HIERARCHICAL COMPUTATIONAL ANATOMY: UNIFYING THE MOLECULAR TO TISSUE CONTINUUM VIA MEASURE REPRESENTATIONS OF THE BRAIN

MICHAEL I. MILLER, DANIEL TWARD, AND ALAIN TROUVÉ

ABSTRACT. This paper presents a unified representation of the brain based on mathematical functional measures integrating the molecular and cellular scale descriptions with continuum tissue scale descriptions. We present a fine-to-coarse recipe for traversing the brain as a hierarchy of measures projecting functional description into stable empirical probability laws that unifies scale-space aggregation. The representation uses measure norms for mapping the brain across scales from different measurement technologies. Brainspace is constructed as a metric space with metric comparison between brains provided by a hierarchy of Hamiltonian geodesic flows of diffeomorphisms connecting the molecular and continuum tissue scales. The diffeomorphisms act on the brain measures via the 3D varifold action representing "copy and paste" so that basic particle quantities that are conserved biologically are combined with greater multiplicity and not geometrically distorted. Two applications are examined, the first histological and tissue scale data in the human brain for studying Alzheimer's disease, and the second the RNA and cell signatures of dense spatial transcriptomics mapped to the meso-scales of brain atlases. The representation unifies the classical formalism of computational anatomy for representing continuum tissue scale with non-classical generalized functions appropriate for molecular particle 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 [10]) which brings the nano and micron scales of single cell measures of RNA via spatial transcriptomics [38, 8, 48, 33, 35] coupled to the tissue scales of mouse atlases. The recent emergence of spatial transcriptomics as method of the year [51] highlights the importance and ascendence of such approaches for understanding the dense metric structure of the brain at the cellular scales. In our Alzheimer's BIOCARD study [36] we are examining pathological Tau at both the micro histological and macro atlas scales of Tau particle detections, from 10-100 μm [45] and to human magnetic resonance millimeter scales for examining entire circuits in the medial temporal lobe [53]. 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 [46, 47].

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 [39]. In contrast, information associated to atlasing methods at gross anatomical tissue and organ scales extend smoothly [43, 12, 24, 32, 22, 23, 31, 9, 42, 11]. 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

Date: April 19, 2021.

for which advection and transport hold [20, 44, 1, 4, 34, 28, 27], principles upon which continuum mechanics and its analogues are based, but not directly applicable for particles.

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 [14, 3, 16] 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. **Theoretical Results.**

2.1.1. *Measure representation of the particle and atomistic structures of the brain.* 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. The 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.

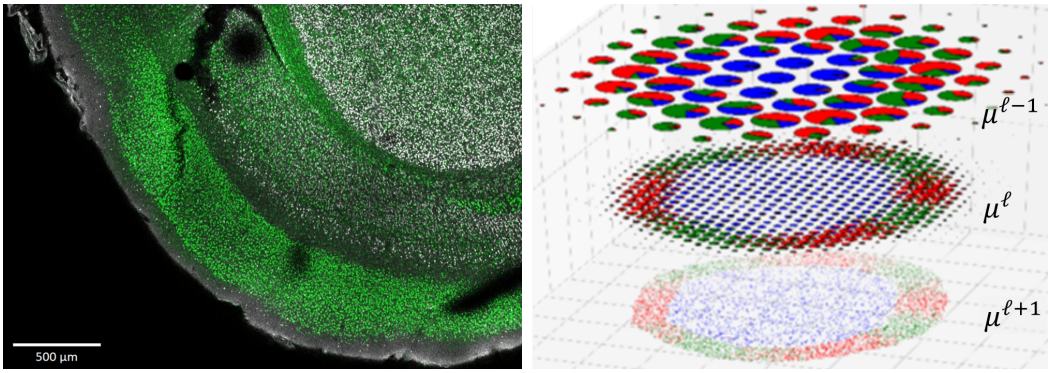

FIGURE 1. Left: Tissue from a NexCre+/-;Brn2fl/+ adult mouse mimicing a wild-type mouse with CUX1 labelling of layer ii/iii,iv and CTIP2 in layers v,vi in green. Shows sections at $2.5^2 \times 50\,\mu\text{m}^3$ 6 tile images, 1433×1973 pixels; taken from Uli Mueller. Right: Showing the abstraction 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 cascade $\mu^{\ell} \overset{T_1}{\to} \nu \overset{T_2}{\to} \mu^{\ell'}$ resampling the hierarchy of measures.

Our representation must exist 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_c 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 $f(x)$ but a probability distribution μ_x of 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 their marginals ρ, $\rho(A) := \mu(A \times \mathcal{F})$ for measurable $A \subset \mathbb{R}^d$, and their family of conditional probability distributions on f given x :

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

with 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 number 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

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

2.1.2. *Transforming scales via empirical laws.* 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 right panel of Figure 1, our model becomes a sequence of measures:

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

The idealization 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.

The idealization of each particle and feature as a generalized function, a measure, requires that it is paired or dual to the measurement process or measurement probes $\delta_z : h \mapsto h(z)$. The scale space of the measurement probes are introduced via kernel functions carrying the resolution scale σ or reciprocally a bandwidth, something analogous to Planck's scale. For 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 bottom of Figure 1 is our construction of a composition of transformations crossing scales, consisting of 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 $\mathcal{Z} = \mathbb{R}^d \times \mathcal{F}$:

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

The second transforms nonlinearly to the new features $\mathcal{F}'$ via the kernel $k_2((x, \alpha), \cdot)$ transforming any feature probability $\alpha \in \mathcal{M}_P(\mathcal{F})$. Decomposing $\nu(dx, df) = \rho(dx)\nu_x(df)$ into its marginal and conditional gives

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

We now examine crossing measure scales $\mu^\ell, \ell = 0, 1, \ldots$ resampling into a hierarchy of lattices for histology and spatial transcriptomics. At every scale μ^ℓ remains a measure allowing for the procedure of (5a), (5b) for scale-space renormalization to be executed recursively.

2.1.3. *Rescaling across Computational Lattices.* Our hierarchy crosses scales by resampling the measures $\mu^\ell = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$ across lattices or grids allowing us to interpolate between the continuum and our discrete computational codes at different scales. Resampling projects one set of particles $(x_i \in \mathbb{R}^d)_{i\in I}$ to the resampling lattice of particles and tiles $(y_j \in Y_j \subset \mathbb{R}^d)_{j\in J}$, with $\cup_j Y_j \subset \mathbb{R}^d$ forming a disjoint partition that for regular lattces are a complete covering. We reweight with $\pi(x_i, Y_j), \pi(x_i, \mathbb{R}^d) = 1$ which computes the fraction that particle x_i shares with the lattice site Y_j for all particles x_i and tiles Y_j.

The kernel transformation k_1 performs space reduction with the weights of π, and k_2 feature transformation via maps from machine learning $\phi : \mathcal{F} \to \mathcal{F}'$:

$$\begin{aligned} k_1((x,f),(\cdot,\cdot)) &= \sum\nolimits_{j\in J} \pi(x, Y_j)\, \delta_{y_j}(\cdot) \otimes \delta_f(\cdot) \\ k_2((x,\nu_x),(\cdot,\cdot)) &= \delta_x(\cdot) \otimes \delta_{\phi(\nu_x)}(\cdot) \end{aligned} \quad . \tag{6a}$$

This gives the resampled measure with new space density $\rho' = \sum_{j\in J} w_j \delta_{y_j}$:

$$\mu' = \sum\nolimits_{j\in J} w_j\, \delta_{y_j} \otimes \delta_{\phi(\nu_{y_j})} \tag{6b}$$

$$\text{with} \begin{cases} w_j &= \sum_{i\in I} w_i\, \pi(x_i, Y_j) \\ \nu_{y_j} &= \frac{1}{w_j} \sum_{i\in I} w_i\, \pi(x_i, Y_j)\, \delta_{f_i} \end{cases} .$$

See Methods 4.1 for derivation. The indexing across scales is defined disjointly $I \cap J = \varnothing$.

2.1.4. *Brainspace as a Hierarchy of Diffeomorphic Flows.* We want to build correspondences at any scale, each scale having different numbers of particles and resolutions. We do this by connecting the discrete particles through the continuum using the diffeomorphism group. Connection between sample brains occurs through diffeomorphic transport using the group of k-times continuously differentiable diffeomorphisms at any layer $\varphi \in G_k$ with group operation function composition $\varphi \circ \varphi'$. For any brain $\mu = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$ the diffeomorphisms act $(\varphi, \mu) \mapsto \varphi \cdot \mu$ according to

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

The 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 increase accordingly and are not conserved, in contrast to classic measure or probability transport. We call this action the varifold action, and it corresponds to "copy and paste".

To accomodate scales we take the product group of diffeomorphisms with elements $\boldsymbol{\varphi} = (\varphi^\ell)_{\ell<\ell_{max}} \in \boldsymbol{G}_k$ acting component-wise (Figure 2), with group action on the hierarchy given by

$$\boldsymbol{\varphi} \cdot \boldsymbol{\mu} := (\varphi^\ell \cdot \mu^\ell)_{\ell<\ell_{max}} \, . \tag{8}$$

Dynamics occurs through the group generated as a dynamical system in which the vector fields $t \mapsto u_t^\ell$ control the diffeomorphisms:

$$\dot{\varphi}_t^\ell = u_t^\ell \circ \varphi_t^\ell \, . \tag{9}$$

As depicted in Figure 2 the controls are coupled across scales by successive refinements $v^\ell, \ell < \ell_{max}$:

$$u^\ell = u^{\ell-1} + v^\ell \, , \tag{10}$$

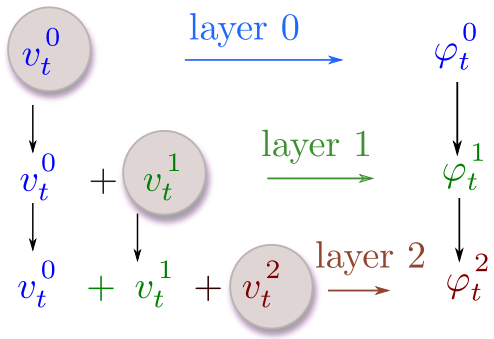

FIGURE 2. Hierarchical system, controls $u^\ell = u^{\ell-1} + v^\ell, u^0 = v^0$ and flows $\varphi^\ell, \ell = 0, 1, \ldots$.

