aligned}$$

where the last equality uses (34), $w_{i,0}p^w_{i,0} = w_{i,t}p^w_{i,t}$. The remaining identity follows $\frac{d}{dt}((d\varphi_t)^{-1}\, d\varphi_t) = 0$:

$$\begin{aligned} \frac{d}{dt}(d\varphi_t)^{-1} \quad &= \quad -(d\varphi_t)^{-1}(d\dot{\varphi}_t)(d\varphi_t)^{-1} \\ &= \quad -(d\varphi_t)^{-1}\, du_t \circ \varphi_t = (d\varphi_t^{-1}) \circ \varphi_t\, du_t \circ \varphi_t\ . \end{aligned}$$

□

We point out that the optimal control (14a) is written in the endpoint with the boundary conditions (14b). Using the form of $\frac{d}{dt}((d\varphi_t)^{-1}\, d\varphi_t)$ gives $p^x_{i,t}$ in the boundary $t = 1$ satisfies the differential equation. For $p^w_{i,t}$ then constancy $p^w_{i,t}w_{i,t} = p^w_{i,1}w_{i,1}$ with $|d\varphi_{t,1}|(x_{i,t}) = |d\varphi_1|(x_i)|d\varphi_t^{-1}|(x_{i,t})$ gives the endpoint:

$$p^w_{i,t} = \frac{p^w_{i,1}w_{i,1}}{w_{i,t}} = \frac{p^w_{i,1}|d\varphi_1(x_i)|}{|d\varphi_t(x_i)|} = p^w_{i,1}|d\varphi_{t,1}|(x_{i,t})\ .$$

References

[1] J. R. Ecker, D. H. Geschwind, A. R. Kriegstein, J. Ngai, P. Osten, D. Polioudakis, A. Regev, N. Sestan, I. R. Wickersham, and H. Zeng, "The brain initiative cell census consortium: Lessons learned toward generating a comprehensive brain cell atlas," *Neuron*, vol. 96, no. 3, pp. 542 –557, 2017, ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2017.10.007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0896627317309364.

[2] S. Shah, E. Lubeck, W. Zhou, and L. Cai, "In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus," *Neuron*, vol. 92, pp. 342–357, Oct. 2016. DOI: 10.1016/j.neuron.2016.10.001.

[3] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, "Spatially resolved, highly multiplexed rna profiling in single cells," *Science*, vol. 348, no. 6233, 2015, ISSN: 0036-8075. DOI: 10.1126/science.aaa6090. eprint: https://science.sciencemag.org/content/348/6233/aaa6090.full.pdf. [Online]. Available: https://science.sciencemag.org/content/348/6233/aaa6090.

[4] X. Wang*, W. Allen*, M Wright, E Sylwestrak, N Samusik, S Vesuna, K Evans, C Liu, C Ramakrishnan, J Liu, G. P. Nolan, F.-A. Bava, and K Deisseroth, "Three-dimensional intact-tissue-sequencing of single-cell transcriptional states.," *Science*, 2018. DOI: eaat5691.

[5] M. P. Nagle, G. S. Tam, E. Maltz, Z. Hemminger, and R. Wollman, "Bridging scales: From cell biology to physiology using in situ single-cell technologies," *Cell systems*, vol. 12, no. 5, 388–400, 2021, ISSN: 2405-4712. DOI: 10.1016/j.cels.2021.03.002. [Online]. Available: https://doi.org/10.1016/j.cels.2021.03.002.

[6] M Albert, A Soldan, R Gottesman, G McKhann, N Sacktor, L Farrington, M Grega, R. Turner, B. Lu, S Li, M.-C. Wang, O Selnes, and B. R. Team, "Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to apoe genotype," *Current Alzheimer Research*, vol. 11, no. 8, 773–784, 2014.

[7] M. Miller, L. Younes, J. Ratnanather, T. Brown, H. Trinh, E. Postell, D. Lee, M.-C. Wang, S. Mori, R. O'Brien, and M. Albert, "The diffeomorphometry of temporal lobe structures in preclinical alzheimer's disease," *NeuroImage: Clinical*, vol. 3, pp. 352–360, Dec. 2013. DOI: 10.1016/j.nicl.2013.09.001.

[8] D. Tward, T. Brown, Y. Kageyama, J. Patel, Z. Hou, S. Mori, M. Albert, J. Troncoso, and M. Miller, "Diffeomorphic registration with intensity transformation and missing data: Application to 3d digital pathology of alzheimer's disease," *Frontiers*, 2019.

