

A multimodal approach for visualizing and identifying electrophysiological cell types in vivo

Received: 29 September 2025

Accepted: 19 March 2026

Cite this article as: Lee, E.K., Gül, A.E., Heller, G. *et al.* A multimodal approach for visualizing and identifying electrophysiological cell types in vivo. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-71331-0>

Eric Kenji Lee, Asım E. Gül, Gregory Heller, Anna Lakunina, Han Yu, Andrew Shelton, Shawn Olsen, Nicholas A. Steinmetz, Cole Hurwitz, Santiago Jaramillo, Pawel F. Przytycki & Chandramouli Chandrasekaran

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A multimodal approach for visualizing and identifying electrophysiological cell types *in vivo*

Eric Kenji Lee¹, Asım E. Gül², Gregory Heller³, Anna Lakunina^{4,5,6}, Han Yu^{7,8}, Andrew Shelton⁶, Shawn Olsen⁶, Nicholas A. Steinmetz⁹, Cole Hurwitz⁸, Santiago Jaramillo^{4,5}, Pawel F. Przytycki¹⁰, and Chandramouli Chandrasekaran^{*1,11,12,13}

¹*Department of Psychological and Brain Sciences, Boston University, Boston, MA 02115, USA*

²*Department of Psychology, Boğaziçi University, Beşiktaş, Istanbul 34342, Türkiye*

³*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

⁴*Department of Biology, University of Oregon, Eugene, OR 97403, USA*

⁵*Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA*

⁶*Electrophysiology Group, Allen Institute for Neural Dynamics, Seattle, WA 98109, USA*

⁷*Department of Electrical Engineering, Columbia University, New York City, NY 10027, USA*

⁸*Zuckerman Institute, Columbia University, New York City, NY 10027, USA*

⁹*Department of Neurobiology and Biophysics, University of Washington, Seattle, WA 98195, USA*

¹⁰*Faculty of Computing & Data Sciences, Boston University, Boston, MA 02215, USA*

¹¹*Center for Systems Neuroscience, Boston University, Boston, MA 02115, USA*

¹²*Department of Biomedical Engineering, Boston University, Boston, MA 02115, USA*

¹³*Department of Anatomy & Neurobiology, Boston University, Boston, MA 02118, USA*

March 18, 2026

Abstract

Neurons of different types perform diverse computations and coordinate their activity during sensation, perception, and action. While electrophysiological recordings can measure the activity of many neurons simultaneously, identifying cell types during these experiments remains difficult. Here we present PhysMAP, a framework adapted from multiomics data analysis that weights multiple electrophysiological modalities simultaneously to obtain interpretable multimodal representations. We apply PhysMAP to seven datasets and demonstrate that these multimodal representations are better aligned with known transcriptomically-defined cell types than any single modality alone. We then show that this alignment allows PhysMAP to better identify putative cell types in the absence of ground truth. We also demonstrate how annotated datasets can transfer labels to unannotated recordings and confirm that inferred cell types exhibit properties consistent with ground truth. Crucially, we show that PhysMAP can also be used to iterably detect batch effects which confound classification. Together, these results establish PhysMAP as a tool for studying multiple cell types simultaneously and gaining insight into neural circuit dynamics.

Keywords: Cell types, Electrophysiology, optotagging, Multimodal Analysis, Dimensionality Reduction

*Corresponding Author (cchandr1@bu.edu)

Introduction

Single-cell transcriptomics, electrophysiology, and volumetric imaging have identified numerous clusters defined by common gene expression, patterns of activity, and morphologies all broadly referred to as neuronal “cell types” in the brain¹⁻⁴. These cell types form microcircuits that perform neural computations and drive behavior⁵⁻⁷. For instance, in mice, interactions between excitatory pyramidal, parvalbumin-positive (PV⁺), somatostatin-positive (SOM⁺), and vasoactive intestinal peptide-positive (VIP⁺) neurons enable computations e.g., gain modulation⁸ and selective enhancement of salient visual stimuli⁹. Despite these advances, most experiments can monitor only one or two cell types simultaneously; this precludes a full understanding of how cell types contribute to the brain’s many computations¹⁰. Current techniques are infeasible or technically challenging in species other than mice and in brain regions outside the neocortex owing to a lack of effective viral constructs^{11;12} limited optical access¹³⁻¹⁵. These constraints prevent the study of cortical cell types in non-human primates and humans (whose cortex is mostly too deep to access optically) and of any cell types in the subcortical structures of any animal.

In contrast, high-density extracellular electrophysiological recordings can simultaneously sample from numerous neurons throughout the brain without genetic or optical access¹⁶⁻²⁰, including in humans^{21;22}. Yet the cellular identities of recorded neurons remain largely invisible because the correspondence between cell types and their *in vivo* electrophysiological properties remains poorly understood, limiting our understanding of circuit dynamics. An analysis approach that delineates cell types within electrophysiological recordings would address this limitation. Such an approach should have the following capabilities: First, it should maximally incorporate electrophysiological information across multiple modalities to identify cell types. Waveform shape²³⁻²⁶, inter-spike interval (ISI) distributions²⁷⁻²⁹, autocorrelogram (ACG)³⁰, and peri-stimulus time histograms (PSTHs)³¹⁻³³ are all informative about cell type. However, which modalities to combine, how to best combine them, and under what conditions they are valid across experiments are not well understood. Second, “cell types” are not discrete, as they often lie along a continuum of function³⁴. Thus, any approach to delineate cell types from electrophysiological data should be able to visualize and explore heterogeneity in an interpretable manner (e.g., how biological features vary within a cell type). Third, when cell type identity is available, representations should logically cluster according to ground truth, providing confidence that similar structure can identify cell types in datasets lacking ground truth. Fourth, the approach should transfer labels from annotated to unannotated datasets (i.e., generalize) after detecting and addressing batch effects. Finally, this method should excel at all these tasks with the computational efficiency needed for rapid iteration across modalities and datasets.

Here we present “PhysMAP,” an approach satisfying all of these requirements. PhysMAP uses Seurat³⁵ and adapts advances in multimodal representation learning from computational biology^{36;37} to electrophysiological data. PhysMAP combines non-linear dimensionality reduction³⁸ with a weighted-nearest neighbor graph (WNN)³⁵, weighting multiple electrophysiological modalities on a per-unit basis to create an interpretable multimodal representation for unsupervised cell type identification. We apply PhysMAP to seven datasets from different neuroscience labs and show that it: 1) combines modalities and identifies cell types when ground truth is available, 2) explores and clusters putative cell types without annotations, 3) identifies cell types in partially-annotated datasets, 4) detects batch effects, and 5) runs on a personal computer in minutes. To demonstrate practical utility, we use a high-quality ground-truth dataset to identify cell types in a dataset with incomplete ground truth and assess the properties of several cell types interacting simultaneously during behavior. PhysMAP thus provides a versatile approach for identifying cell types in electrophysiological recordings, advancing our ability to study the circuit dynamics of multiple neuronal types during behavior.

Results

PhysMAP combines multiple electrophysiological modalities to uncover cell type and laminar structure

Ex vivo experiments have shown that the relationship between cell type and electrophysiological signature spans multiple properties^{2,39,40}, suggesting that cell types could be identified from *in vivo* extracellular recordings. We examined our first desired property of a multimodal tool: whether combining multiple electrophysiological modalities could better align with cell types than any single modality. PhysMAP unifies different modalities by linearly combining high-dimensional UMAP graphs using a weighted nearest neighbor approach³⁵ on an arbitrary number of modalities⁴¹ (see Supplementary Note 1 and Fig. S1 for details).

We applied PhysMAP to *in vivo* juxtacellular recordings from the mouse somatosensory cortex, which provided low-noise somatic waveforms and ground truth cell type information ($n = 246$) from eight cell types³². UMAP applied to waveform shape^{25,42} produced 2D embeddings that aligned well with underlying cell types and laminar structure (Fig. 1A), though some cell types (like layer 4 excitatory cells [E-4]) appeared split. In contrast, this split disappeared in the ISI modality (Fig. 1B), but there was less clear delineation between PV⁺ and SOM⁺ cells. However, E-4 cells were well separated from E-5 (layer 5 excitatory cells) and other cell types. These complementary views suggest that optimally combining multiple modalities could provide more robust cell type delineation, as unsupervised multimodal approaches benefit when signals correlate across modalities while noise remains uncorrelated⁴³.

We examined how PhysMAP weights each modality per-neuron (Fig. 1C). For most units, waveform shape was most informative (58% vs. 42% for ISI)²⁹, though certain regions of the projection showed greater ISI importance. The multimodal representation obtained by weighting each modality per-neuron appeared to best separate cell types visually (Fig. 1D). This visualization revealed that PV⁺ cells cluster tightly and segregate by layer, while SOM⁺ cells occupy a larger area, perhaps a result of underlying physiological diversity⁴⁴⁻⁴⁶.

To test whether PhysMAP's improvements simply resulted from including more information, we applied UMAP to concatenated waveform shape and ISI features (unweighted combination of all modalities; Fig. 1E). The resulting representation showed a weaker structure than PhysMAP (compare Fig. 1E to Fig. 1D), with poorer clustering of E-5, SOM⁺, and E-4 neurons. These results suggest PhysMAP's weighted averaging approach better utilizes multiple modalities.

We also replicated this analysis with three modalities by adding peri-stimulus time histograms (PSTHs) aligned to whisker deflection (Fig. S2A, B, and C, respectively). PSTH could also separate cell types, but was not the most informative modality for any units (Fig. S2D). Again, the three-modality PhysMAP representation showed better clustering of PV⁺ and E-4 cells visually (Fig. S2E vs. in Fig. 1D) while the concatenated three-modality representation remained poorly clustered (Fig. S2F) again reinforcing that naïve combination doesn't improve representation quality. Since PSTH contributed minimally and its inclusion would preclude assessment of other functional properties due to analytic circularity⁴⁷, we excluded it from subsequent analyses.

These results show that PhysMAP satisfies the first requirement of an ideal tool: it can combine multiple modalities to obtain a representation that is superior to any single modality.

Interpretable representations from PhysMAP enable biological insight

Even when cell types have been identified through optotagging, they often encompass multiple subclasses⁴⁴; physiological variation within a class is the norm³⁴. Thus, a second requirement is the ability to visualize physiological diversity within and across defined types.