with our convention taking $u^{-1} = 0$. The controls are in reproducing kernel spaces (RKHS's) V_ℓ, norm $\|\cdot\|_{V_\ell}$ with the hierarchy $\boldsymbol{u} := (u^\ell)_{\ell<\ell_{max}}$ in the product $\boldsymbol{V} = \prod_{\ell<\ell_{max}} V_\ell$. The kernels are taken as diagonal $K^\ell(\cdot,\cdot) = g^\ell(\cdot,\cdot)\mathrm{id}_d$, id_d a $d \times d$ identity matrix (see [29] for non-diagonal kernels),The accumulation across scale of Green's functions becomes

$$\bar{g}^\ell := \sum_{k=0}^{\ell} g^k \,. \tag{11}$$

2.1.5. *Geodesic structure of Brainspace.* Dynamics translates into a navigation in the orbit $\boldsymbol{G}_k \cdot \boldsymbol{\mu}$. Geodesic mapping $t \mapsto \boldsymbol{\varphi}_t \cdot \boldsymbol{\mu}$ flows the control along minimum energy onto $\boldsymbol{\mu}_{obs}$ encoded as a endpoint matching condition. The matching condition measures when two brains are equal, defined using the measure norm. Brains with 0 norm difference are equal; brains with small normed difference are similar. Every brain has a variable number of particles, with no apriori correspondence between particle dimensions. Measure norms accomodate these variabilities. The measure norm at each scale is denoted $\|\mu\|^2_{W_\ell^*}$; across scales it becomes

$$\|\boldsymbol{\mu}\|^2_{\boldsymbol{W}^*} = \sum_{\ell<\ell_{max}} \|\mu^\ell\|^2_{W_\ell^*} \,. \tag{12}$$

See the Methods section 4.4 for the construction of the measure norm.

Geodesic mapping solves for the optimal $\boldsymbol{u}_\cdot := (\boldsymbol{u}_t, t \in [0,1])$ with $\alpha > 0$:

$$\begin{cases} \min_{\boldsymbol{u}_\cdot \in L^2([0,1],\boldsymbol{V})} \frac{1}{2} \sum_{\ell=0}^{\ell_{max}-1} \int_0^1 \|u_t^\ell - u_t^{\ell-1}\|^2_{V_\ell} dt + \frac{\alpha}{2} \|\boldsymbol{\varphi}_1 \cdot \boldsymbol{\mu} - \boldsymbol{\mu}_{obs}\|^2_{\boldsymbol{W}^*} \\ \dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t \circ \boldsymbol{\varphi}_t, \ \boldsymbol{\varphi}_0 = \mathbf{Id} \end{cases} . \tag{13}$$

Hamiltonian control reparameterizes the flow via the state $\boldsymbol{q}_t := (q_t^\ell)_{\ell<\ell_{max}}$:

$$q_t^\ell := (x_{i,t}^\ell = \varphi_t^\ell(x_i^\ell), w_{i,t}^\ell = w_i^\ell |d\varphi_t^\ell|(x_i^\ell))_{i \in I_\ell} \,. \tag{14}$$

The state encodes the action of (8) as $\boldsymbol{\mu}(\boldsymbol{q}_t) := \boldsymbol{\varphi}_t \cdot \boldsymbol{\mu}$, allowing us to rewrite the endpoint in the state:

$$\boldsymbol{U}(\boldsymbol{q}_1) := \frac{\alpha}{2} \|\boldsymbol{\mu}(\boldsymbol{q}_1) - \boldsymbol{\mu}_{obs}\|^2_{\boldsymbol{W}^*} = \frac{\alpha}{2} \|\boldsymbol{\varphi}_1 \cdot \boldsymbol{\mu} - \boldsymbol{\mu}_{obs}\|^2_{\boldsymbol{W}^*} \,. \tag{15}$$

Define $\varphi_{t,s}^\ell = \varphi_s^\ell \circ (\varphi_t^\ell)^{-1}$ at any scale. To write the optimal control explicitly, we require two smoothness conditions: (i) the controls are in RKHS's V_ℓ at least twice continuously differentiable $\mathscr{C}_0^m$, $m \geqslant 2$, and (ii) the endpoint function $\boldsymbol{U}(\boldsymbol{q}_1)$ as a function of the state $\boldsymbol{q}$ is smoothly differentiable $\mathscr{C}^1$.

Then the optimal control of (13) satisfies for every $\ell < \ell_{max}$:

$$u_t^\ell(\cdot) = \sum_{\ell' < \ell_{max}} \sum_{i \in I_{\ell'}} \left(\bar{g}^{\ell \wedge \ell'}(x_{i,t}^{\ell'}, \cdot) p_{i,t}^{x,\ell'} + p_{i,t}^{w,\ell'} w_{i,t}^{\ell'} \nabla_1 \bar{g}^{\ell \wedge \ell'}(x_{i,t}^{\ell'}, \cdot) \right) , \tag{16a}$$

$$\text{with} \quad \begin{cases} p_{i,t}^{x,\ell} = (d\varphi_{t,1}^\ell)^T(x_{i,t}^\ell) p_{i,1}^{x,\ell} + p_{i,1}^{w,\ell} w_{i,1}^\ell \int_t^1 (d\varphi_{t,s}^\ell)^T(x_{i,t}^\ell) \nabla(\mathrm{div} u_s^\ell)(x_{i,s}^\ell) ds \\ p_{i,t}^{w,\ell} = p_{i,1}^{w,\ell} |d\varphi_{t,1}^\ell|(x_{i,t}^\ell) \\ \text{b.c.'s} \quad \begin{array}{l} p_{i,1}^{x,\ell} = -\nabla_{x_i^\ell} \boldsymbol{U}(\boldsymbol{q}_1) \\ p_{i,1}^{w,\ell} = -\nabla_{w_i^\ell} \boldsymbol{U}(\boldsymbol{q}_1) \end{array} \end{cases} . \tag{16b}$$

Here $\bar{g}$ is from (11) and $\nabla_1 \bar{g}(a,b)$ denotes the gradient with respect to the first variable. The optimal controls couple the hierarchy across scales with the refininements given by $v^\ell = u^\ell - u^{\ell-1}, u^{-1} = 0$. Classical optimal control theory gives the gradient matching endpoints of (16b), $\boldsymbol{p}_1 + \nabla_{\boldsymbol{q}} \boldsymbol{U}(\boldsymbol{q}_1) = 0$. Equation (26) of the Methods section 4.4 gives the explicit calculation of the gradients of the endpoint conditions based on the Hilbert space construction of the measure norm.

2.2. **Experimental Results.**

2.2.1. *Gaussian Scale-Space Resampling of Tau Histology.* Figure 3 shows the multi-scale data from the clinical BIOCARD study [2] in which we are studying the spatio-temporal flow of pre-clinical biomarkers of Alzheimer's disease within the medial temporal lobe [26, 45, 21, 54]. The top row (left panel) shows the clinical magnetic resonance imaging (MRI) collected for subjects longitudinally over the decades of the study. The top row (middle) shows the high-field $200\,\mu$m MRI scale depicting the medial temporal lobe including the collateral sulcus and lateral bank of the entorhinal cortex for an Alzheimer's subject along with their histology (right) at μm scale associated to the high-field MRI section.

The molecular histology measure encodes the detected Tau and amyloid particles with the size and geometric variation $\mu = \sum_i \delta_{x_i} \otimes \delta_{f_i}, \mathcal{F} = \mathbb{R}^+$. We take the weights as identically $w_i = 1$ for fine scale particles with the functional features $f_i \in \mathcal{F}$ either size and geometric variation of particles. Crossing from the histology projects the Tau particles $(x_j \in \mathbb{R}^d)_{i \in I}$ accumulating Tau on the resampling tissue lattice $(y_j \in Y_j \subset \mathbb{R}^d)_{j \in J}$ using Gaussian scale-spaces based on normal reweighting in $\mathbb{R}^2$. At the original scale the $w_i = 1$ for all particles. Resampling the feature space projects onto moments of Tau, shown are two:

$$\pi_\sigma(x_i, Y_j) = 1/(2\pi\sigma^2) \int_{Y_j} e^{-|y - x_i|^2/2\sigma^2} dy \, .$$

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

$$\text{with} \quad \begin{cases} w_j = \sum_{i \in I} \pi_\sigma(x_i, Y_j), \; \nu_{y_j} = \frac{1}{w_j} \sum_{i \in I} \pi_\sigma(x_i, Y_j) \, \delta_{f_i} \\ \phi(\nu_{y_j}) = \left(\int_{\mathcal{F}} f \nu_{y_j}(df), \int_{\mathcal{F}} (f)^2 \nu_{y_j}(df) \right) \in \mathcal{F}' \end{cases} .$$

Figure 3 (bottom two rows) shows the detected Tau pathology from the output of a machine learning algorithm depicting the Tau particles as red dots (left column) for two sections at 4μ m scale. The right two columns shows the first several moments of the whole sections at the tissue scale depicting the mean and variance of the particle size. Feature reduction for matching brains across scales projects the conditional feature distributions ν_y associated to the tissue lattice onto their moments. The measure representation and measure norm in navigation takes these features as components in calculating distances.