[9] M. I. Miller, J. Fan, and D. J. Tward, "Multi scale diffeomorphic metric mapping of spatial transcriptomics datasets," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021, pp. 4472–4480.

[10] D. Snyder and M. Miller, *Random Point Processes in Time and Space*, ISBN-13: 978-1461278214. Springer-Verlag, 1991.

[11] J. Talairach and P. Tournoux, *Co-Planar Stereotaxis Atlas of the Human Brain*. Stuttgart: Georg Thiem Verlag, 1988.

[12] J. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "Probabilistic atlas of the human brain: Theory and rationale for its development," *NeuroImage*, vol. 2, pp. 89–101, 1995.

[13] M. Ono, S. Kubik, and C. D. Abernathey, *Atlas of the Cerebral Sulci*. Georg Thieme Verlag, 1990.

[14] N. Chuang, S. Mori, A. Yamamoto, H. Jiang, X. Ye, X. Xu, L. J. Richards, J. Nathans, M. I. Miller, A. W. Toga, *et al.*, "An mri-based atlas and database of the developing mouse brain," *Neuroimage*, vol. 54, no. 1, pp. 80–89, 2011.

[15] S. M. Sunkin, L. Ng, C. Lau, T. D. andTerri L. Gilbert, C. L. Thompson, M. Hawrylycz, and C. Dang, "Allen brain atlas: An integrated spatio-temporal portal for exploring the central nervous system," *Nucleic Acids Res.*, vol. D996-D1008, no. 41(Database issue), 2013.

[16] D. C. V. Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, and K. Ugurbil, "The wu-minn human connectome project: An overview," *NeuroImage*, vol. 80, pp. 62 –79, 2013, Mapping the Connectome, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2013.05.041. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811913005351.

[17] U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quarterly of Applied Mathematics*, vol. 56, no. 4, pp. 617–694, 1998, ISSN: 0033569X. [Online]. Available: http://portal.acm.org/citation.cfm?id=309082.309089.

[18] P. M. Thompson and A. W. Toga, "A framework for computational anatomy," *Computing and Visualization in Science*, vol. 5, no. 1, pp. 13–34, 2002, ISSN: 14329360. DOI: 10.1007/s00791-002-0084-6. [Online]. Available: http://www.springerlink.com/openurl.asp?genre=article\&id=doi:10.1007/s00791-002-0084-6.

[19] J Ashburner, "Computational anatomy with the spm software," *Magnetic Resonance Imaging*, vol. 27, pp. 1163–1174, 8 2009.

[20] X Pennec, "From Riemannian Geometry to Computational Anatomy," *Elements*, 2011.

[21] M. I. Miller, A. Trouvé, and L. Younes, "Hamiltonian systems and optimal control in computational anatomy: 100 years since d'arcy thompson.," *Annual Review of Biomed Engineering*, no. 17, pp. 447–509, 2015.

[22] W. Allard, "On the first variation of a varifold," *Ann. of Math.*, vol. 95, pp. 417–491, 1972.

[23] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139–157, 2005, ISSN: 09205691. DOI: 10.1023/B:VISI.0000043755.93987.aa. [Online]. Available: http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:VISI.0000043755.93987.aa.

[24] M. I. Miller, D. J. Tward, and A. Trouvé, "Coarse-to-fine hamiltonian dynamics of hierarchical flows in computational anatomy," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 3760–3765. DOI: 10.1109/CVPRW50498.2020.00438.

[25] S. Kulason, D. J. Tward, T. Brown, C. S. Sicat, C.-F. Liu, J. T. Ratnanather, L. Younes, A. Bakker, M. Gallagher, M. Albert, M. I. Miller, and the Alzheimer's Disease Neuroimaging Initiative, "Cortical thickness atrophy in the transentorhinal cortex in mild cognitive impairment," *Neuroimage Clin.*, vol. 21:101617, 2019. [Online]. Available: doi:10.1016/j.nicl.2018.101617.

[26] W. Wells, W. Grimson, R Kikinis, and F. Jolesz, "Adaptive segmentation of mri data," *IEEE Trans Med Imaging*, vol. 4, no. 15, pp. 429–42, 1996. DOI: doi:10.1109/42.511747.