PhysMAP's visualizations fulfill this requirement. Fig. 2A shows the PhysMAP plot from Fig. 1D with marker radius scaled to reflect waveform spike width, peak-to-trough ratio, and onset latency. Examining spike width (Fig. 2A, left), clear patterns emerge: PV⁺ cells are uniformly narrow regardless of layer; SOM⁺ cells span a gradient of widths overlapping with E-5 cells at their widest; and E-4 cells have the widest spikes. This heterogeneity in SOM⁺ cells likely reflects several subtypes with varying widths⁴⁸ and distinct computational roles⁴⁵. Waveform peak-to-trough ratios also differ between cell types (Fig. 2A, center) with PV⁺ and SOM⁺ cells having the smallest ratios and both E-4 and E-5 cells occupying the largest ratios. Firing rate onset latency (Fig. 2A, right) is fastest for PV⁺ cells and slowest for SOM⁺ cells, consistent with³².

These interpretations also held when the trial-averaged firing rate response to a whisker deflection (peristimulus time histogram, PSTH) was included (Fig. S5A). We note that PSTH was excluded here to avoid analytic circularity⁴⁷ as, in the next section, we examine the PSTH of each unsupervised cluster.

These insights are not apparent when examining metrics without the PhysMAP representation. A scatter plot of spike width versus peak-to-trough ratio (Fig. S3) shows some structure but appears as one continuous relationship, failing to clearly separate SOM⁺, PV⁺, and excitatory cells. Histograms of these metrics show considerable overlap between cell types (Fig. S3B).

Thus, PhysMAP satisfies the second property of an ideal tool as it provides an interpretable visualization that effectively combines multimodal electrophysiological data, facilitating exploration of physiological diversity while maintaining the relationship to underlying cell types.

Combining multiple modalities improves unsupervised clustering

Annotations are often unavailable, as in electrophysiological datasets without optotagging. In such cases it is desirable to uncover cell types via unsupervised methods^{1;2;21;23;24;26;29;49}. Thus, a third requirement is that unsupervised clustering should align with ground truth cell types. To do this, we evaluated whether PhysMAP representations align with underlying molecular cell types in a fully unsupervised analysis, asking the question: "without access to ground truth, how well do unsupervised methods identify cell types?"

We assessed PhysMAP-to-ground-truth alignment using the modified adjusted Rand index (MARI;⁵⁰), to estimate if ground truth cell types align with individual unsupervised clusters. We applied Leiden clustering to the high-dimensional graphs produced by PhysMAP and each individual modality and then calculated MARI scores between clustering results and ground truth labels across a range of resolution parameters (cluster size priors; see Methods: MARI Calculation for details).

Fig. 2B shows neurons in PhysMAP colored by cell type (left) alongside a sample Leiden clustering with resolution set to 2 (right) which produced the same number of clusters as prominent cell types (greater than 10 samples). PhysMAP produced higher MARI scores than any individual modality or unweighted combination across all Leiden resolutions (Fig. 2C), indicating superior unsupervised capture of cell type structures at all levels of granularity (see Fig. S4A for clustering at other resolutions). This result also held when PSTH was included as a modality (Fig. S5B) and was not sensitive to changing the

neighborhood size used by UMAP (Fig. S4B). Choosing a resolution of 2 (the maximum value before MARI dropped), six clusters were produced with varying proportions of cell types within each (Fig. 2D).

We found that while not all clusters captured single cell types, many showed high specificity: E-4 cells dominated cluster 1, PV⁺ cells (especially PV-4) dominated cluster 2, and SOM⁺ cells were substantially represented in cluster 4. With PSTH included, clusters were also similar in composition to the two modality case, ostensibly because PSTH contributed little to overall clustering (Fig. S5C).

To confirm this advantage was not a peculiar feature of Leiden clustering, we also applied a Gaussian mixture model (GMM) to both individual modalities and their unweighted concatenation (Fig. S5D). The GMM was applied to either 30-dimensional PC space for individual modalities or the 30-dimensional multimodal embedding for PhysMAP. Across all comparisons, PhysMAP's weighted multimodal representation achieved higher MARI scores than GMM clustering on individual modalities or unweighted combinations, indicating better capture of underlying ground truth cell types (Fig. S5C, D).

To assess the functional properties of these unsupervised clusters, we examined their responses to whisker deflection stimuli (Fig. 2E). Several notable features aligned with previous findings in the literature: cluster 2 (predominantly PV⁺) showed large, low-latency (peak at 7 ms) firing rate increases, matching properties of PV⁺ cells reported by³²; cluster 1 (predominantly E-4) responded with similar latency but much smaller firing rate changes consistent with excitatory cells receiving direct thalamo-cortical input; and cluster 3 (containing many SOM⁺ cells) was the slowest responding (peak at 16 ms), consistent with previous characterizations of SOM⁺ cells. While SOM⁺ cells were split among several clusters, this is unsurprising given their diversity^{44,45}. Traditional waveform metrics similarly struggled to cluster SOM⁺ cells, as shown by their distribution split across separate GMM clusters (Fig. S3B and C).

These results show that PhysMAP has another property desired of an ideal tool—provide unsupervised clustering methods an excellent opportunity to capture underlying cell types, even when annotations are unavailable.

PhysMAP delineates cell types better than any modality alone or in unweighted combination

The previous section showed that PhysMAP's multimodal representations align better with underlying cell types from an unsupervised perspective. However, there are also datasets that contain cell type information; do representations continue to identify these cell types assessed via a classifier trained on PhysMAP's structure? We now examine whether these advantages persist in a supervised analysis by training a classification algorithm on PhysMAP's representation and seeing if it outperformed other classifiers trained on individual modalities or in unweighted combinations. In this way, we examine PhysMAP's performance on datasets that both completely lack ground truth (as in the previous section) or are also fully characterized.

We trained classifiers on high-dimensional graphs to identify cell types using individual modality representations, unweighted concatenation, and PhysMAP. For comparison, we also trained classifiers on derived electrophysiological metrics and raw data without dimensionality reduction. Using a 15-85% test-train split, we trained gradient-boosted tree models (GBM) with five-fold cross-validation on high-dimensional embeddings to identify the five cell types with over 10 examples each (so that at least one data point from each class could be included in the test set; see Methods: Classifier Analysis for details).

PhysMAP matched or exceeded the performance of all other modalities, whether individual or in unweighted combination, across all cell types (Fig. 2F). Waveform shape performed nearly as well as

PhysMAP for PV-4, E-4, and SOM⁺ cells, confirming our hypothesis that waveform shape is the most cell type-informative modality (Fig. 1C) and validating our previous findings with WaveMAP^{24,25}. The benefit of additional modalities was most evident in improved classification of E-5 and PV-5 neurons.

PhysMAP also demonstrated better classification performance than the modalities in an unweighted combination (concatenated) for all cell types (Fig. S5E, teal line). PhysMAP and waveform shape far outperformed traditional derived waveform metrics (spike width and peak-to-trough ratio; Fig. S5E, gray dashed line), confirming that hand-derived metrics lose important cell type-distinguishing information. Notably, PhysMAP matched or performed better than classifiers trained on raw data, despite the dimensionality reduction step: the GBM classifier trained directly on full-dimensional data from all modalities underperformed both PhysMAP and waveform shape alone (Fig. 1C, black dashed line). While seemingly counterintuitive, this reflects how dimensionality reduction helps “denoise” latent clusters in smaller datasets. Moreover, training directly on raw data sacrifices interpretability, providing only class labels and probabilities without the visualization benefits of PhysMAP.

These results were robust to choices of input dimensionality and classifier type. Classification accuracy patterns remained consistent across embedding dimensions from high (50-dim.) down to low (5-dim.; Fig. S6A) and across classifier types (Fig. S6B). This preservation of performance trends is likely because the input modalities are low dimensional. Principal component analysis showed that the intrinsic dimensionality of each dataset was relatively low, with “elbows” in the scree plots at the third, sixth, and third PC’s representing 91%, 90%, and 93% of total variance for waveform shape, ISI distribution, and PSTH respectively (Fig. S7).

Thus, PhysMAP satisfies our third property of an ideal cell type identification tool as it intelligently combines multiple modalities to draw out cell type structure better than any modality alone or in unweighted combination, while providing valuable visualizations for exploring physiological diversity and maintaining high classification performance.

PhysMAP identifies cell types in extracellular electrophysiological recordings

The juxtacellular recordings analyzed above represent an ideal scenario: unattenuated somatic waveforms, unambiguous single-unit isolation, and preferential targeting of specific cell types. We next asked whether PhysMAP could identify cell types in extracellular probe recordings, which introduce challenges including slow-pass signal filtering by neuropil, poorer single-unit isolation, and non-somatic waveforms—all contributing to increased ambiguity in cell type properties. We analyzed an extracellular dataset from mouse primary auditory cortex with optotagged cells.

In this study, researchers recorded from neurons in mouse primary auditory cortex ($n = 373$) with extracellular probes while mice were presented with auditory stimuli. They identified both SOM⁺ and PV⁺ units via optogenetic tagging and recorded each unit’s waveform shape and ISI distribution. When PhysMAP was applied to this data, visualization shows clear separation of both SOM⁺ and PV⁺ cell types (purple and teal, respectively) from untagged, presumably excitatory cells (gray; Fig. 3A).

We quantified this separation by training a classifier to identify each cell type using PhysMAP and compared it against classifiers trained on individual modalities or waveform metrics. PhysMAP outperformed ISI distributions and waveform metrics but matched performance on waveform shape and raw data for PV⁺, SOM⁺, and excitatory cells (Fig. 3B). This equivalence to waveform shape likely occurs because this modality is highly informative of underlying cell types, with median per-unit weights of 0.71 for

waveform shape versus 0.29 for ISI distribution. This is a demonstration of PhysMAP's ability to focus only on informative modalities and this prioritization of waveform shape is evident in the visualizations: PhysMAP's projection closely resembles the waveform shape visualization (Fig. 3C), while the same for ISI distribution shows far less structure (Fig. 3D).

The PhysMAP approach also revealed interpretable cell type heterogeneity. We identified two sub-populations of SOM⁺ cells not documented in the original study: wide-spiking SOM⁺ (WS-SOM⁺, Fig. 3A, left) and narrow-spiking SOM⁺ (NS-SOM⁺, Fig. 3A, lower right) cells, which occupied separate regions of the embedding. These sub-populations align with previous observations of SOM⁺ subtypes with specific synaptic targeting⁴⁵ and differential behavioral functions^{51;52}. Importantly, these SOM⁺ sub-populations would be difficult to identify using spike width alone, as SOM⁺ cells as a whole form a continuum on this metric (Fig. 3D, purple) that overlaps with both PV⁺ cells (Fig. 3D, teal) and numerous untagged cells, presumably including excitatory neurons (Fig. 3E). PhysMAP's multimodal approach disambiguates these SOM⁺ subtypes while simultaneously distinguishing PV⁺ and putative excitatory cells.