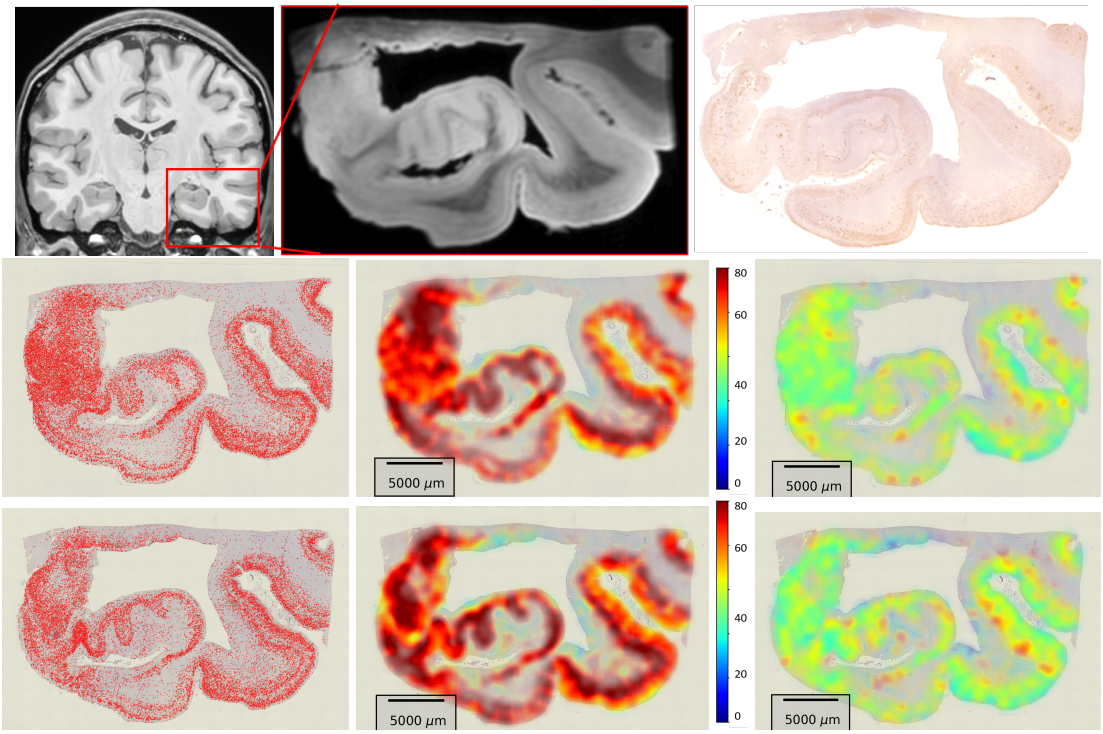

FIGURE 3. Top: Medial lobe at 1 mm and high-field 200 μm MRI along with 4 μm Tau histology (right). Bottom rows: Panels from the Alzheimer brain (top row) showing the 4 μm Tau histology section with the first moments of Tau particle size (columns 2,3) shown at tissue scales; deepest red color bar denotes 80 μm^2 Tau particle area.

Shown in Figure 4 is the 4 μm histology (rows 1,2, column 1) are several sections with the output of a machine learning algorithm depicted as red dots showing the estimation of the positions of the Tau particles. Columns 2 and 3 depict the mathematical measure representation of the perirhinal cortex constructed from the positions and sizes at the 4μ m scale (column 2) and the tissue scale (column 3) using Gaussian resampling onto the tissue lattice. The color bar indicates the largest particle size as deep red at 80 μm^2. The gradients in tau tangle area between superficial and deep layers is apparent with the deep red as high as 80 μm^2 for the innermost layers of tissue.

The rightmost panel (top) shows the metric comparison of the perirhinal cortex mapping the two sections in column 1. The vector fields shown in the right column encodes the geodesic transformation showing the narrowing of the banks of the perirhinal cortex exhibiting tissue scale motions order 1000 μm (brightness on color bar).

The bottom three rows of Figure 4 shows these measures navigating through the orbit mapping one fold onto the other in the collateral sulcus at the boundary of the trans entorhinal cortex region. Associated to the motions are order 1000 μm fold deformation. Notice the multiple scales of transformation depicted between row 3 and rows 4 and 5. Rows 4 and 5 show the transformation based on the first two moments of the feature size encoded in ν_y based on the measure norm distances.

2.2.2. *Spatial Transcriptomics.* Methods in spatial-transcriptomics which have emerged for localizing and identifying cell-types via marker genes and across different cellular resolutions [37, 13, 41, 30, 48] presents the opportunity of localizing in spatial coordinates the transcriptionally distinct cell-types.

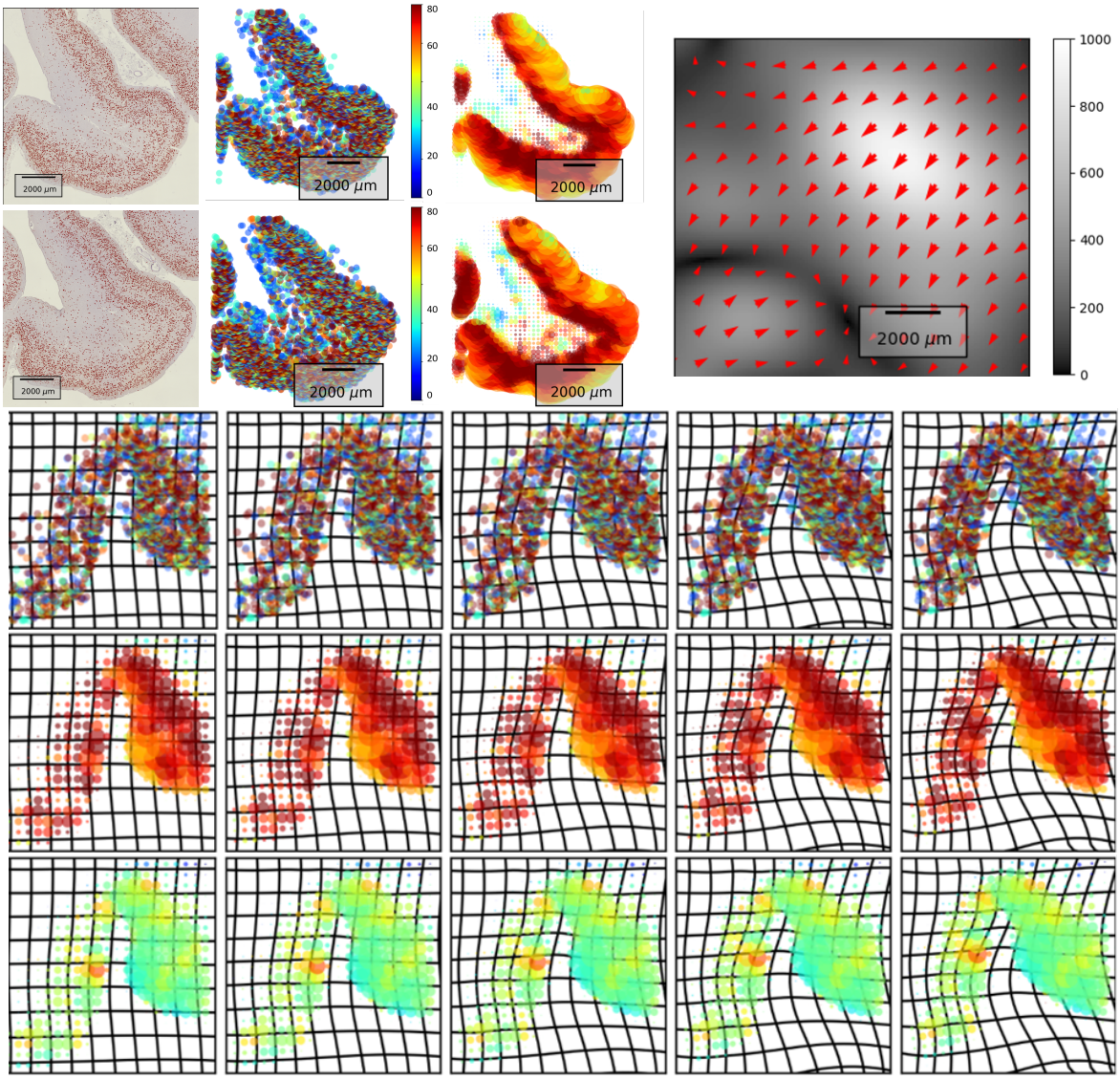

FIGURE 4. Top: Showing detected Tau 4μ m histology (column 1) and mathematical measures depicting molecular and tissue scales (column 2, 3) showing the first moments of Tau particle size on the perirhinal cortex; saturated red color denotes $80\,\mu$m^2 Tau area; rows show sections. Top right: Showing the metric size of the deformation via the vector field mapping of the perirhinal sulcus narrowing with motions having maximal value $1000\,\mu$m. Bottom: Grids depicting geodesic navigation $\varphi_t \cdot \mu, t \in [0,1]$ from collateral sulcus (left column, Figure 3) showing widening of folds order $1000\ \mu$m motion across molecular and tissue scales depicting first two moments.

We now examine the molecular and tissue scale brain measures transferring from RNA molecules to cells examining the building blocks at cellular scales of mixtures of RNA functional types measured with MERFISH [50]. We examine the spatial transcriptomics across two-scales going from the molecular RNA to cellular and then tissue as depicted in Figure 5.

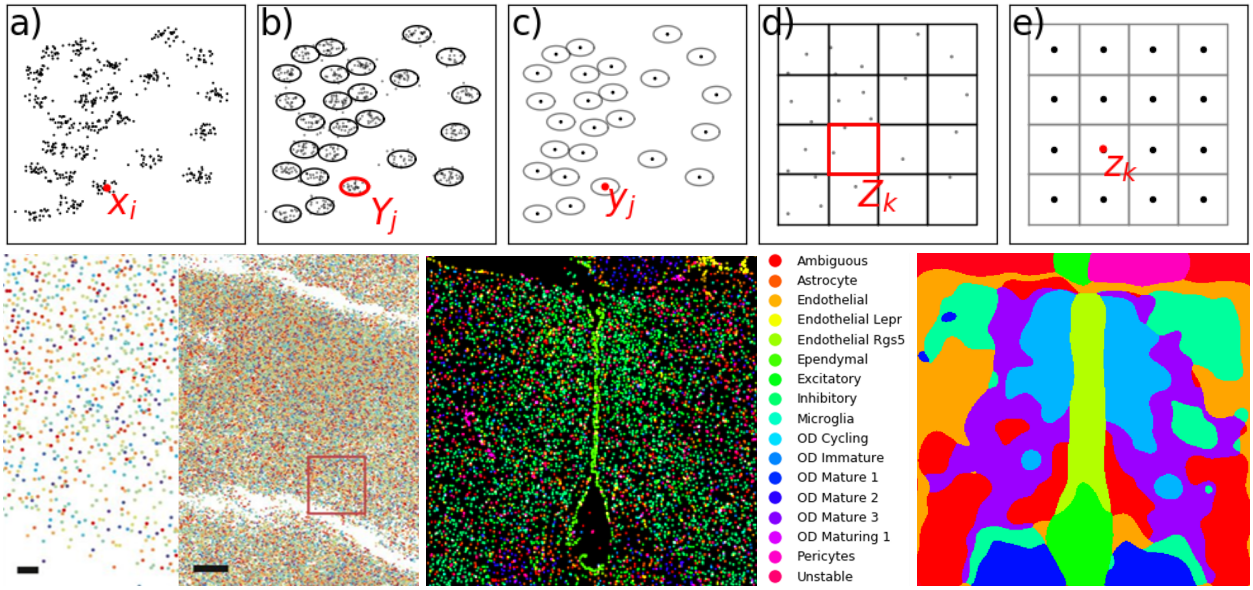

FIGURE 5. Top: Showing the multi-scale renormalization from particle measure $(\delta_{x_i})_{i\in I}$ to cells and cell-centers $(\delta_{y_j}, y_j \in Y_j)_{j\in J}$ to tissue tiling $(\delta_{z_k}, z_k \in Z_k)_{k\in K}$. Bottom: MERFISH [50] showing RNA molecules (left) shown as colored markers depicting locations of the 167 RNA gene species (bar scales $1, 10\,\mu$m); middle shows 17 cell types as colored particles clustered on cell centers with feature vectors eigen function reduction of 167 dimension RNA; right shows $k-$means clustering to tissue type.