[27] G. Gerstein and N. Kiang, "An approach to the quantitative analysis of electrophysiological data from single neurons.," *Biophysical journal*, vol. 1, no. 1, 15–28, 1960. [Online]. Available: https://doi.org/10.1016/s0006-3495(60)86872-5.

[28] E. D. Boer and P. Kuyper, "Triggered correlation," *IEEE Transactions on Bio-Medical Engineering*, vol. 15, no. 3, 1968.

[29] R. Satij, J. A. Farrell, D. Gennert, F Schier Alexander, and A. Regev, "Spatial reconstruction of single-cell gene expression data," *Nature Biotechnology*, vol. 33, 2015. [Online]. Available: https://doi.org/10.1038/nbt.3192.

[30] J. Fan, N. Salathia, R. Liu, G. E. Kaeser, Y. C. Yung, J. L. Herma, F. Kaper, J.-B. Fan, K. Zhang, J. Chun, and P. V. Kharchenko, "Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis," *Nature Methods*, vol. 13, 241–244, 2016.

[31] P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, and J. Frisén, "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics," *Science*, vol. 353, no. 6294, pp. 78–82, 2016, ISSN: 0036-8075. DOI: 10.1126/science.aaf2403. eprint: http://science.sciencemag.org/content/353/6294/78.full.pdf. [Online]. Available: http://science.sciencemag.org/content/353/6294/78.

[32] J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, E. Vaughn, K. Shekhar, J. D. Perez, N. D. Rubinstein, J. Hao, A. Regev, C. Dulac, and X. Zhuang, "Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region," *Science*, vol. 362, no. 6416, 2018, ISSN: 0036-8075. DOI: 10.1126/science.aau5324. eprint: http://science.sciencemag.org/content/362/6416/eaau5324.full.pdf. [Online]. Available: http://science.sciencemag.org/content/362/6416/eaau5324.

[33] C. Xia, J. Fan, G. Emanuel, J. Hao, and X. Zhuang, "Spatial transcriptome profiling by merfish reveals subcellular rna compartmentalization and cell cycle-dependent gene expression," *Proceedings of the National Academy of Sciences*, vol. 116, no. 39, pp. 19 490–19 499, 2019, ISSN: 0027-8424. DOI: 10.1073/pnas.1912459116. eprint: https://www.pnas.org/content/116/39/19490.full.pdf. [Online]. Available: https://www.pnas.org/content/116/39/19490.

[34] A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron, *Registration of functional data using fisher-rao metric*, 2011. arXiv: 1103.3817 [math.ST].

[35] N. Charon and A. Trouvé, "The varifold representation of nonoriented shapes for diffeomorphic registration," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2547–2580, 2013. DOI: 10.1137/130918885. eprint: http://dx.doi.org/10.1137/130918885. [Online]. Available: http://dx.doi.org/10.1137/130918885.

[36] H.-W. Hsieh and N. Charon, "Metrics, quantization and registration in varifold spaces," *Foundations of Computational Mathematics*, pp. 1–45, 2021.

[37] L. Risser, F.-X. Vialard, R. Wolz, M. Murgasova, D. D. Holm, and D. Rueckert, "Simultaneous Multi-scale Registration Using Large Deformation Diffeomorphic Metric Mapping.," *IEEE Transactions on Medical Imaging*, vol. 30, no. 10, pp. 1746–59, 2011, ISSN: 15580062. DOI: 10.1109/TMI.2011.2146787. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21521665.

[38] S. Sommer, F. Lauze, M. Nielsen, and X. Pennec, "Sparse multi-scale diffeomorphic registration: The kernel bundle framework," *Journal of Mathematical Imaging and Vision*, vol. 46, no. 292, 2013. DOI: 10.1007/s10851-012-0409-0. [Online]. Available: https://doi.org/10.1007/s10851-012-0409-0.

[39] M. I. Miller, A. Trouvé, and L. Younes, "Geodesic Shooting for Computational Anatomy.," *Journal of Mathematical Imaging and Vision*, vol. 24, no. 2, pp. 209–228, 2006, ISSN: 15737683. DOI: 10.1007/s10851-005-3624-0. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2897162\&tool=pmcentrez\&rendertype=abstract.

[40] Y Zhang, M Brady, and S Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Trans Med Imaging*, vol. 20(1), pp. 45–57, 2001. DOI: 10.1109/42.906424.