This ability to optimally identify underlying cell types extracellularly also held for a separate heterogeneous dataset from CellExplorer⁵³ spanning multiple brain areas. Here, the multimodal approach again separated cell types better than unimodal visualizations (Fig. S8A to E) and demonstrated superior performance from both supervised classification (Fig. S8F) and unsupervised clustering perspectives (Fig. S8G).

These results establish that PhysMAP can identify cell types in extracellular recordings by effectively leveraging informative modalities while discounting less reliable ones.

PhysMAP identifies inhibitory cell types with accuracy greater than or equal to other unsupervised methods

Targeted optogenetic experiments like those in the previous section offer ideal conditions for cell type identification: during recordings, single electrodes are advanced until optically-responsive cells are found and isolated. However, in most electrophysiological recordings, electrodes take an unbiased sampling of all neurons. Typical recordings also capture neurons with poorer isolation, at variable distances to somata, and more variable light intensities than targeted tagging experiments. Cell identification in such settings faces additional challenges: opsin-expressing cells can show diverse responses including "paradoxical suppression"⁵⁴, and non-expressing cells may show excitation due to polysynaptic effects. In order to demonstrate whether or not PhysMAP satisfies the fourth property of an ideal tool (identifying cell types in real use cases), we must first examine how it performs when optotagged cells aren't specifically searched for and enriched within a dataset. We demonstrate that despite these challenges, PhysMAP outperforms single modalities and other unsupervised methods in identifying cell type structures without optogenetic targeting, while requiring less computation time and minimal architecture tuning.

We applied PhysMAP to a large dataset of optotagged inhibitory neurons collected using Neuropixels Ultra probes from mouse visual cortex⁵⁵, containing waveforms and 3D-autocorrelograms (ACGs) from 8,953 putative single units including 237 PV⁺, 107 SOM⁺, and 118 VIP⁺ cells. PhysMAP's embedding (Fig. 4A, top left) clearly separated PV⁺ cells (teal) from SOM⁺ cells (purple), while VIP⁺ cells (pink) were scattered among untagged presumptive excitatory cells (gray). In contrast, the 3D-ACG visualization showed poor clustering (Fig. 4A, top right) but waveform shape clustering was ostensibly comparable (Fig. 4A, bottom).

A GBM classifier trained on PhysMAP's multimodal representation outperformed both unimodal representations of 3D-ACG and waveform shape across all three optotagged cell types (Fig. 4B; $p < 0.001$ and $p = 0.016$ respectively, two-sided Friedman test). PhysMAP also outperformed classifiers trained directly on the raw data for most cell types (80%, 64%, and 71% for PV⁺, SOM⁺, and VIP⁺, respectively; $p < 0.001$, two-sided Friedman test), demonstrating that its cell type-informative structure isn't merely an artifact of optogenetic targeting.

We also compared our approach against a β -variational autoencoder (VAE) method, representative of current unsupervised deep learning approaches. Although the VAE successfully reconstructed both waveform shapes and 3D-ACGs (Fig. S9A) and produced sensible embedding structures (Fig. S9B), a classifier trained on its embedding failed to outperform the GBM classifier trained on PhysMAP (Fig. 4B, bottom right; $p < 0.001$, two-sided Friedman test). It is important to note that we used a single VAE architecture trained for each modality separately (Fig. S10), in contrast to extant ensemble approaches with multiple VAEs^{30;56}, as the latter would be computationally infeasible to compare.

Furthermore, although these alternative identification approaches are often perceived as powerful (however, compare Fig. S11B, left to Supp. Fig. 6G of³⁰), they are not ideal for the use case we intend for PhysMAP of being an iterable tool to quickly explore different datasets across several modalities: the ensemble approach would have entailed training ten separate models for each unit in the dataset. We find that PhysMAP is computationally more efficient. The β -VAE pair needed an expensive, high-performance GPU (NVIDIA L40S GPU; 500 epochs, 81 dimensions; 3.2 minutes) to achieve comparable training times to PhysMAP on a laptop (Apple M2 CPU; 500 epochs, 81 dimensions; 3.0 minutes) at inferior performance.

Thus, PhysMAP provides superior multimodal representations with significantly less computational overhead compared to deep learning methods, making it ideally suited to data exploration and iteration across many modalities and datasets.

Multimodal visualization identifies batch effects

We have demonstrated that PhysMAP can uncover cell types within single datasets. However, for experimental utility, PhysMAP must consistently identify cell types across different animals and conditions. In real-world use cases, distributional shifts between datasets can alter the relationship between cell type and electrophysiological properties, impeding transfer learning. These shifts most commonly arise from lab-specific differences in data processing or experimental procedures⁵⁷. Critically, different modalities may be differentially affected: different signal processing filters can alter waveform shape while preserving spike timing patterns (see Fig. 5 of⁴²); conversely, when comparing recordings from two behavioral paradigms, waveform shape may be conserved while spike timing statistics change.

The danger of batch effects for cell type identification arises when a classifier is applied to data in which the feature-label relationship differs from training. This danger increases with classifiers that output predictions without exposing underlying distributional alignments. Thus, any cell type identification tool must detect batch effects and assess which modalities to include^{58;59}. PhysMAP's multimodal visualization provides exactly this capability.

The danger of batch effects in ACG was something we experienced firsthand: anchor alignment⁶⁰ (used in a previous iteration of PhysMAP) dramatically reduced classification accuracy by introducing random distributional shifts in noisy modalities (in this case, ACG; Fig. S11A). These effects reduced classifier performance by, on average, 22% across the three cell types investigated. This negative effect also appeared when PhysMAP was applied to cerebellar cell types (Fig. S11B;^{30;61}).

Batch effects can be equally dangerous for waveform shape. To demonstrate this, we took half of the Ultras dataset and filtered the waveforms with either a 250 Hz or 500 Hz high-pass 4th-order Butterworth filter. Visualizing this modality with PhysMAP at the 250 Hz cutoff, we easily observed the presence of a batch effect as shown in the “mirroring” of broad- and some narrow-spiking waveforms (Fig. S12A). While the waveforms split according to cell type were nearly identical before and after filtration (Fig. S12B), a classifier trained on the unfiltered waveforms but tested on the filtered ones failed to correctly classify any VIP⁺ cells (Fig. S12C): the classifier falsely identified nearly every VIP⁺ cell as a SOM⁺ cell. The deleterious effect on classification increased with a higher cutoff frequency and commensurately larger batch effect. Visualizing the waveforms filtered at the 500 Hz cutoff, the two sets of waveforms were completely separated (Fig. S12D). Again, although the waveforms themselves did not appear much different (Fig. S12E), classification performance was reduced by 10 to 15% for PV⁺ and SOM⁺ cell types. The classifier for VIP⁺ cells again failed entirely.

The preceding examples used controlled manipulations to demonstrate how batch effects corrupt classification. In practice, however, batch effects are not known ahead of time and thus the challenge is in identifying *which* modalities are affected when combining independently collected datasets. To demonstrate PhysMAP’s ability to diagnose batch effects and assess modalities for inclusion, we examined two independently collected large-scale datasets: the previously investigated Ultras dataset and the Visual Behavior (VB) dataset⁶². To assess distributional alignment between these two datasets, we jointly projected them into PhysMAP’s multimodal embedding using normalized waveform shape and 3D-ACG (Fig. 5A), color-coding by dataset (Ultras in green, VB in yellow). While some units overlapped (gold units within the green cluster at upper left), most were completely separable—a batch effect made visible by PhysMAP’s unsupervised approach. Examining individual modalities (Fig. 5B), waveform shapes largely coincided between datasets (Fig. 5B, top), with minor differences: Ultras waveforms concentrated at borders relative to VB, and some Ultras units formed a separate cluster (middle right) composed of tri-phasic waveforms. These waveforms consisted of large pre-hyperpolarization peaks—likely waveforms from neurites that are visible only on very high-density probes⁵⁵. In contrast, 3D-ACGs showed complete separation between datasets (Fig. 5B, bottom), likely because Ultras ACGs contained no spikes at certain timescales due to masking during optotagging stimulation. Although both datasets used matched window and bin sizes for ACG calculation, and the VB dataset excluded spiked times from during the opto-tagging epochs, other sources of batch effects may have persisted. For instance, the stimulus presentation pattern and frequency in the VB dataset may have systematically influenced spike timing patterns in ways absent from the Ultras dataset that lacked stimuli. Similarly, in the Ultras dataset, the lack of a long contiguous activity periods meant that spikes were drawn from inter-trial intervals; in addition, the optotagging stimulation may have induced longer timescale effects persisting into the inter-trial epoch via second-order effects⁵⁴. Ideally, both experiments would contain either spikes from a “spontaneous activity” period or with matched visual stimuli to enable integration, a point we return to in the Discussion. This real-world use case underscores that batch effects in spike timing modalities can arise from differences in experimental design but can be detected by PhysMAP. Thus, PhysMAP’s visualization-first approach represents a key advantage over purely classifier-based approaches that would have failed silently and led to erroneous conclusions within downstream analyses.

Note, while this dataset may have been reduced to just one modality, we do not mean to imply that any two experiments will not align on modalities such as 3D-ACG. The VB dataset is itself also divided into two experiments, each following the same protocol but using different natural image sets. These two datasets were indistinguishable both in the multimodal embedding space (Fig. 5C) and within each individual constituent modality (Fig. 5D). We emphasize that there are no extant datasets that we know of that have high-quality optotagging of all major inhibitory cell types during behavior, so while

our analysis reduces to the unimodal case when examining the VB and Ultras datasets, it is an example of PhysMAP's graceful handling of inappropriate modalities that would have instead failed silently if a purely classifier approach had been taken.