The molecular RNA measured at the sub micron scale is aggregated to represents the cells and tissue at the micron scales. This is depicted in the cartoon idealization in the top row of Figure 5.

The molecular measures represent RNA locations with features sparse RNA vectors, $\mu = \sum_{i\in I} \delta_{x_i} \otimes \delta_{f_i}$, $\mathcal{F} = \mathbb{R}^{167}$. Crossing to cells $(Y_j \subset \mathbb{R}^2)_{j\in J}$ we partition $I = \cup_{j\in J}\mathscr{I}_j$ into subsets of particles closest to the cells:

$$\mathscr{I}_j = \{i \in I : d(x_i, Y_j) < d(x_i, Y_{j'}), j' \neq j\}\ , j \in J\ .$$

The RNA particles $(x_i \in \mathbb{R}^d)_{i\in I}$ are resampled to the cell centers $(\delta_{y_j}, y_j \in Y_j \subset \mathbb{R}^d)_{j\in J}$ accumulating to nonsparse mixtures of RNA with probabilities computed on the 17 cell-types $\mathcal{F}' \subset [0,1]^{17}$:

$$\pi(x_i, Y_j) = \begin{cases} 1 & \text{for } x_i, i \in \mathscr{I}_j \\ 0 & \text{for } x_i, i \notin \mathscr{I}_j \end{cases}\ ,$$

$$\mu' = \sum\nolimits_{j\in J} w_j\, \delta_{y_j} \otimes \delta_{\phi'(\nu_{y_j})}$$

$$\text{with } \begin{cases} w_j = |\mathscr{I}_j|,\ \nu_{y_j} = \frac{1}{|\mathscr{I}_j|}\sum_{i\in\mathscr{I}_j}\delta_{f_i} \\ \phi'(\nu_{y_j}) = (f'_{j,1}, \ldots, f'_{j,17}) \in \mathcal{F}' \subset [0,1]^{17} \end{cases}\ ,$$

To calculate probability vectors in $\mathcal{F}'$ we cluster the 167-dimensional RNA features into a statistical summary of reduced dimension based on principle components (PCA), orthogonal PCA functions $E_1, E_2, \ldots$. We project the conditional feature laws, $\varepsilon_{y_j}(n) = \int E_n(f)\nu_{y_j}(df)$ and compute the conditional probabilities

$$f'_{j,c} = \Pr\{C = c | \varepsilon_{y_j}(1), \varepsilon_{y_j}(2), \ldots\}\ , c = 1, \ldots, 17\ .$$

Crossing to tissue resamples the cells to the lattice $(\delta_{z_k}, z_k \in Z_k \subset \mathbb{R}^2)_{k\in K}$ using Gaussian resampling:

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

$$\mu'' = \sum\nolimits_{k\in K} w_k \delta_{z_k} \otimes \delta_{\phi''(\nu_{z_k})}$$

$$\text{with } \begin{cases} w_k = \sum_{j\in J} |\mathscr{J}_j|\, \pi_\sigma(y_j, Z_k), \nu_{z_k} = \frac{1}{w_k}\sum_{j\in J} |\mathscr{J}_j|\, \pi_\sigma(y_j, Z_k)\, \delta_{f''_j} \\ \phi''(\nu_{z_k}) = \left(f''_{k,1}, \ldots, f''_{k,10}\right) \in \mathcal{F}'' \subset [0,1]^{10} \end{cases} .$$

The tissue scale conditional probabilities are modelled on 10 types with feature vectors computed using K=10-means clustering on ν_{z_k}. The output of K-means is a set of orthogonal indicator functions $E_1, E_2, \ldots, E_{10}$, where $E_n(f) = 1$ if f is closer to the center of cluster n than the center of any other cluster, and 0 otherwise. Cluster distance in K-means is computed using the Fisher-Rao metric computed as the angle between the square root of pairs of probability mass functions $\nu_{z_k}, \nu_{z_{k'}}$, each viewed as points on an 18 dimensional sphere (17 cell types plus one backround). The projection onto the E_n, $\varepsilon_{z_k}(n) = \int E_n(f)\nu_{z_k}(df)$ gives conditional probabilities

$$f''_k(c) = \Pr\{C = c | \varepsilon_{z_k}(1), \ldots, \varepsilon_{z_k}(10)\} \ .$$

In both scales depicted in Figure 5 the probabilities are modeled as concentrated on one class with probability 1 and the others 0. The bottom row of Figure 5 (left panels) shows the RNA measure μ depicted as colored markers with different colored dots the features corresponding to the 100's of different gene species (bar scale $1, 10$ microns). The middle panel μ' shows the labels of the cell type feature space of 17 cell types 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, and the right panel shows the tissue type feature space of 10 tissue types associated to the K means procedure.

Shown in Figure 6 are results from Wang et. al. 2018 [48] depicting neuronal cell types classified via high dimensional features vectors of as many as $20,000$ genes. Cell types depicted as colors include ecitatory

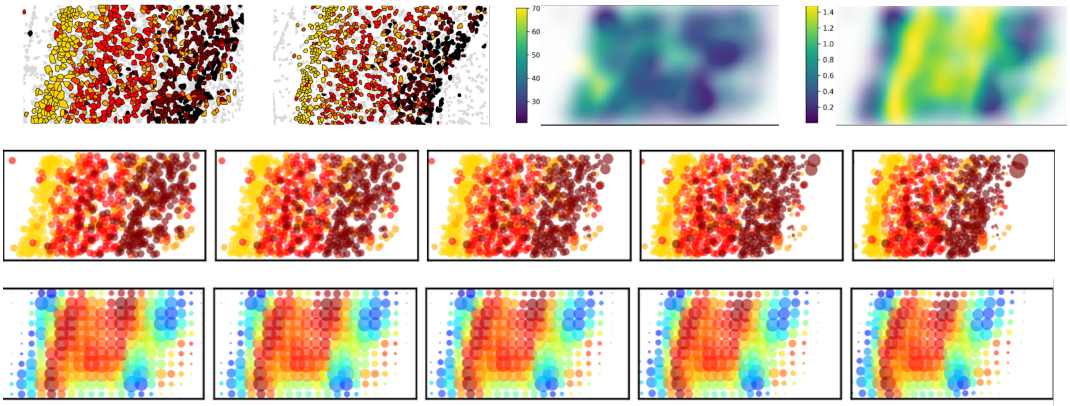

FIGURE 6. Top: Cells from Wang et. al. 2018 [48] depicting neuronal cell types; right panels show cell size and entropy of the conditional distributions ν_x for the single section in panel 1. Rows 2 and 3: Mapping $\varphi_t \cdot \mu$ at the fine cell scales showing the cell types mapping with the bottom row mapping the entropy measure of each particle.

cells yellow eL2/3, orange eL4, red eL5, and inhibitory cell green SST, light blue VIP, for example. Figure

6 (top panels 1 and 2) shows the cell types in the cortical tissue coordinates [48]. The right two panels show features from the mathematical measure $\sum_i w_i \delta_{x_i} \otimes \delta_{f_i}$, $\mathcal{F} = \{\text{cell types}\}$. Panel 3 shows the tissue scale with the conditional probabilities projected onto the features of mean cell type size with panel 4 showing the conditional entropy of cell type. The entropy is a measure of dispersion across the cell identities given by the expectation of the log probability function. An entropy of zero means that space location feature disribution ν_x has all its mass on 1 cell type. The bottom two rows of Figure 6 show the mapping $\varphi_t \cdot \mu$ of the spatial transcriptome measure at the fine scale showing the particles (middle row); the bottom row depicts the entropy of the measure of each particle.

2.2.3. *Cellular Neurophysiology:* Single unit neurophysiology uses space-time models of spiking neurons as inhomogeneous Poisson counting processes [39]. Neurons x_i are modelled as counting measures in time $N_i(t), t \geqslant 0$ with intensity $\lambda(t), t \geqslant 0$, so that

$$\Pr(N_i(t) = n) = \frac{(\int_{t_0}^t \lambda(s)ds)^n}{n!} e^{-\int_{t_0}^t \lambda(s)ds} .$$

We associate to each neuron the set of spike times $f_i = (t_k)_{1 \leqslant k \leqslant n_{f_i}}$ with feature space $\mathcal{F} = \{f = (t_k)_{1 \leqslant k \leqslant n_f} : n_f \geqslant 1, t_1 < t_2 \cdots < t_{n_f}\}$. A neural network becomes

$$\mu = \sum_i \delta_{x_i} \otimes \delta_{f_i} \ , x_i \in \mathbb{R}^d, f_i \in \mathcal{F} .$$

Post-stimulus time (PST) [17] and interval histograms are used often in neurophysiology to examine the instantaneous discharge rates and inter-spike interval statistics [6, 55], the first a linear dimension reduction on feature space and the second non-linear. The interval histogram abandons the requirement of maintaining the absolute phase of the signal for average empirical statistics measuring temporal periodicity and phase locking [55, 40]. Synchronized discharge rates are computed using binning functions $[b_i, b_{i+1})$, $i = 1, \cdots, B$ and Fourier weighting j=$\sqrt{-1}$, frequencies $n = 0, 1, \ldots$:

$$\begin{cases} \phi_n^{pst}(\nu_x) = \sum_i e^{\mathrm{j}\omega_n i} \int_{\mathcal{F}} \frac{1}{n_f} \sum_{k=1}^{n_f} \mathbf{1}_{[b_i, b_{i+1})}(t_k) \nu_x(df) \ , \\ \phi_n^{int}(\nu_x) = \sum_i e^{\mathrm{j}\omega_n i} \int_{\mathcal{F}} \frac{1}{n_f} \sum_{k=1}^{n_f} \mathbf{1}_{[b_i, b_i+1)}(t_k - t_{k-1}) \nu_x(df) \ . \end{cases}$$

The $n = 0$ frequency computes integrated rate; each phase-locked feature is complex $\phi_n \in \mathbb{C}$. Interval histograms provide an exponentially weighted representation of the autocorrelation function (see Appendix **??**) and therefore an asymptotically biased estimator of the power-spectrum.

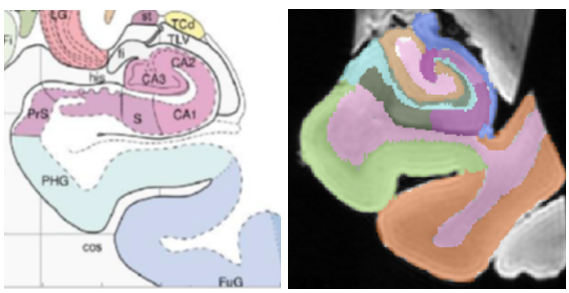

FIGURE 7. Left shows Mai-Paxino atlas of the high-field MRI section of the Medial Temporal Lobe with hippocampus and entorhinal cortex compartments shown on the right.

2.2.4. *Dense MRI images and Large Deformation Diffeomorphic Metric Mapping (LDDMM).* The tissue continuum has been studied extensively in MRI and Computational Anatomy as depicted in Figure 7. Take as the dense tissue model $\rho(dx) = \rho_c(x)dx$ with

$$\mu = \int_{\mathbb{R}^d} \rho_c(x)\delta_x \otimes \mu_x dx \ . \tag{17}$$

Lattice implementations on voxels in imagery has $\rho = \sum_{i\in I} \delta_{x_i}$ reducing the integrals to sums as in (1), (7).

For real-valued images such as for atlasing and 1 mm scale MRI (see Figure 7), the functional feature is identified with the scalar image $f(x) \in \mathbb{R}^+ = \mathcal{F}, x \in \mathbb{R}^3$ with the feature value often taken as deterministic $\mu_x = \delta_{f(x)}$ with $\rho_c = 1$. The action becomes equivalent to the action for LDDMM [5]:

$$\varphi_t \cdot \mu := \int_{\mathbb{R}^d} |d\varphi_t(x)|\delta_{\varphi_t(x)} \otimes \delta_{f(x)} dx = \int_{\mathbb{R}^d} \delta_x \otimes \delta_{f\circ\varphi_t^{-1}(x)} dx \ .$$

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 [25]. 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 necesarily 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 [7]. We call this property "copy and paste", the notion being that the brain is built on basic structuring elements that are conserved.

The use of feature reduction as a key step in crossing scales is commonly done in segmentation techniques at tissue scales [49, 15]. Bayes classifiers are constructed for parcellating the brain for studying neurodevelopmental and neurodegenerative disease. Feature reduction maps gray levels in $\mathcal{F} = [0, 255]$ to decision regions $\theta_n \subset [0, 255]$. Feature reduction maps to a probability vector on tissue types, $\phi : \mathcal{F} \to \mathcal{F}' = (R^+)^N$ integrating over the decision regions, $\phi_n(\nu_x) = \int_{\theta_n} \nu_x(df) = p_n, n = 1, \ldots, N$.

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 [56] and tissue segmentation [52]. Building neighborhood relations as conditional probabilities between lattice sites from which global probabilites laws are constructed with the Hammersley-Clifford theorem [18] links us to Grenander's metric pattern theory [19] formalisms with the atoms and conditional laws $(x_i, \nu_{x_i})_{i\in I}$ at any scale playing the roles of the generators.

4. METHODS

4.1. **Method of Resampling Across Scales of Lattices.** Taking $\mu = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$ with kernels (6a) gives our two-step transformation and (6b). Define $w_j = \sum_{i\in I} w_i \pi(x_i, Y_j)$, $\nu_{y_j} = \sum_{i\in I} w_i \pi(x_i, Y_j)\delta_{f_i}/w_j$, then the calculation with $\mu = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$ gives

$$\begin{aligned} T_1 : \mu \mapsto \quad \nu \quad &= \int_{\mathbb{R}^d \times \mathcal{F}} k_1((x, f), \cdot)\, \mu(dx, df) \\ &= \int_{\mathbb{R}^d \times \mathcal{F}} \sum_{j\in J} \pi(x, Y_j)\, \delta_{y_j} \otimes \delta_f \sum_{i\in I} w_i\, \delta_{x_i} \otimes \delta_{f_i}\, (dx, df) \\ &= \sum_{j\in J} \delta_{y_j} \otimes \left(\sum_{i\in I} w_i \pi(x_i, Y_j) \delta_{f_i} \right) = \sum_{j\in J} w_j \delta_{y_j} \otimes \nu_{y_j} \ . \end{aligned}$$

The space density follows from

$$\rho = \int_{\mathcal{F}} \sum_{j\in J} w_j \delta_{y_j} \otimes \nu_{y_j}(df) = \sum_{j\in J} w_j \delta_{y_j} \ .$$

The feature transformation $\phi : \mathcal{M}_P \to \mathcal{F}'$ gives (6b):

$$\begin{aligned} T_2 : \nu \ \mapsto \ \mu' \ &= \int_{\mathbb{R}^d} k_2((x, \nu_x), \cdot)\, \rho(dx) \\ &= \int_{\mathbb{R}^d} \delta_x \otimes \delta_{\phi(\nu_x)} \left(\sum_{j\in J} w_j \delta_{y_j}(dx) \right) = \sum_{j\in J} w_j \delta_{y_j} \otimes \delta_{\phi(\nu_{y_j})} \ . \end{aligned}$$

4.2. **Method of Comparison via the Geodesic Metric.** For any $\mu = \sum_{i\in I} w_i \delta_{x_i} \otimes \delta_{f_i}$, the diffeomorphisms act $(\varphi, \mu) \mapsto \varphi \cdot \mu$ according to (7). To accomodate scale the product group for the hierarchy becomes

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

with elements $\boldsymbol{\varphi} \in \boldsymbol{G}_k$ satisfying law of composition component-wise $\boldsymbol{\varphi} \circ \boldsymbol{\varphi}' = (\varphi^\ell \circ \varphi'^\ell)_{\ell<\ell_{max}}$ (Figure 2). The group acts on the hierarchy componentwise according to (7), (8) $\boldsymbol{\varphi} \cdot \boldsymbol{\mu}$.

Dynamics occurs through the group action on the brain measures generated as a dynamical system in which the hierarchical control $t \mapsto \boldsymbol{u}_t := (u_t^\ell)_{\ell<\ell_{max}}$ flows the hierarchy $t \mapsto \boldsymbol{\varphi}_t$ with ODE of (9) $\dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t \circ \boldsymbol{\varphi}_t$.

Each control u^ℓ is constrained to be an element of a reproducing kernel Hilbert space (RKHS) V_ℓ with inner-product $\langle\,,\,\rangle_{V_\ell}$. The total $\boldsymbol{u}$ is in the product $\boldsymbol{u} \in \boldsymbol{V} = \prod_{\ell<\ell_{max}} V_\ell$ with inner-product determining the norm $\langle\,,\,\rangle_{\boldsymbol{V}} = \sum_{\ell<\ell_{max}} \langle\,,\,\rangle_{V_\ell}$. The spaces are organised as a sequence of continous embeddings:

$$V^0 \hookrightarrow \cdots \hookrightarrow V^{\ell_{max}} ,$$

where $V^{\ell_{max}}$ is an additional layer containing the others. The hierarchy is connected via the continuous linear operator $\boldsymbol{A} : \boldsymbol{V} \to \boldsymbol{V}$ with $\boldsymbol{v} = \boldsymbol{A}\boldsymbol{u}$ for $u^\ell = u^{\ell-1} + v^\ell, u^0 = v^0$. 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

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

Optimal curves which minimize the integrated energy $\mathcal{E}_{\boldsymbol{A}}$ between any two fixed boundary conditions (BC) $\boldsymbol{\varphi}_0$ and $\boldsymbol{\varphi}_1$ where $\boldsymbol{\varphi}_1$ is accessible with a path of finite energy extends the LDDMM setting [5] to a hierarchy of diffeomorphisms and describes a geodesic for an associated Riemannian metric [29] on $\boldsymbol{G}_k$:

$$d_{\boldsymbol{G}_k}(\boldsymbol{\varphi}_0, \boldsymbol{\varphi}_1)^2 := \min_{\substack{\boldsymbol{u}_\cdot \in L^2([0,1],\boldsymbol{V}):\, \dot{\boldsymbol{\varphi}}_t = \boldsymbol{u}_t \circ \boldsymbol{\varphi}_t \\ \text{with BC } \boldsymbol{\varphi}_0, \boldsymbol{\varphi}_1}} \mathcal{E}_{\boldsymbol{A}}(\boldsymbol{u}_\cdot) \tag{18}$$

The metric distance from $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_1$ in the orbit $\boldsymbol{G}_k \cdot \boldsymbol{\mu}_0$ is given by the shortest length paths of (18) with boundary conditions $\boldsymbol{\varphi}_1 \cdot \boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$. Existence of solutions can be established for minimizers of (18). Specifically, for the group G_k properly defined and $V^{\ell_{max}}$ the space of m-times continuously differentiable vector fields vanishing at infinity with $m \geqslant k \geqslant 1$, the metric holds. See A for technical statement.

4.3. **Method via Hamiltonian Control.** The Hamiltonian method reduces the parameterization of the vector field to the essential dynamics of the particles that encode the flow. Towards this end, Hamiltonian control introduces state processes of (14) $q_t = (x_{i,t} = \varphi_t(x_i), w_{i,t} = w_i |d\varphi_t|(x_i))_{i\in I}$ that reparameterize the controls. State dynamics are linear in the control since $\dot{\varphi}_t = u_t \circ \varphi_t$ gives

$$\dot{q}_t = (u_t(x_{i,t}), w_{i,t}\, \mathrm{div} u_t(x_{i,t}))_{i\in I} \;,\; q_0 = (x_i, w_i)_{i\in I} \;, \tag{19a}$$

and rewriting the right hand side gives

$$\xi_{q_t}(u_t) = (u_t(x_{i,t}), w_{i,t}\, \mathrm{div} u_t(x_{i,t}))_{i\in I} \tag{19b}$$

with the hierarchy sastisfying $\dot{\boldsymbol{q}}_t = \boldsymbol{\xi}_{\boldsymbol{q}_t}(\boldsymbol{u}_t) := (\xi_{q_t^\ell}(u_t^\ell))_{\ell<\ell_{max}}$.