PhysMAP can infer cell types using reference datasets

Guided by the previous section, we used waveform shape alone and visualized cells by optotagged identity. While most PV⁺ and SOM⁺ units occupied halves of one narrow-spiking cluster (Fig. 6A, lower right), numerous SOM⁺ and VIP⁺ cells were scattered throughout the larger cluster of wide-spiking excitatory cell types. The vast majority of neurons and cell types remain untagged since experimentalists can typically label only a few cell types, leaving most uncharacterized. This experimental constraint creates variable confidence across the neural population: regions with many tagged cells offer high classification confidence, whereas sparsely labeled regions remain uncharacterized (the “open-world problem” in machine learning⁶³). Yet nearly all previous cell type identification methods train on datasets containing only labeled cells, assigning labels uniformly without considering out-of-distribution cell types or varying confidence based on local ground truth density. What is needed is an identification approach that acknowledges incomplete knowledge and identifies cell types only where sufficient evidence exists. We address this need by creating a classifier that uses local densities of ground truth information exposed by PhysMAP.

We built a nearest-neighbor classifier based on the amount of ground truth labels surrounding every untagged neuron in PhysMAP's representation. For each untagged VB unit, we calculated the proportion of nearest neighbors from the Ultras dataset belonging to each labeled cell type (including other untagged cells) on the multimodal graph constructed by PhysMAP (Fig. S1J). This approach reduces the number of identified neurons but increases accuracy and avoids low-confidence regions (presumably composed of uncharacterized cell types). Based on this tradeoff (Fig. S13), we determined thresholds of 20% for nearest PV⁺ and 7.5% for nearest SOM⁺ neighbors, yielding 63 PV⁺ and 62 SOM⁺ units with 95% and 67% accuracy, respectively, all localizing to the narrow-spiking cluster shown by Fig. 6E and Fig. 6F, respectively. Note, these classification accuracies are higher than in Fig. 4B because the selection of ground truth nearest neighbors acts as a “confidence threshold”³⁰. As the proportion of ground truth neighbors increases, the accuracy of identification remains either roughly constant (SOM⁺) or monotonically increases (PV⁺) except for when the number of untagged cells with a certain ground truth neighbor proportion reaches zero.

As expected, inferred SOM⁺ cells aggregated strongly in SOM⁺-labeled areas and showed intermediate to narrow waveform widths matching ground truth SOM⁺ cells from previous analyses in this work. Similarly, inferred PV⁺ cells were more common in regions surrounded by Ultras PV⁺ cells and showed narrow waveforms consistent with known PV⁺ properties.

Thus, while we only identified a subset of these inhibitory neurons, we did so with higher confidence than possible from a simple classifier. PhysMAP's visualization-first approach allows direct assessment of where and why units are classified, building trust through transparency rather than relying solely on blind predictions from classifiers trained on datasets potentially containing batch effects.

Inferred types from PhysMAP match ground truth properties

In the previous section, we inferred PV⁺ and SOM⁺ cells by projecting VB and Ultras datasets into a shared multimodal representation and then using Ultras ground truth to predict cell types in VB. But do

the properties of the inferred cells in VB match known properties of these cell types? To determine this, we examined the properties of these inferred PV⁺ and SOM⁺ cells.

Inferred PV⁺ cells exhibited vigorous low-latency stimulus responses to the onset and offset of full-field flashes and spike frequency spectra peaks in the gamma frequency band (Fig. 7A). These are known characteristics of these types of interneuron^{31;32;64–66}. In contrast, inferred SOM⁺ cells (Fig. 7B) displayed lower maximum firing rates, suppressed or delayed responses to stimuli, and more variable responses than PV⁺ cells, matching previous findings^{31;67;68}. Furthermore, their spike spectra showed peaks at lower frequencies consistent with their role in driving visually induced oscillations at lower LFP frequencies (< 30 Hz)⁶⁶. These cell type-specific patterns were largely consistent across all cells of each inferred type, with PV⁺ cells showing strong responses to the onset and offset of stimuli (Fig. 7C) whereas SOM⁺ cells were more variable and sustained in response (Fig. 7D).

Collectively, inferred PV⁺ and SOM⁺ cells match their expected physiological properties *in vivo*. These results provide compelling evidence that PhysMAP can leverage ground truth datasets to enable the identification of cell types in datasets where ground truth is not available.

Case study: Using PhysMAP to identify microcircuits *in vivo*

Currently, technical limitations mean that circuit dynamics are inferred by examining only single cell types at a time. However, neural computation is not the product of cell types operating in isolation but instead emerges from the interactions between different cells⁶⁹. Having established that PhysMAP can identify several cell types simultaneously during behavior, we now demonstrate how this capability can be extended to the identification of microcircuits *in vivo*.

To demonstrate PhysMAP's ability to identify microcircuits during behavior, we examined the relationships between PV⁺, SOM⁺ cells and directly optotagged VIP⁺ cells all simultaneously within the same experiment (Fig. 8A; ^{70;71}). Note that since VIP⁺ cells were not identifiable given the waveform shape modality alone (only with 40% accuracy [Fig. S13]), here they needed to be directly identified. As additional verification, the firing rate patterns produced by this VIP⁺ cell (Fig. 8A, top) are very similar to the PSTH's of VIP⁺ cells documented elsewhere in the literature (compare Fig. 5G in⁷² to Fig. 8A, bottom).

Using pair-wise cross-correlations fitted with a generalized linear model of spiking interactions (GLMCC)⁷³, we found that a SOM⁺ and VIP⁺ cell pair showed symmetric decreases in firing around time lag zero (Fig. 8B), indicating possible mutual inhibition (a ubiquitous motif)^{72;74;75} and/or simultaneous inhibition from an upstream interneuron. Investigating this latter hypothesis, we examined the cross-correlogram between a third neuron and the SOM⁺ and VIP⁺ cells and found that it suggested that the VIP⁺-SOM⁺ pair were strongly inhibited following this upstream cell's spikes (Fig. 8C; top and bottom, respectively). The third neuron had fast firing properties and a vigorous response to stimuli (Fig. 8D, top), a waveform with an extremely short trough-to-peak width (0.23 ms; Fig. 8D, middle), with a predominant peak in the autocorrelogram around 67 Hz (gamma band; Fig. 8D, bottom) strongly suggestive of this being a PV⁺ cell (although it was not identified by our algorithm).

These types of connections were not hard to find in high-density probe recordings. In this particular experiment, there were 78 putative monosynaptic connections (Fig. S14A) including both inhibitory and excitatory types (Fig. S14B, top and middle respectively), with some also being bidirectionally connected neurons mixing both inhibitory and excitatory connections (Fig. S14B, bottom). The VIP⁺ and SOM⁺ units made several inhibitory connections each, furthering our confidence that they have been positively identified as inhibitory types (units 6 and 7 in Fig. S14A).

These results suggest that not only can PhysMAP deal with batch effects and identify multiple untagged cell types, but it can also identify microcircuit interactions in-vivo.

Discussion

We developed PhysMAP, an approach that uses multimodal electrophysiological information to compute a representation for simultaneous cell type identification *in vivo*. PhysMAP optimally combines single modalities across seven datasets, outperforms comparable deep-learning methods with substantially less computation³⁰, and enables analysis of interactions between multiple cell types during behavior. We also present a case study demonstrating how microcircuit dynamics can be identified without supervision during behavior.

While pure classifier approaches might seem sufficient for cell type identification, three major barriers make them unsuitable: ubiquitous hidden batch effects, scarcity of ground truth datasets, and uncertainty about which modalities to include. PhysMAP addresses all three prerequisites for trustworthy cell type identification.

First, PhysMAP visualizes dataset misalignments, providing a critical assessment of how appropriate a dataset is for cell type identification. This sanity-checking prevents erroneous conclusions from misapplied identification. Our experiments demonstrated dramatic accuracy decreases (by 22% and 36% in Fig. S11A and B respectively and by 19% and 26% in Fig. S12C and F respectively) when using datasets with misalignment due to batch effects. Direct identification, without visualization, does not make these effects visible. Other cell type identification tools have not grappled with batch effects because they assess performance on datasets without distributional shifts—training and test sets are randomly sampled from the same overall dataset, so by definition the distributional shift between sets tends to approach zero with increasing sample size. In the real world, batch effects always exist and they are invisible to users of direct classification approaches.

Second, most electrophysiological recordings lack ground truth. Of the large-scale mouse experiments with optotagging, each covers only a single brain region. For primates and other relevant species, no such datasets exist. PhysMAP offers exploratory visualization of neuronal heterogeneity—for example, revealing SOM⁺ cell diversity across cortical areas, with both broad- and narrow-spiking waveforms as predicted by the literature^{44;45;48}. Similarly, it revealed distinct properties of excitatory neurons across layers, aligning with observations from slice electrophysiological experiments¹. This physiological variation within “discrete” cell types is invisible to many pure classifiers, but is readily apparent with PhysMAP. And in cases where partial ground truth is available, PhysMAP’s interpretable visualizations offer insight into which cell types are identified and where.

Third, it is not well understood which modalities (provided they are not affected by batch effects) are most informative to use when building a reference dataset. PhysMAP offers an intuitive head-to-head comparison through visualization of each modality that pure classifier-based methods do not readily provide. For instance, in both juxtacellular and extracellular recordings, we show waveform shape is the most reliable modality for delineating ground truth cell type and ISI distribution, the least. These results reaffirm our previous work^{24;25} and are inconsistent with other work that emphasizes spike timing information. Moreover, PhysMAP provides this natively on a per-modality and per-unit basis, while interpretability in pure classifier methods necessitates retraining on data with modalities and units censored one by one.

One trade-off of this approach is that PhysMAP does not admit a reusable function for classifying un-

tagged cells once trained; instead, the WNN representation must be recomputed by combining both new and old data. A potential concern is that there are data leakage effects from pooling new and old data into a WNN transformation, wherein information from test data could influence the unsupervised transformation itself and bias cross-validation results. We evaluated this using the control analysis from⁷⁶ (See Supplementary Note 2) and found that any potential bias from applying the unsupervised transformation before train-test splitting is negligible and likely pessimistic. Specifically, validation accuracy (A_{val} : performance on held-out test data within the transformed subset) consistently matched generalization accuracy (A_{gen} : performance when the trained classifier is applied to completely held-out data not included in the transformation) across different data subset proportions, with the bias direction being consistently negative ($A_{\text{val}} \leq A_{\text{gen}}$), meaning our cross-validation results are likely slightly conservative estimates of true performance. In⁷⁶, preprocessing bias is only incurred through data leakage for datasets of around a hundred samples or fewer; ours number several thousand.