Measure evolution is encoded as $\boldsymbol{\mu}(\boldsymbol{q}_t) := \boldsymbol{\varphi}_t \cdot \boldsymbol{\mu}_0$, with

$$\mu(q_t^\ell) := \varphi_t^\ell \cdot \mu^\ell = \sum\nolimits_{i\in I_\ell} w_{i,t}^\ell \delta_{x_{i,t}^\ell} \otimes \delta_{f_i^\ell} .$$

The optimal control problem satisfying (13) reparameterized in the states becomes, for $\alpha > 0$:

$$\begin{cases} \min_{\boldsymbol{u}_\cdot \in L^2([0,1],\boldsymbol{V})} \mathcal{E}_{\boldsymbol{A}}(\boldsymbol{u}_\cdot) + \frac{\alpha}{2}\|\boldsymbol{\mu}(\boldsymbol{q}_1) - \boldsymbol{\mu}_{obs}\|^2_{\boldsymbol{W}^*} \\ \dot{\boldsymbol{q}}_t = \boldsymbol{\xi}_{\boldsymbol{q}_t}(\boldsymbol{u}_t),\ \boldsymbol{\mu}(\boldsymbol{q}_0) = \boldsymbol{\mu}_0 \end{cases} . \tag{20}$$

The Hamiltonian controls the states via co-states $\boldsymbol{p} = (\boldsymbol{p}^x, \boldsymbol{p}^w)$ as Lagrange multipliers:

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{u}) = (\ \boldsymbol{p} \mid \boldsymbol{\xi}_{\boldsymbol{q}}(\boldsymbol{u})) - \mathcal{E}_{\boldsymbol{A}}(\boldsymbol{u}_\cdot)$$
$$\text{with } (\ \boldsymbol{p} \mid \boldsymbol{\xi}_{\boldsymbol{q}}(\boldsymbol{u})) = \sum\nolimits_{\ell < \ell_{max}} \left(\sum\nolimits_{i \in I_\ell} \langle p_i^{x,\ell}, u^\ell(x_i^\ell) \rangle_{\mathbb{R}^d} + p_i^{w,\ell} w_i^\ell \mathrm{div} u^\ell(x_i^\ell) \right) .$$

The Pontryagin maximum [28] gives the optimum control $\partial_{u^\ell} \mathcal{H}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{u}) = 0$, with dynamics for every ℓ,

$$\begin{cases} \dot{q}^\ell = \ \ \partial_{p^\ell} \mathcal{H}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{u}) \\ \dot{p}^\ell = -\partial_{q^\ell} \mathcal{H}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{u}) \end{cases} . . \tag{21}$$

Statement 1. *We assume that $V^{\ell_{max}} = \mathscr{C}_0^m(\mathbb{R}^d, \mathbb{R}^d)$ with $m \geqslant 2$. If $\boldsymbol{u}_\cdot$ is a solution of the optimal control problem* (20) *then there exists time-dependent co-state such that for every $i \in I_\ell$ and $\ell < \ell_{max}$:*

$$\begin{cases} \dot{q}_{i,t}^\ell = (u_t(x_{i,t}^\ell), w_{i,t}^\ell \operatorname{div} u_t(x_{i,t}^\ell)) \\ \dot{p}_{i,t}^{x,\ell} = -(du_t)^T(x_{i,t}^\ell) p_{i,t}^{x,\ell} - p_{i,t}^{w,\ell} w_{i,t}^\ell \nabla(\operatorname{div} u_t)(x_{i,t}^\ell) \\ \dot{p}_{i,t}^{w,\ell} = -p_{i,t}^{w,\ell} \operatorname{div} u_t(x_{i,t}^\ell) \\ u_t^\ell(\cdot) = \sum\nolimits_{\ell' < \ell_{max}} \sum\nolimits_{i \in I_{\ell'}} \left(\bar{g}^{\ell \wedge \ell'}(x_{i,t}^{\ell'}, \cdot) p_{i,t}^{x,\ell'} + p_{i,t}^{w,\ell'} w_{i,t}^{\ell'} \nabla_1 \bar{g}^{\ell \wedge \ell'}(x_{i,t}^{\ell'}, \cdot) \right) \end{cases} . \tag{22}$$

The Hamiltonian differential equations are derived in Appendix B using calculus of variations; the averaged Green's function $\bar{g}$ is Eqn. (11). The endpoint condition $\boldsymbol{U}(\boldsymbol{q}_1)$ is added to the Hamiltonian with smoothness of $\boldsymbol{U}(\boldsymbol{q}_1)$ in the state giving absolutely integrable solutions to the Hamitonian dynamics. We omit the superscripts ℓ below since co-states and states and flows are at the same scale.

Statement 2. *Assume $\boldsymbol{q} \to \boldsymbol{U}(\boldsymbol{q})$ is $\mathscr{C}^1$ in $\boldsymbol{q}$, then the co-state integral equations* (16b) *flowing from the endpoints $t = 1$ solve the Hamiltonian differential equations* (22)*; the integral equations flowing from the initial conditions $t = 0$ satisfy:*

$$\begin{cases} p_{i,t}^x = (d\varphi_{t,0})^T(x_{i,t}) p_{i,0}^x - p_{i,0}^w w_{i,0} \int_0^t (d\varphi_{t,s})^T(x_{i,t})\, \nabla(\mathrm{div} u_s)(x_{i,s}) ds \\ p_{i,t}^w = p_{i,0}^w |d\varphi_{t,0}|(x_{i,t}) \end{cases} . \tag{23}$$

Proof. For $p_{i,t}^w = p_{i,0}^w |d\varphi_{t,0}|(x_{i,t})$, $w_{i,t} = w_i |d\varphi_t|(x_i)$, $x_{i,t} = \varphi_t(x_i)$ which implies $w_{i,t} p_{i,t}^w = \text{const}$:

$$\frac{d}{dt} w_{i,t} p_{i,t}^w \ = \ \frac{d}{dt} w_{i,0} |d\varphi_t|(x_i) p_{i,0}^w |d\varphi_t^{-1}|(x_{i,t}) = 0$$

since $(d\varphi_t^{-1})(x_{i,t}) = (d\varphi_t)^{-1}(x_i)$. Evaluating $\frac{d}{dt} w_{i,t} p_{i,t}^w = 0$ gives $p_{i,t}^w = p_{i,0}^w |d\varphi_{t,0}|(x_{i,t})$ satisfying (22):

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

Constancy $p_{i,t}^w w_{i,t} = p_{i,1}^w w_{i,1}$ with $|d\varphi_{t,1}|(x_{i,t}) = |d\varphi_1|(x_i) |d\varphi_t^{-1}|(x_{i,t})$ gives the co-state in the endpoint

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

The $p^x_{i,t}$ satisfies (22) since it is a linear inhomogeneous differential equation of the form:

$$\dot{a}_t = -(du)^T(x_{i,t})a_t - b_t \ .$$

Since $\frac{d}{dt}((d\varphi_t)^{-1}\, d\varphi_t) = 0$ gives $\frac{d}{dt}(d\varphi_t)^{-1} = -(d\varphi_t)^{-1}(d\dot{\varphi}_t)(d\varphi_t)^{-1}$; we use $d\dot{\varphi}_t = du_t \circ \varphi_t d\varphi_t$ giving

$$\frac{d}{dt}(d\varphi_t)^{-1} = -(d\varphi_t)^{-1}(d\dot{\varphi}_t)(d\varphi_t)^{-1} = -(d\varphi_t)^{-1}(du_t)\circ\varphi_t \ ,$$

which since $(d\varphi_t)^{-1} = (d\varphi_t^{-1})\circ\varphi_t$ gives our result:

$$\frac{d}{dt}\,(d\varphi_t^{-1})^T \circ \varphi_t = -(du_t)^T \circ \varphi_t\,(d\varphi_t^{-1})^T \circ \varphi_t \ .$$

Rewrite the integral solution in the initial condition using $d\varphi_{t,s} = d\varphi_s \circ \varphi_t^{-1} d\varphi_t^{-1}$:

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

which using $w_{i,t}p^w_{i,t} = \text{const}$ shows p^x_t of (23) satisfies the Hamiltonian differential equation:

$$\begin{aligned} \dot{p}^x_{i,t} &= -(du_t)^T(x_{i,t})p^x_{i,t} - p^w_{i,0}w_{i,0}\nabla(\operatorname{div} u_t)(x_{i,t}) \\ &= -(du_t)^T(x_{i,t})p^x_{i,t} - p^w_{i,t}w_{i,t}\nabla(\operatorname{div} u_t)(x_{i,t}) \ . \end{aligned}$$

□

4.4. **Method of Geodesic Shooting Based on the Measure Norm Matching Condition.** The p^x, p^w are termed the "co-state" variables as they control the velocities of the states in the Hamiltonian Method. As long as the endpoint condition for the Hamiltonian control $\boldsymbol{U}(\boldsymbol{q}_1) := \frac{\alpha}{2}\|\boldsymbol{\mu}(\boldsymbol{q}_1) - \boldsymbol{\mu}_{obs}\|^2_{\boldsymbol{W}^*}$ is smooth, then the co-states satisfy the integral equations steered by the endpoint gradients $\nabla_{\boldsymbol{q}}\boldsymbol{U}$. To explicit the gradients we construct the measure norm associating RKHSs W with smooth kernels K_W (see Appendix A for technical condition). The norm-square is defined by integrating the measure against the kernel function:

$$\|\mu\|^2_{W^*_\ell} := (\mu \mid K_{W_\ell}[\mu]) = \int_{\mathbb{R}^d\times\mathcal{F}} K_{W_\ell}[\mu](x,f)\mu(dx,df) \ .$$

The notation $(\cdot \mid \cdot)$ is called the duality bracket allowing the rewrite of the endpoint condition:

$$\boldsymbol{U}(\boldsymbol{q}_1) = \frac{1}{2}\sum_{\ell<\ell_{max}} \left(\mu(q^\ell_1) - \mu^\ell_{obs} \mid \alpha K_{W_\ell}[\mu(q^\ell_1) - \mu^\ell_{obs}]\right) \ . \tag{24}$$

This highlights the role the smoothed measure mismatch plays for every scale $\ell < \ell_{max}$ defined as

$$h_{q^\ell}(x,f) := \alpha K_{W_\ell}[\mu(q^\ell) - \mu^\ell_{obs}](x,f), \quad (x,f)\in\mathbb{R}^d\times\mathcal{F} \ . \tag{25}$$

Assume h_q is spatially smooth then the explicit formula for the gradients follow.