In light of these issues, how then should reference datasets be collected? We first recommend that ground truth generating experiments include standardized periods of “spontaneous activity” (involving no behavior) from which to generate informative spike timing modalities. These are then easily replicable as they can be appended onto a lab’s existing behavioral workflow. Thus, they are far less likely to be affected by batch effects (and this can be evaluated with PhysMAP). Additionally, other behavior-independent modalities should be explored, such as spike-triggered LFPs⁵⁶ or spike-LFP coupling, which may separate PV⁺ and SOM⁺ types^{77;78}. Other modalities might include specific “localizer” stimuli^{24;56} that generate distinctive activity patterns useful for identifying certain cell types. Other multimodal work has also shown that the multi-channel waveform “footprint” improves identification performance⁷⁹. The multi-channel waveform was not investigated in this work as that would preclude cross-dataset cell type identification (different probe geometries between Neuropixels 1.0 and Ultras).

Our analysis also revealed that few existing datasets have sufficient single unit isolation quality and comprehensive cell type coverage to serve as reference datasets. Only the Ultras dataset contained reliable tagging for all three major inhibitory types, and even this dataset covers only the visual cortex of the mouse. High-quality single-unit recordings with optotagging remain challenging due to light attenuation, transgene expression specificity⁸⁰, and low yield for rarer cell types like VIP⁺ neurons which compose only ~13% of inhibitory neurons⁸¹. In addition, single unit isolation also remains challenging even with modern technologies⁸² and sorting methods⁸³. A recent study⁸⁴ predicts that, across publicly available datasets, isolation quality varies widely with between 3.1 and 50.0% of spikes from purported single units being false positives from nearby units. While this may not significantly impact population-level analyses⁸⁵, it can blur physiological differences between cell types which demand high-quality unit isolation. For example, mixing fast-firing narrow-spiking PV⁺ cells with slow-firing broad-spiking excitatory cells can potentially create artifactual intermediate-property units, thus underscoring the importance of high-quality single unit isolation²⁴. This mixing effect is evident when comparing unambiguous juxtacellular recordings (showing stronger cell type structure) with extracellular recordings (showing less structure) even in highly curated datasets. Despite these issues, recent work has demonstrated several ways to improved dataset quality: higher-density recordings⁵⁵ and improved light delivery via waveguide optrodes⁸⁶ have been shown to increase yields. Additionally, improved protocols for isolating directly tagged cells, such as using synaptic transmission blockers³⁰, further improve ground truth quality.

Given these barriers presented, it might be tempting to ask why to choose to examine cell types with electrophysiological over optical approaches. Electrophysiology offers several advantages over optical approaches for cell type characterization: simultaneous monitoring of multiple cell types on single trials, scalable access to many cells across the entire brain, monitoring of circuit interactions, and exploration of cell types in primates and non-model organisms. While optical methods allow cell type-specific population

monitoring with perfect single unit isolation, most currently limit imaging to only two fluorophores simultaneously⁸⁷. Electrophysiology enables monitoring all cell types together, helping understand single-trial circuit dynamics^{88;89}, and bridging anatomy and dynamics⁹⁰. Optical methods also require optical access, limiting them primarily to cortical regions due to light attenuation in tissue⁸⁰. Deeper imaging requires GRIN lenses or cortical aspiration, causing significant tissue damage. Furthermore, optical methods are limited by their “photon budget”⁹¹: only a fixed number of photons are deliverable into the brain and so there exists a trade-off between number of units recorded and sampling rate^{92;93}. Electrophysiology avoids these limitations, allowing arbitrary probe placement and simultaneous recording of thousands of neurons at high sampling rates²⁰, with technological improvements promising even greater scaling.

Optical methods also only offer their advantages at scale when applied to calcium imaging. However, calcium transients do not allow for the observation of interactions between cell types as this instead requires the imaging of action potentials. Even though voltage imaging allows individual spikes to be seen, this technique samples from even fewer neurons and only at shallower depths relative to calcium imaging due to hard technological barriers⁹³. As we have shown, electrophysiology can identify dozens of putative monosynaptic connections within a single recording.

Lastly, cell type-specific imaging requires transgene delivery and expression, limiting its use to certain model organisms. PhysMAP requires only electrical activity recordings, making it potentially valuable for studying cell types in non-human primates (pending generation of ground truth datasets), where viral methods and optical access are limited by technical constraints^{13;14;94}. PhysMAP provides a potential low-risk solution for obtaining cell type information from routine electrophysiology in primates.

In summary, PhysMAP offers robust exploration of putative cell types in unlabeled data and fast, efficient identification of cell types with ground truth. When used judiciously, it enables unsupervised identification of multiple cell types, in real-time, in behaving animals.

Methods

The methods are organized as follows: dataset descriptions, the weighted-nearest-neighbor method, performance analyses, and label transfer methods. Finally, we describe the label transfer method to identify cell types along with various analyses to verify their properties.

Statistics & Reproducibility

In each analysis, an open dataset was used (Table 1). We refer the reader to each dataset's respective publication for additional methodological information and for justification of sample size or data exclusion criteria. In some cases, data was additionally excluded relative to the originating dataset: for the dataset from³², we excluded spiking from ventral posteromedial thalamic neurons to keep the analysis and conclusions focused on cortex; for the dataset from⁹⁵, the signal-to-noise ratio threshold (SNR or spike quality index) was increased to 4.0 relative to the value of 2.5 used in the original publication; and for the Visual Behavior dataset, only units with an SNR above 3, an ISI violation rate below 5%, a presence ratio above 98%, and a baseline firing rate above 2 spikes/s were chosen. In addition, for VIP⁺-optotagged cells, we only selected those whose firing rate rose to at least twice their baseline in the first 10 ms after laser onset and were driven to at least a firing rate of 25 spikes/s. These metrics were empirically determined such that the projected space of the waveform shapes matched between Visual Behavior and the Neuropixels Ultra datasets.

Juxtacellular Mouse S1 Dataset

The juxtacellular dataset used in Fig. 1 and Fig. 2 and analyzed in section was collected by³² and downloaded from the associated file sharing portal (see Table 1). Recordings were performed in the primary somatosensory (barrel) cortex of Sst-IRES-Cre \times Ai32, Pvalb-IRES-Cre/Pvalb-Cre \times Ai32, or Vip-IRES-Cre \times Ai32 mice. These experiments used *in vivo* juxtacellular electrophysiological recordings with glass micropipettes. After recording, cells were filled with biocytin/neurobiotin. Filled neurons both acted as a verification of the recorded cell type via their morphology and also served as landmarks to align recording depths of each cell with cortical layers as determined by histology.

ChR2-expressing neurons were located during recording by observing laser-evoked spikes that occurred 1-2 ms after the onset of brief light pulses (5 ms) delivered at low frequency (5 Hz) for 3 out of every 10 s. Spikes were examined by eye and non-spike events were removed after projection in PCA-space. To control for drift, only waveforms between the 25th and 75th percentile of peak-to-trough duration for all spikes from a cell were used in further waveform shape analyses (average waveform and waveform shape metrics). Averaged waveform shapes and ISI distribution were extracted for each neuron from the *spkwaveform_all.mat* file in the corresponding cell type structs. In addition, the waveform shape metrics of spike width and peak-to-trough ratio were also collected from this file. Cells labeled "fake SOM⁺" were reassigned to the PV⁺ population and cells labeled "putative SOM⁺" were reassigned to SOM⁺. This reassignment was based on a careful investigation in³² confirming earlier findings that off-target recombination occurs in fast-spiking neurons of somatostatin-IRES-Cre mice possibly due to transient Cre expression during development^{96;97}. PSTHs were calculated only from whisker touches during active whisking and activity resulting from both whisker protraction and retraction were used. Spiking was aligned to touch-onset and a PSTH calculated using 1 ms time bins. As not all cells contained PSTHs, only the subset of units containing all three modalities (waveform shape, ISI distribution, and PSTH) were used. To create the "concatenated" modality, we simply concatenated the waveform shape, ISI distribution, and PSTH for a given unit into a single long feature vector. Waveform metrics—peak-to-

trough duration and peak-to-trough ratio—were calculated in the originating publication as the time from the first peak to the first trough and the ratio of absolute values of the peak to the trough respectively.

Extracellular Mouse A1 Dataset

The extracellular data we analyzed from⁹⁵ was downloaded from GitHub (see Table 1). Recordings were collected via silicon probes (A4x2-tet configuration from NeuroNexus) and light was delivered via optic fiber at around 200 μm from the first recording site. PV⁺ and SOM⁺ cells were identified by optotagging in Pvalb-Cre or Sst-Cre mice respectively that were the progeny of crossing LSL-ChR2 mice. Photoidentification of ChR2-expressing cells was relatively conservative with positively identified cells being those that exhibited significant firing rate changes ($p < 0.001$, nonparametric Wilcoxon signed-rank test) in firing rate during the first 10 ms of stimulation-onset. Putative excitatory units were identified by a significant decrease in firing rate during the experiments in Sst-Cre mice. Recordings were collected from primary auditory cortex via silicon probe and stimulation was via an optical fiber located at the top of the probe recording sites. Spikes were sorted offline in Klustakwik and the units used in further analyses only if ISI violation rate (number of spike intervals < 2 ms) was less than 2% per cluster. In the original publication, only cells with spike quality index (SQI; ratio between the peak amplitude of the waveform and the average variance, calculated using the channel with the largest amplitude) above 2.5 were used; in our analysis, we only used cells with SQI above 4.0. Mean waveforms were calculated for a unit using the channel with greatest amplitude of spikes.

Extracellular Mouse Visual Cortex and Hippocampus Dataset (CellExplorer, Petersen et al.)

The third dataset was obtained from the CellExplorer package⁹⁸ and is heterogeneous, being composed of two separate extracellular electrophysiological datasets from the Allen Institute's Brain Observatory and the Buzsaki lab, and included waveform shape, ISI distribution, autocorrelogram (ACG), and various derived electrophysiological metrics. We used the derived electrophysiological metrics that all units shared and this included the following eleven:

- Spike width (*troughToPeak*)
- Peak-to-trough derivative (*troughToPeakDerivative*)
- Pre-hyperpolarization peak-to-post depolarization peak ratio (*ab_ratio*)
- Coefficient of variation (*cv2*) measuring ISI distribution
- ACG tau rise (*acg_tau_rise*)
- ACG tau decay (*acg_tau_decay*)
- ACG tau bursts (*acg_tau_burst*)
- ACG refractory period (*acg_refrac*)
- ACG decay amplitude (*acg_c*)
- ACG rise amplitude (*acg_d*)
- ACG burst amplitude (*acg_h*).