Statement 3. *Take h_{q^ℓ} for every scale $\ell < \ell_{max}$ as continuously differentiable in x and bounded $\mathscr{C}^{1,0}_b$. For endpoint $\boldsymbol{U}(\boldsymbol{q}_1)$ of* (24) *the gradients are given by,*

$$\begin{aligned} \nabla_{x^\ell_i}\boldsymbol{U}(\boldsymbol{q}_1) &= w^\ell_{i,1}\,\nabla_{x_i}h_{q^\ell_1}(x^\ell_{i,1}, f^\ell_i) \\ \nabla_{w^\ell_i}\boldsymbol{U}(\boldsymbol{q}_1) &= h_{q^\ell_1}(x^\ell_{i,1}, f^\ell_i) \end{aligned} \ . \tag{26}$$

Proof. The gradient requires the perturbation over all the scales simultaneously $\boldsymbol{q}(\varepsilon) = (q^\ell(\varepsilon))_{\ell<\ell_{max}}$, with $q_1^\ell(\varepsilon) = (x_{i,1}^\ell + \varepsilon\xi_i^{x,\ell}, w_{i,1}^\ell + \varepsilon\xi_i^{w,\ell})_{i\in I_\ell}$. This breaks apart for each scale independently. The dependence on scale ℓ is implied in the notation written as $q_1 = (x_{i,1}, w_{i,1})_{i\in I}$, $\mu(q_1) = \sum_{i\in I} w_{i,1}\delta_{x_{i,1}} \otimes \delta_{f_i}$:

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

□

4.4.1. *The Dense Tissue Continuum.* The tissue continuum has been studied extensively in MRI and Computational Anatomy. This continuum representation for the endpoint gradient $\nabla_{\boldsymbol{q}}\boldsymbol{U}$ is often used in image matching in its discretized form on the computational lattices. The continuum measure has $\rho(dx) = \rho_c(x)dx$ and is given according to (17). The measure is parameterized in the state $q_t := (\varphi_t(x), w_t(x))_{x\in\mathbb{R}^d}$, $q_0 = (Id, w_0 = 1)$ with $\mu(q_t) := \varphi_t \cdot \mu$ given as

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

The average of h_q over the feature space determines the boundary term variation in the dense setting.

Statement 4. *Denote $\nabla_\varphi h_{q_1}(a,b)$ as the gradient of h_{q_1} with respect to variable a, with*

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

For endpoint $\boldsymbol{U}(\boldsymbol{q}_1)$ of (24)*, the gradients are given by*

$$\begin{aligned}
\nabla_{\varphi^\ell}\boldsymbol{U}(\boldsymbol{\varphi}_1, \boldsymbol{w}_1) &= w_1^\ell\rho_c^\ell\bar{h}^\nabla_{q_1^\ell} \\
\nabla_{w^\ell}\boldsymbol{U}(\boldsymbol{\varphi}_1, \boldsymbol{w}_1) &= \rho_c^\ell\bar{h}_{q_1^\ell}
\end{aligned}\quad . \tag{27}$$

Proof. With $q_1 = (\varphi_1, w_1)$ take the variation $\varphi_1 \to \varphi_1(\varepsilon) = \varphi_1 + \varepsilon\delta\varphi$:

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

Take the variation $w_1 \rightarrow w_1(\varepsilon) = w_1 + \varepsilon\delta w$:

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

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 $k \geqslant 2$. □

5. Aknowledgement

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). 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.

Appendix A. Technical Descriptions for Geodesics and Spaces

A.1. **Technical statement for existence of solution for the geoedesic metric:** Existence of solutions can be established under a broad mathematical setting. Specifically, if $\mathscr{C}^k_0(\mathbb{R}^d, \mathbb{R}^d)$ is a space of $\mathscr{C}^k$ vector fields vanishing at infinity as well all its partial derivatives of order $p \leqslant k$, then for $G_k = (\mathrm{id} + \mathscr{C}^k_0(\mathbb{R}^d, \mathbb{R}^d)) \cap \mathrm{Diff}^1(\mathbb{R}^d, \mathbb{R}^d)$ and $V^{\ell_{max}} = \mathscr{C}^m_0(\mathbb{R}^d, \mathbb{R}^d)$ with $m \geqslant k$ the metric holds.

A.2. **Technical statement for the measure norm:** The technical condition to ensure the brain measures are elements of W^* is the kernels K_W of the RKHS W are densely and continuously embedded 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 brain varifolds are continuously embedded in the dual spaces $\mathscr{C}_b(\mathbb{R}^d \times \mathcal{F}, \mathbb{R})^*$.

Appendix B. Proof of Statement 1

Since the states (x_t, w_t), co-states (p^x_t, p^w_t) controls u_t, v_t and flows φ_t are all functions of time in the statement we suppress it in the notation.

Appendix Statement 1. *We assume that $V^{\ell_{max}} = \mathscr{C}^m_0(\mathbb{R}^d, \mathbb{R}^d)$ with $m \geqslant 2$. If $\boldsymbol{u}_\cdot$ is a solution of the optimal control problem* (20) *then there exists time-dependent co-state $(t \mapsto \boldsymbol{p}_t)$ such that for any $i \in I_\ell$ for all ℓ:*

$$\text{(28a)} \qquad \left\{ \begin{array}{rcl} \dot{q}^\ell_{i,t} & = & (u^\ell_t(x^\ell_{i,t}), w^\ell_{i,t} \operatorname{div} u^\ell_t(x^\ell_{i,t})) \\ \dot{p}^{x,\ell}_{i,t} & = & -(du_t)^T(x^\ell_{i,t}) p^{x,\ell}_{i,t} - p^{w,\ell}_{i,t} w^\ell_{i,t} \nabla(\operatorname{div} u_t)(x^\ell_{i,t}) \\ \dot{p}^{w,\ell}_{i,t} & = & -p^{w,\ell}_{i,t} \operatorname{div} u_t(x^\ell_{i,t}) \end{array} \right.$$

The optimal control satisfies $\partial_{u^\ell} \mathcal{H}(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{u}) = 0$ and $\boldsymbol{v} = \boldsymbol{A}\boldsymbol{u}$ satisfies for any $\ell_0 < \ell_{max}$ (with convention $v^{\ell_{max}} = 0$):

$$\text{(28b)} \qquad \left\{ \begin{array}{l} v^{\ell_0}(x) = u^{\ell_0}(x) - u^{\ell_0 - 1}(x) = \displaystyle\sum_{\ell_0 \leqslant \ell < \ell_{max}} \sum_{i \in I_\ell} \Big(g^{\ell_0}(x^\ell_i, x) p^{x,\ell}_i + p^{w,\ell}_i w^\ell_i \, \nabla_1 g^{\ell_0}(x^\ell_i, x) \Big) \\ u^{\ell_0}(x) = \displaystyle\sum_{\ell < \ell_{max}} \sum_{i \in I_\ell} \Big(\bar{g}^{\ell_0 \wedge \ell}(x^\ell_i, x) p^{x,\ell}_i + p^{w,\ell}_i w^\ell_i \, \nabla_1 \bar{g}^{\ell_0 \wedge \ell}(x^\ell_i, x) \Big) \, . \end{array} \right.$$

where for any $\ell < \ell_{max}$, $L^\ell : V_\ell \to V_\ell{}^*$ is the isometry such that $\langle \tilde{u}^\ell, \tilde{v}^\ell \rangle_{V_\ell} = (L^\ell \tilde{u}^\ell \mid \tilde{v}^\ell)$ for any $\tilde{u}^\ell,\ \tilde{v}^\ell \in V_\ell$ with $(\cdot \mid \cdot)$ the dual bracket and $\nabla_1 g(a, b)$ denotes the gradient of g^ℓ with respect to the first variable a.

B.1. **Proof of Eqns.** (28a)**,** (28b)**.**

Proof. Under the assumption $V^{\ell_{max}} \hookrightarrow \mathscr{C}^2_0(\mathbb{R}^d, \mathbb{R}^d)$ then we have $(\boldsymbol{u}, \boldsymbol{q}) \mapsto \boldsymbol{\xi}_{\boldsymbol{q}}(\boldsymbol{u})$ is $\mathscr{C}^2$ and standard results of optimal control theory apply establishing that $\dot{\boldsymbol{q}} = \partial_{\boldsymbol{p}} \mathcal{H}, \dot{\boldsymbol{p}} = -\partial_{\boldsymbol{q}} \mathcal{H}, \partial_{\boldsymbol{u}} \mathcal{H} = 0$. Taking the variation for $\partial_{q^\ell} \mathcal{H}$ of the Hamiltonian for scale ℓ varies $x^\ell_i(\varepsilon) = x^\ell_i + \varepsilon \xi^{x,\ell}_i$, $w^\ell_i(\varepsilon) = w^\ell_i + \varepsilon \xi^{w,\ell}_i$:

$$\text{(29)} \qquad \frac{d}{d\varepsilon} \big(p^\ell \ \big| \, \xi(q^\ell(\varepsilon), u^\ell) \big)_{|\varepsilon = 0} = \sum_{i \in I_\ell} \Big(\langle p^{x,\ell}_i, du^\ell(x^\ell_i) \xi^{x,\ell}_i \rangle + p^{w,\ell}_i w^\ell_i \langle \nabla(\operatorname{div} u^\ell)(x^\ell_i), \xi^{x,\ell}_i \rangle + p^{w,\ell}_i (\operatorname{div} u^\ell)(x^\ell_i) \xi^{w,\ell}_i \Big)$$

so that we get the last two equations of (22). To calculate $\partial_{u^\ell} \mathcal{H} = 0$, define $u^\ell(\varepsilon) = u^\ell + \varepsilon \xi$ implying $v^{\ell+1}(\varepsilon) = v^{\ell+1} - \varepsilon \xi$, $v^\ell(\varepsilon) = v^\ell + \varepsilon \xi$ for $\xi \in V_\ell$. We have

$$\text{(30)} \qquad \begin{aligned} & \frac{d}{d\varepsilon} \left(\big(p^\ell \ \big| \, \xi(q^\ell, u^\ell(\varepsilon)) \big) - \frac{1}{2} \sum_{j=\ell}^{\ell+1} \big(L^j v^j(\varepsilon) \ \big| \, v^j(\varepsilon) \big) \right)_{|\varepsilon = 0} \\ = \ & \sum_{i \in I_\ell} \Big(\langle p^{x,\ell}_i, \xi(x^\ell_i) \rangle + p^{w,\ell}_i w^\ell_i (\operatorname{div} \xi)(x^\ell_i) \Big) - \big(L^\ell v^\ell - L^{\ell+1} v^{\ell+1} \ \big| \, \xi^u \big) = 0 \, . \end{aligned}$$