More information on these metrics is available on the CellExplorer website docs ([link](#)). For the analyses, these metrics were concatenated for each neuron into a single vector per unit to form the electrophysiological metrics modality. Given the large number of publications and protocols that have contributed to the CellExplorer dataset, we only provide a brief overview here.

Neuropixels Ultra Dataset (Ye & Shelton et al.)

This dataset was collected as part of⁵⁵ and was collected using Neuropixels Ultra probes inserted into mouse primary visual cortex. Unlike the originating dataset, we only used the peak amplitude channel which produced single-channel waveforms—this was for the purpose of being able to co-project both Visual Behavior and Neuropixels Ultra into the same embedding space and using a classifier to perform out-of-dataset classification. Unique to this dataset, an unsupervised procedure for opto-tagged unit identification was conducted by first concatenating the average PSTH for each stimulation pattern and intensity of a 1 second cosine ramp and a 10 ms square wave pulse at 0.2, 4.1, and 10.0 mW/mm². This formed a neuron-by-response vector matrix which was then normalized and to which PCA was applied followed by UMAP on each time series PC. Then, DBSCAN (density-based spatial clustering of applications with noise,⁹⁹) was applied with cluster radius (ϵ) set to the inflection point in distance distributions of the ten nearest neighbors. The cluster found by DBSCAN that contained a pattern of optogenetic excitation consistent with the pattern produced predominantly in response to the short 10 ms stimulus was then said to contain optotagged units. Only spikes that occurred during the “stim off” periods were used to calculate average normalized waveforms and 3D-ACGs.

Visual Behavior Dataset (Allen Institute)

The Visual Behavior dataset, containing the Images H and Images G experiments, was downloaded using the Allen Institute’s *AllenSDK* package. Units were selected only if they were from primary visual cortex and higher visual areas (HVAs). Furthermore, we were quite stringent with quality control: units were only selected for analysis if they had a signal-to-noise ratio (maximum amplitude of the mean spike waveform divided by the standard deviation of background noise) exceeding 3; a maximum of inter-spike interval violation over firing rate ratio of 0.05 (to limit the false discovery rate;⁸⁴); a minimum presence ratio of 0.98 (the proportion of time a unit was present in a recording); and a baseline firing rate of 2 spikes/s over the entire recording. In addition, optotagged units were selected if their firing rose to twice their baseline during the first 10 ms of laser onset and were driven to a firing rate of at least 25 spikes/s.

The 3D-ACGs were calculated by taking the spike times across the entire experiment except for the optotagging epoch which occurred at the end of the session. The 3D-ACGs used a 1000 ms window size and a 10 ms bin size and were log-transformed (matching the Ultras dataset) and were constructed using the *NeuroPyxels* function, *fast_acg3d*⁶¹.

PhysMAP Preprocessing

Modalities from all neurons were preprocessed via the same steps required by and performed using Seurat v4:

1. **PCA reduction:** All modalities were initially dimensionality reduced to the smallest input modality dimension since all modalities into Seurat must be the same dimensionality. For the S1 dataset, this is 30 dimensions; for the A1 dataset, this is 20 dimensions; and for Ultras and Visual Behavior, this is 81 dimensions.
2. **Normalization:** Each modality is separately normalized using a centered log ratio transform.
3. **Rescaling and centering:** Each feature of each modality is linearly rescaled to occupy the same unit variance. Each feature is then centered by mean subtraction.

Weighted Nearest Neighbor (WNN) Algorithm

The PhysMAP approach uses the WNN algorithm available from Seurat v4³⁵ which is summarized as the following steps.

1. **Nearest neighbor identification:** Within the preprocessed, high-dimensional space of a given modality (here, waveform shape Fig. S1A denoted “wf”), an i^{th} single unit (x_i^{wf}) is selected (Fig. S1B, blue sphere) and its nearest neighbors (indexed by j) identified via Euclidean distance in the ambient (un-dimensionality reduced) space (Fig. S1B, red spheres).
2. **Nearest neighbor prediction (within modality):** A prediction of the selected single unit’s waveform $\hat{x}_{i,\text{knn}_{\text{wf}}}^{\text{wf}} = \frac{\sum_{j=1}^k x_{i,j}^{\text{wf}}}{k}$ is made via unweighted averaging the k -nearest waveform neighbors (knn_{wf}) identified in Fig. S1B. By default, we use $k = 20$ and Euclidean distances in the ambient space to locate the nearest neighbors.
3. **Within-modal affinity calculation:** The unit’s nearest neighbor predicted waveform shape ($\hat{x}_{i,\text{knn}_{\text{wf}}}^{\text{wf}}$) generated by k -nearest neighbors vs. its actual waveform shape (x_i^{wf}) are both then passed into a modified UMAP distance kernel,

$$d_{\text{UMAP}}(x_i, \hat{x}_i) \equiv \exp \frac{-\max\left(0, d_{L^2}(x_i, \hat{x}_{i,\text{knn}=20}) - d_{L^2}(x_i, x_{i,\text{knn}=1})\right)}{\sigma_i - d_{L^2}(x_i, x_{i,\text{knn}=1})} \quad (1)$$

where $d_{L^2}(\cdot)$ is the Euclidean (L^2 -norm) distance metric; $d_{L^2}(x_i, x_{i,\text{knn}=1})$ is the Euclidean distance from x_i to the nearest single neighbor; and σ_i is a modified UMAP bandwidth equal to the average of the Euclidean distances from the i^{th} unit to the $k=20$ nearest units with lowest non-zero Jaccard index. Note that non-zero Jaccard index is the same as saying that two data points must have at least one graph node in common. This equation is used to calculate the “within-modal affinity” which gives a measure of how predictive a certain modality is of this unit. This forms the numerator of the “affinity ratio” shown in Fig. S1D.

4. **Nearest neighbor identification (cross modality):** Examining normalized single unit ISI distributions (the “cross-modal” space) generated from the same neurons in Fig. S1E, we locate the same unit and nearest neighbors previously identified but now in ISI distribution-space (Fig. S1F).
5. **Nearest neighbor prediction (cross modality):** As in Fig. S1C, a nearest neighbor prediction is made but this time, in a different modality albeit using the same neurons identified in the original modality (Fig. S1G, red dashed circle).

$$\hat{x}_{i,\text{knn}_{\text{wf}}}^{\text{isi}} = \frac{\sum_{j=1}^k x_{i,j}^{\text{isi}}}{k} \quad (2)$$

6. **Affinity ratio calculation:** This unit’s ISI dist. (x_i^{isi}) and the prediction of its ISI dist. ($\hat{x}_{i,\text{knn}_{\text{wf}}}^{\text{isi}}$) is computed by unweighted average of the waveform-space nearest neighbor units’ ISI dist. This is then used to compute the “cross-modal affinity” by passing it into a modified UMAP distance kernel; this forms the denominator of the waveform affinity ratio for a given unit $S_{\text{Wave}}(i)$ defined as,

$$S_{\text{Wave}}(i) = \frac{d_{\text{UMAP}}(x_i^{\text{wf}}, \hat{x}_{i,\text{knn}_{\text{wf}}}^{\text{wf}})}{d_{\text{UMAP}}(x_i^{\text{isi}}, \hat{x}_{i,\text{knn}_{\text{wf}}}^{\text{isi}})} \quad (3)$$

(Fig. S1D). This process is repeated for every unit in the dataset and also in the reverse manner: beginning with the ISI distribution-space as the within modality and with the waveform shape-space as the cross modality to find the ISI dist. affinity ratio $S_{\text{ISI}}(i)$ defined as

$$S_{\text{ISI}}(i) = \frac{d_{\text{UMAP}}(x_i^{\text{isi}}, \hat{x}_{i, \text{knn}_{\text{isi}}}^{\text{isi}})}{d_{\text{UMAP}}(x_i^{\text{wf}}, \hat{x}_{i, \text{knn}_{\text{isi}}}^{\text{wf}})} \quad (4)$$

For datasets with more than two modalities, the affinity ratio is calculated for each modality versus every other and for every unit. Thus for a dataset with d data points and n feature modalities, there will be $2d \cdot \binom{n}{2} = 2d \cdot \frac{n!}{2!(n-2)!} = d \cdot n(n-1)$ affinity ratios to calculate.

7. **Converting affinity ratios to modality weights:** With affinity ratios calculated for both waveform and ISI distribution across all units, modality weights ($\beta_{\text{Wave}}(i)$ and $\beta_{\text{ISI}}(i)$) for each unit can be calculated. For a given unit, this weight is the ratio of a given modality's affinity ratio for that unit divided by the sum of all other modality's affinity ratios for that unit (the softmax normalization). For the two-modality case, this is $\beta_{\text{Wave}}(i) = \frac{\exp(S_{\text{Wave}}(i))}{\exp(S_{\text{Wave}}(i)) + \exp(S_{\text{ISI}}(i))}$ and $\beta_{\text{ISI}}(i) = \frac{\exp(S_{\text{ISI}}(i))}{\exp(S_{\text{Wave}}(i)) + \exp(S_{\text{ISI}}(i))}$ (Fig. S1H). For m numbers of modalities, the additional affinity ratios are added to the denominator such that for any n^{th} modality's affinity ratio $\beta_n(i)$, it is

$$\beta_n(i) = \frac{\exp(S_n(i))}{\sum_1^m \exp(S_n(i))} \quad (5)$$

8. **Calculating unit pair-wise affinities:** For every pair of units (i and j) in each modality (n), an additional "pair-wise affinity" is also calculated by passing both of them and their coordinates in some modality into UMAP's distance kernel ($\theta_n = d_{\text{UMAP}}(i_n, j_n)$). This occurs for the same pair of units in both the waveform shape-space (Fig. S1I, top) and the ISI distribution-space (Fig. S1I, bottom) to determine pair-wise affinities $\theta_{\text{Wave}}(i, j)$ and $\theta_{\text{ISI}}(i, j)$ respectively. This is done for all pairs of points and all modalities.
9. **Creating the WNN:** To create the weighted-nearest neighbors representation, a modality-weighted sum of the pair-wise affinities (Fig. S1J, top) is calculated to get the multimodal affinity between any pair of points ($\theta_{\text{WNN}}(i, j)$) from m modalities

$$\theta_{\text{WNN}}(i, j) = \sum_{n=1}^m \beta_n(i) \theta_n(i, j) \quad (6)$$

is taken to produce a "connectivity matrix" (Fig. S1J, bottom). It should be noted that $\theta_{\text{WNN}}(i, j) \neq \theta_{\text{WNN}}(j, i)$ and thus the pair-wise affinities are symmetrized by calculating the fuzzy union³⁸: if A is the weighted adjacency matrix formed by $\theta_{\text{WNN}}(i, j)$ across all pairs of points, then the symmetrized adjacency matrix (B) is

$$B = A + A^T - A \odot A^T \quad (7)$$

where \odot is the Hadamard product. The k -nearest neighbor algorithm is applied to this symmetrized matrix to form the final symmetrized WNN (default number of neighbors is $k = 200$) which is then visualized into two dimensions with UMAP's force-directed graph layout projection for visualization.

Leiden Community Detection

To calculate the MARI score for each dataset across different numbers of clusters we used Leiden clustering¹⁰⁰ on each dataset's WNN graphs with resolution between 0.1 and 3.0 in 0.1 resolution intervals. The Leiden algorithm is a method for "community detection" which finds highly inter-connected nodes on a network graph akin to clustering in a metric space. Algorithms like Leiden (and the simpler Louvain algorithm) attempt to find a partitioning of the network into a set of communities that maximizes the total modularity of the network. This modularity, \mathcal{H} , is a measure of how inter-connected the nodes are within a community versus outside of it. In the below definition, m is the total number of edges in the network, e_c is the number of edges in a community c , γ is the resolution parameter, and K_c is the sum of the degrees of the nodes in community c .

$$\mathcal{H} \equiv \frac{1}{2m} \sum_c \left(e_c - \gamma \frac{K_c^2}{2m} \right) \quad (8)$$

Thus, the resolution parameter is effectively a prior on the number of expected communities that should be found during the optimization; with lower resolution, less communities are found and with higher resolution, more communities are found.

Modified Adjusted Rand Index (MARI) Calculation

Once this partitioning of a graph into communities is computed, a MARI score⁵⁰ is calculated between each of these neuron's community membership and the ground truth cluster identity to determine how closely they correspond. To understand MARI, which is a modification of the adjusted Rand index (ARI), we begin with an explanation of the Rand index. Given two clusterings, X and Y of the same dataset, the number of pair-wise elements that share cluster membership in both X and Y is a and the number of pair-wise elements that share different cluster memberships is b . The Rand index (RI) is $RI = \frac{a+b}{\binom{n}{2}}$

where n is the total number of pairs. The Rand index can also be interpreted as the sum of the number of true positive and true negative correspondences divided by the total number of guesses if one clustering is regarded as a classifier prediction of the other. However, in the classifier interpretation, the Rand index is not corrected for chance predictions. The ARI corrects for these chance predictions by subtracting the expected Rand index under the hypergeometric distribution given two independent clusterings and fixed number of clusters in each. MARI further refines this metric by instead incorporating a multinomial distribution which does not enforce cluster size and is a better assumption given that most clustering algorithms (including Leiden) do not fix cluster sizes. This was used to compare each of the Leiden clusterings at different resolutions to underlying ground truth clusters.

Peri-stimulus Time Histogram Calculation

For each Leiden cluster in Fig. 2D, a peri-stimulus time histogram (PSTH) was calculated by aligning times to each whisker deflection onset and binning spikes into 10 ms bins. This was then normalized by subtracting the baseline spike rate for each neuron and then dividing by the standard deviation of each unit's PSTH. These were then smoothed by a 5-point moving average and then grouped according to Leiden cluster membership for between -10 and 50 ms before and after the whisker deflection.

Identification Algorithm

After a WNN is constructed for a dataset, we used UMAP's force-directed graph layout procedure to project the graph into a high-dimensional embedded space. Dimensionality for this embedding was maximized to provide the greatest amount of information available for classifier comparison (the WNN algorithm only allows for a maximal output dimensionality equal to the highest dimensional input modality). We then used a random 15-85% test-train split and trained a stochastic gradient boosted tree model (GBM) classifier on the training set with five-fold cross-validation using the *caret* package in R¹⁰¹ to identify each underlying cell type with a multiclass (one-vs-rest) objective function. In some datasets, cell type classes with less than 10 examples were excluded because this would not guarantee that a cell would appear in both the training and test sets at the test-train split used. We then repeated this train-test split 20 times and averaged the balanced accuracy over these independent runs. Each classifier also underwent hyperparameter tuning via a grid search. This procedure was run identically for the controls in Fig. S6 except with varying graph embedding dimensions and using different classifiers available in the *caret* package. This same approach was used for the three modality case in Fig. S5E.

Gaussian mixture model analysis of spike metrics

A Gaussian mixture model Fig. S3 was fit to units in³² by applying the *mclust* library's *Mclust* function to the log-scaled peak-to-trough and log-scaled spike width metrics for each unit as done previously in the literature^{23;25}. This function simultaneously compares many GMM models and finds the optimal number of clusters for each via the Bayesian information criterion (BIC). It arrived at a VVE (ellipsoidal with equal orientation) model and three clusters.

β -variational autoencoder and multi-layer perceptron identification architectures

Using the Ultras dataset, a β -variational autoencoder (VAE) was constructed using a standardized architecture for each modality (Fig. S10): a first/second 1-D convolutional layer ($N = 64$, $C_{in} = 1/32$, $C_{out} = 32/64$, $K_{size} = 3$) followed each by ReLU non-linearities and flattened; passed into an 81-dimensional latent embedding parametrized by mean μ and log variance $\log \sigma^2$ (linear transform layers, latent dim. = 81, hidden dim. = 64); then unflattened and passed into a decoder composed of a first/second 1-D convolutional layer ($N = 64$, $C_{in} = 64/32$, $C_{out} = 32/1$, $K_{size} = 3$) with a ReLU in-between. Networks were trained with the Adam optimizer over 200 epochs, grid search optimized for β based upon classification accuracy, and with logistic Kullback-Leibler (KL) divergence annealing. The β -VAE was optimized using the evidence lower bound objective function (ELBO) balancing both L^2 norm reconstruction error and KL-divergence multiplied by the tunable β parameter. For the waveform shape VAE, $\beta = 0.05$ and for the 3D-ACG VAE, $\beta = 0.1$; notice that these optimal β reduce the VAE to nearly a vanilla autoencoder. The multi-layer perceptron (MLP) classifier (scikit-learn's *MLPClassifier*) was composed of an input layer (10 units), hidden layer (100 units), ReLU non-linearity, final output layer (4 units), and softmax output for cell type prediction. The MLP classifier was trained for 1000 iterations with early stopping. Each architecture was trained on an NVIDIA L40S GPU with CUDA-optimized code.

Identification comparison

For each confusion matrix in Fig. 4B, the projection and identification process was trained 25 times for each of PhysMAP, waveform shape, 3D-ACG, and the β -VAE. For each iterate, the transformation was applied to the entire dataset and then a seed selected to determine a random 90-10 train-test data split. Across ten random splittings, five-fold repeated cross-validation was applied and averaged (across

repeated CV's and splittings) to determine each cell type's raw classification accuracy. We also followed this same procedure for the confusion matrices in Fig. S11 with and without erroneous alignment. This procedure was also applied to PhysMAP with and without alignment on the C4 dataset shown in Fig. S11B⁶¹. The alignment procedure is not used in further iterations of PhysMAP but is documented for comparison and to demonstrate the deleterious effects of misaligned data distributions. The alignment procedure is explained in the next section for clarity.

Distributional shift in waveform shape

To analyze the effects of distributional shift on classifier accuracy, we applied both 250 Hz- and 500 Hz-high-pass 4th order Butterworth filters to half of the normalized mean waveforms in the Ultras dataset (Fig. S12A and D, respectively for each filter). A random forest classifier with 5-fold cross-validation ($n_estimators = 100$, $max_depth = 10$) was trained on the half of the data that was unfiltered and then tested on the other half of the data (Fig. S12C and F, respectively for each filter). This analysis simulates the usage of a cell type classifier that was trained on a dataset with different but reasonably chosen preprocessing filters applied to waveform shapes^{42;102}.

Reference mapping with anchors and cell type identification

Unlike the previous identification analyses, an 80-20 train-test split was conducted *before* applying PhysMAP. This was done to combat bias in cross-validation because of data leakage due to pre-processing transformations⁷⁶. The training set was used to create the reference mapping upon, and the test set was used to create the query dataset. To construct the reference mapping, a WNN was constructed from each neuron's waveform shape, ISI distribution, autocorrelogram, and derived electrophysiological metrics as before. Following the procedure for multimodal reference mapping recommended in Hao & Hao et al. 2021, we first ran supervised principal component analysis (SPCA) on the reference data. Next, the query dataset is projected onto the reference using the previously computed SPCA transform. Now that reference and query datasets occupy the same space, "anchors" can be calculated between the reference and query. These anchors are pairs of cells between the reference and query that are located within each other's neighborhoods; this concept is also referred to as mutual nearest neighbors (MNN). These neighborhoods are defined by computing k -nearest neighbors with $k = 5$. This nearest neighbor search is conducted in only the top-30 SPC's. Once all anchors are found, they are scored, which is an assessment of how confident we are in their correspondences. This score is calculated for each anchor given the 30 nearest neighbors in each reference and query dataset for both the reference and query data points of the anchor, respectively. The overlap of these nearest neighbor matrices between the 0.01 and 0.90 quantiles is then linearly rescaled to be between 0 and 1 for all reference-query anchor pairs; this provides a score for each anchor. A weight matrix is then constructed between each query cell c and each i th nearest anchor a_i (where $i \in [1, 50]$) based upon the distance to each query cell and the corresponding anchor score. These weighted distances $D_{c,i}$ are calculated as

$$D_{c,i} = S_{a_i} \left(1 - \frac{\|c, a_i\|}{\|c, a_{50}\|} \right) \quad (9)$$

where S_{a_i} is the score of the i th anchor where $\|\cdot\|$ is the Euclidean distance. These distances are passed through a Gaussian distance kernel $\tilde{D}_{c,i}$ as,

$$\tilde{D}_{c,i} = 1 - \frac{-D_{c,i}}{e^{(2/sd)^2}} \quad (10)$$

with $sd = 1$ by default. To form the entries of a weight matrix

$$W_{c,i} = \frac{\tilde{D}_{c,i}}{\sum_{j=50} \tilde{D}_{c,j}}. \quad (11)$$

For cell type identification, a classification matrix L is created where each row corresponds to a ground truth cell type class and each column corresponds to each reference anchor. If a certain reference anchor pertains to a certain cell type, it is given a 1 and a 0 if otherwise. Label predictions P are then computed simply as,

$$P = LW^T. \quad (12)$$

This returns a prediction score for each cell in the query dataset. A query cell is thus given a predicted cell type of whichever class has the highest prediction score.