After summation of (30) for $\ell \geqslant \ell_0$, we get for any $\xi \in V^{\ell_0}$ that

$$\text{(31)} \qquad (L^{\ell_0} v^{\ell_0} \mid \xi) = \sum_{\ell \geqslant \ell_0} \sum_{i \in I_\ell} \Big(\langle p^{x,\ell}_i, \xi(x^\ell_i) \rangle + p^{w,\ell}_i w^\ell_i (\operatorname{div} \xi)(x^\ell_i) \Big)$$

Now, for any $x, \alpha \in \mathbb{R}^d$, consider $\delta^\alpha_x \in V^{\ell_0}{}^*$ such that $(\delta^\alpha_x \mid \tilde{v}^\ell) = \langle \tilde{v}^\ell(x), \alpha \rangle_{\mathbb{R}^d}$ for any $\tilde{v}^{\ell_0} \in V^{\ell_0}$. The reproducing property on V^{ℓ_0} gives $\langle K^{\ell_0} \delta^\alpha_x, \ K^{\ell_0} \delta^\beta_y \rangle_{V^{\ell_0}} = g^{\ell_0}(x, y) \langle \alpha, \beta \rangle$. We get from (31) for $\xi = K^{\ell_0} \delta^\alpha_x = g^{\ell_0}(., x) \alpha$ that

$$\langle v^{\ell_0}(x), \alpha \rangle = (L^{\ell_0} v^{\ell_0} \mid K^{\ell_0} \delta^\alpha_x) = \sum_{\ell \geqslant \ell_0} \sum_{i \in I_\ell} \Big(g^{\ell_0}(x^\ell_i, x) \langle p^{x,\ell}_i, \alpha \rangle + p^{w,\ell}_i w^\ell_i \langle \nabla_1 g^{\ell_0}(x^\ell_i, x), \alpha \rangle \Big)$$

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 (16a). □

REFERENCES

[1] Special issue on computational anatomy and functional imagery. *NeuroImage*, 23, 2004.

[2] M. Albert, A. Soldan, R. Gottesman, M. G McKhann, M. N Sacktor, L. Farrington, M. Grega, R. Turner, B. Lu, S. Li, M.-C. Wang, O. Selnes, and the BIOCARD Research Team. Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to apoe genotype. *Current Alzheimer Research*, 11(8):773âĂŞ784, 2014.

[3] W. Allard. On the first variation of a varifold. *Ann. of Math.*, 95:417–491, 1972.

[4] J. Ashburner. Computational anatomy with the spm software. *Magnetic Resonance Imaging*, 27:1163–1174, October 2009.

[5] 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*, 61(2):139–157, 2005.

[6] E. D. Boer and P. Kuyper. Triggered correlation. *IEEE Transactions on Bio-Medical Engineering*, 15(3), July 1968.

[7] N. Charon and A. Trouvé. The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580, 2013.

[8] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*, 348(6233), 2015.

[9] 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*, 54(1):80–89, 2011.

[10] 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*, 96(3):542 – 557, 2017.

[11] 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*, 80:62 – 79, 2013. Mapping the Connectome.

[12] A. Evans, D. Collins, S. Mills, E. Brown, R. Kelly, and T. Peters. 3d statistical neuroanatomical models from 305 mri volumes. page 1813âĂŞ1817, 1993.

[13] 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*, 13:241âĂŞ244, 2016.

[14] H. Federer. *Geometric measure theory*. Number 153. Springer, 1969.

[15] B. Fischl, D. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–55, January 2002.

[16] J. F.J. Almgren and J. Taylor. The geometry of soap bubbles and soap films. *Scientific American*, pages 82–93, July 1976.

[17] . K. N. Y. GERSTEIN, G. L. An approach to the quantitative analysis of electrophysiological data from single neurons. *Biophysical journal*, 1(1):15âĂŞ28, 1960.

[18] G. GR. A theorem about random fields. *Bulletin of the London Mathematical Society*, 5(1):81–84, 1973.

[19] U. Grenander. *General Pattern Theory*. Number ISBN 978-0198536710. Oxford Science Publications, 1994.

[20] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, 56(4):617–694, 1998.

[21] 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.*, 21:101617, 2019.

[22] J. W. Lewis and D. C. Van Essen. Mapping of architectonic subdivisions in the macaque monkey, with emphasis on parieto-occipital cortex. *J Comp Neurol*, 428(1):79–111, 2000.

[23] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, D. L. Collins, P. Thompson, M. MacDonald, D. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, , B. Kawashima, and R. Mazoyer. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Phil. Trans. R. Soc. Lond. B Biol. Sci.*, 356:1293–1322, 2001.

[24] 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*, 2:89–101, 1995.

[25] M. Miller and L. Younes. Group Actions, Homeomorphisms, and Matching: A General Framework. *International Journal of Computer Vision*, 41(4):61–84, 2001.

[26] 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*, 3:352–360, 12 2013.

[27] M. I. Miller, S. Arguillere, D. J. Tward, and L. Younes. Computational anatomy and diffeomorphometry: A dynamical systems model of neuroanatomy in the soft condensed matter continuum. pages 1–24, 2018.

[28] 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*, (17):447–509, November 4 2015.

[29] M. I. Miller, D. J. Tward, and A. TrouvÃl'. Coarse-to-fine hamiltonian dynamics of hierarchical flows in computational anatomy. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3760–3765, 2020.

[30] 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*, 362(6416), 2018.

[31] S. Mori, S. Wakana, L. Nagae-Poetscher, and P. C. V. Zijl. Elsevier, 2005.

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

[33] I. Peikon, J. Kebschull, V. Vagin, D. Ravens, Y. Sun, E. Brouzes, I. J. CorrÃła, D. Bressan, and A. Zador. Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Res.*, 12(45), July 7 2017.

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

[35] S. G. Rodriques, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, Vanderburg, C. R., J. Welch, L. M. Chen, F. Chen, and E. Z. Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

[36] N. Sacktor, A. Soldan, M. Grega, F. L, Q. Cai, M. Wang, R. Gottesman, R. Turner, A. M, and B. R. Team. The biocard index: A summary measure to predict onset of mild cognitive impairment. alzheimer dis assoc disord. *Alzheimer Disease Association Disord*, 31(2):114–119, Apr-Jun 2017.

[37] R. Satij, J. A. Farrell, D. Gennert, F. Schier, Alexander, and A. Regev. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, 33, 2015.

[38] 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*, 92:342–357, 10 2016.

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

[40] G. J. Srulovicz P. A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *J Acoust Soc Am.*, 4(73):1266–76, April 1983.

[41] 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*, 353(6294):78–82, 2016.

[42] 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.*, D996âĂŞD1008(41(Database issue)), Jan 2013.

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

[44] P. M. Thompson and A. W. Toga. A framework for computational anatomy. *Computing and Visualization in Science*, 5(1):13–34, 2002.

[45] 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*.

[46] D. Tward, X. Li, B. Huo, B. Lee, P. Mitra, and M. Miller. 3d mapping of serial histology sections with anomalies using a novel robust deformable registration algorithm. volume LNCS, volume 11846, October 10 2019.

[47] Q. Wang, S.-L. Ding, Y. Li, J. Royall, D. Feng, P. Lesnar, N. Graddis, M. Naeemi, B. Facer, A. Ho, et al. The allen mouse brain common coordinate framework: a 3d reference atlas. *Cell*, 181(4):936–953, 2020.

[48] 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.

[49] W. WM, G. WL, K. R, and J. FA. Adaptive segmentation of mri data. *IEEE Trans Med Imaging*, 4(15):429–42, 1996.

[50] 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*, 116(39):19490–19499, 2019.

[51] A. Xiaowei. Method of the year 2020: spatially resolved transcriptomics. *Nature Methods*, 18(1), 2021.

[52] Z. Y, B. M, and S. S. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging*, 20(1):45–57, January 2001.

[53] L. Younes, M. Albert, and M. I. Miller. Inferring changepoint times of medial temporal lobe morphometric change in preclinical alzheimer's disease. *NeuroImage: Clinical*, 5:178 – 187, 2014.

[54] L. Younes, M. Albert, A. Moghekar, A. Soldan, C. Pettigrew, and M. I. Miller. Identifying changepoints in biomarkers during the preclinical phase of alzheimerâĂŹs disease. *Frontiers in Aging Neuroscience*, 11:74, 2019.

[55] S. M. Young ED. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, (66):1381–1403.

[56] E. Zhao, M. R. Stone, X. Ren, T. Pulliam, P. Nghiem, J. H. Bielas, and R. Gottardo. Bayesspace enables the robust characterization of spatial gene expression architecture in tissue sections at increased resolution. *bioRxiv*, 2020.

CENTER OF IMAGING SCIENCE AND & DEPARTMENT OF BIOMEDICAL ENGINEERING, THE JOHN HOPKINS UNIVERSITY, 301 CLARK HALL, BALTIMORE, MD 21218

E-mail address, M. I. Miller: mim@cis.jhu.edu

UNIVERSITY OF CALIFORNIA AT LOS ANGELES

E-mail address, D. Tward: dtward@mednet....

ECOLE NORMALE SUPÉRIEURE PARIS-SACLAY, UNIVERSITÉ PARIS-SACLAY, 4 RUE DES SCIENCES, F-91190 GIF SUR YVETTE, FRANCE

E-mail address, A. Trouvé: alain.trouve@ens-paris-saclay.fr

CENTER OF IMAGING SCIENCE & DEPARTMENT OF APPLIED MATHEMATICS, THE JOHN HOPKINS UNIVERSITY, 301 CLARK HALL, BALTIMORE, MD 21218