Joint embedding and label transfer of Ultras and Visual Behavior

Both Ultras' and Visual Behavior's waveform shape and 3D-ACG modalities were concatenated and passed into PhysMAP. The multimodal representation was generated using *caret* with centered log ratio transform normalization on a per-unit basis. UMAP was calculated for 20 components, but the 2 component projection is used for the visualizations shown in Fig. 5A. The same approach was used for the individual modalities shown in Fig. 5B, top and bottom. For the multi-dataset nonlinear embedding of waveform shape (unimodal PhysMAP), shown are the optotagged cell type identities (PV^+ , SOM^+ , and VIP^+) from each experiment amidst a background of untagged neurons Fig. 6E. Labels for Visual Behavior were generated using the previously described procedure (Methods: Pre-processing of the Visual Behavior dataset); labels for the Ultras dataset were generated by applying density-based spatial clustering of applications with noise (DBSCAN) to a UMAP embedding of PSTH responses in the first 10 ms after laser-onset⁵⁵.

For label transfer from Ultras tags to Visual Behavior, we used a nearest neighbor approach: for a fixed number of neighbors ($k = 33$) for each Visual Behavior unit within the high-dimensional graph constructed within PhysMAP, we calculated what percentage of neighbors (including unlabeled cells) contains a given label from the Ultras dataset and set a minimum percentage threshold belonging to a unitary cell type label (20% for PV^+ cells and 7.5% for SOM^+ cells) in order for a label to transfer. These inferred cell types are shown in (Fig. 6F and G). This percentage can be adjusted to the desired level of confidence; accuracy increases with an increasing percentage of neighbors in the Ultras dataset with accuracy measured as the number of correct identifications if only the Ultras dataset is considered Fig. S13. The procedure was fixed at $k = 33$ and repeated 25 times for each percentile.

This method worked well for both PV^+ and SOM^+ cells, but because VIP^+ cells were only very sparsely labeled and seemingly non-differentiated from broad-spiking excitatory cells, we used the direct labels obtained during the Visual Behavior experiments. Furthermore, VIP^+ cells were poorly identifiable even at the highest confidence thresholds (a maximum accuracy of 40% at 8% nearest neighbor threshold; Fig. S13, red lines).

Firing rate analysis of inferred cell types

For each peri-stimulus time histogram in Fig. 8A, bottom, firing rates were determined by collecting spikes into 100 ms time bins, summed across trials, and then smoothed by a Gaussian kernel with a standard deviation of the number of total bins divided by 6. The same procedure was done for each

individual inferred unit in the Visual Behavior dataset and normalized (dividing by the maximum firing rate) to generate Fig. 8B. The stimuli used were a gray to black flash on a computer monitor for 100 ms across most of the visual field contralateral to the recorded hemisphere. In Fig. 8D, a natural image was used as the stimulus aligned instead. This image was one that was correctly identified as a change from the previously shown natural images indicated by the mouse licking a lickspout during the reward window.

For the autocorrelograms in Fig. 8D, bottom, spikes were gathered into 5 ms time bins and shown for a window 150 ms before and after 0 lag. To generate the trial-averaged spike spectra in Fig. 8E, the procedure in the chapter titled “Analysis of Rhythmic Spike Train Data” from¹⁰³ was used. Specifically, this is a multitaper spectral density estimation method using 4 Slepian sequences (*scipy.signal*). The “pre-stimulus period” was taken as the 350 ms before to 100 ms after stimulus onset and the “post-stimulus period” is taken as the 100 ms after to 550 ms after stimulus onset. The reason for not aligning the pre- and post-stimulus periods directly with stimulus onset is the approximately 100 ms propagation time of stimulus information from retina to visual cortex (intuitively seen as a change in activity across most cell types). These were divided into both “hit” and “miss” trials Fig. 8E, middle and bottom in which the animal correctly identified the natural image as different from the ones before or else failed to respond to the changing image (indicated by a lick or lack thereof) respectively.

To generate the CCGs in Fig. 8F and G, we used a window/bin size of 3 ms/0.1 ms for the SOM⁺ to VIP⁺ CCG and 10 ms/0.2 ms for the VIP⁺ and SOM⁺ onto putative PV⁺ respectively. The ACG for the putative PV⁺ cell is shown with 100 ms window size and 5 ms bin sizes.

Identifying putative monosynapses with a GLMCC

To identify putative monosynaptic connections between neurons, we examined each pair of neurons for low latency responses on the timescale of synaptic transmission delays. We used the revised general linear model on cross-correlogram (GLMCC) approach by⁷³ which tends to perform better than both using classical cross-correlogram analysis¹⁰⁴ and jittered cross-correlations¹⁰⁵ in terms of minimizing both false positives and false negative rates. Furthermore, the GLMCC also yields an estimate of the strength of synaptic connectivity other than the classical and jittered methods, which only yield a binary identification or non-identification of a monosynaptic connection. More specifically, the underlying spike rate of a given neuron over time $c(t)$ is given as the product of several exponential distributions modeling respectively slow time-varying fluctuations affecting each neuron in a pair $a(t)$, the connection from the first neuron to the second J_{ij} , and the connection of the second neuron to the first J_{ji} .

$$c(t) = \exp\left(a(t) + J_{ij}f(t) + J_{ji}f(-t)\right) \quad (13)$$

Here, $f(t)$ is modeled as,

$$f(t) = \exp\left(-\frac{t-d}{\tau}\right) \quad (14)$$

where d is the synaptic delay and τ is the synaptic timescale; the former is set to be between 1 and 4 ms while the latter is set to be 4 ms. Note that the underlying spike rate $c(t)$ is modeled as a Poisson point process such that the spike probability $p(\{t_k\}|\theta)$ over time with,

$$p(\{t_k\}|\theta) = \prod_k c(t_k) \exp\left(-\int_{-W}^W c(t) dt\right) \quad (15)$$

with a parametrization of $\theta = \{a(t), J_{ij}, J_{ji}\}$ and W being the integration window (set to 50 ms in our analysis). The prior distribution $p(\theta)$ is proportional to the following,

$$p(\theta) \sim \exp\left(-\frac{1}{\gamma} \int_{-W}^W \left(\frac{da}{dt}\right)^2 dt\right) \quad (16)$$

with the parameter γ setting the “flatness” of the $a(t)$ is set to 2×10^{-4} in our analysis. The posterior distribution $p(\theta|\{t_k\})$ was calculated given Bayes’ theorem and a maximum *a priori* (MAP) procedure used to calculate the parameters θ .

A monosynaptic connection is said to exist if the null hypothesis is rejected; the null hypothesis is that two neurons generate spikes at their baseline firing rate independently of each other. The variance of a Poisson process within a time interval Δ is equal to its mean and thus the mean number of spikes n is $n = c(0)\Delta$. If J is small then the average number of spikes caused by a connection over an interval Δ is,

$$\delta n = c(0)J\tau(1 - e^{-\Delta/\tau}). \quad (17)$$

The condition that a monosynaptic connection is found is given by $|\delta n| > z_\alpha \sqrt{n}$ where $z_\alpha = 3.891$ for the $\alpha = 0.0001$ used in this analysis. Isolating the estimated connection strength \hat{J} by substituting $|\delta n| > z_\alpha \sqrt{n}$ into the previous equation yields,

$$|\hat{J}| > z_\alpha \frac{\Delta^{1/2}}{\tau(1 - e^{-\Delta/\tau})} \cdot \frac{1}{c(0)^{1/2}}. \quad (18)$$

The variable Δ reaches its minimum at $\Delta = 1.26\tau$ which reduces the previous equation to the following,

$$|\hat{J}| > 1.57z_\alpha(\tau c(0))^{-1/2} \quad (19)$$

for which a connection is determined to be monosynaptic.

Data Availability Statement

Post-processed data used in analyses and figure generation are provided in the repository https://github.com/EricKenjiLee/PhysMAP_Manuscript. Raw datasets are available from each source publication’s open data repositories in Table 1. Additionally, source data for all non-schematic main and supplementary figures are provided as pages in a “Source Data” Excel file.

Code Availability Statement

MATLAB, Python, and R for open dataset analysis and figure generation are available at https://github.com/EricKenjiLee/PhysMAP_Manuscript and is citable at <https://doi.org/10.5281/zenodo.18239113>.

References

- [1] Scala F, et al. (2021) Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* 598(7879):144–150.
- [2] Gouwens NW, et al. (2020) Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. *Cell* 183(4):935–953.
- [3] (BICCN) BICCN (2021) A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* 598(7879):86–102.
- [4] Bomkamp C, et al. (2019) Transcriptomic correlates of electrophysiological and morphological diversity within and across excitatory and inhibitory neuron classes. *PLoS Computational Biology* 15(6):e1007113.
- [5] Keller AJ, et al. (2020) A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex. *Neuron*.
- [6] Kullander K, Topolnik L (2021) Cortical disinhibitory circuits: cell types, connectivity and function. *Trends in Neurosciences* 44(8):643–657.
- [7] Christensen AJ, Ott T, Kepecs A (2022) Cognition and the single neuron: How cell types construct the dynamic computations of frontal cortex. *Current Opinion in Neurobiology* 77:102630.
- [8] Ferguson KA, Cardin JA (2020) Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience* 21(2):80–92.
- [9] Millman DJ, et al. (2020) VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli. *eLife* 9:e55130.
- [10] Adesnik H, Naka A (2018) Cracking the function of layers in the sensory cortex. *Neuron* 100(5):1028–1043.
- [11] Bohlen M, Tremblay S (2023) Vectorology for Optogenetics and Chemogenetics. *Neuromethods* pp. 293–329.
- [12] Bliss-Moreau E, Costa VD, Baxter MG (2022) A pragmatic reevaluation of the efficacy of nonhuman primate optogenetics for psychiatry. *Oxford Open Neuroscience* 1.
- [13] O'Shea DJ, et al. (2017) The need for calcium imaging in nonhuman primates: New motor neuroscience and brain-machine interfaces. *Experimental Neurology* 287:437–451.
- [14] Trautmann EM, et al. (2021) Dendritic calcium signals in rhesus macaque motor cortex drive an optical brain-computer interface. *Nature Communications* 12(1):3689.
- [15] Takasaki K, Abbasi-Asl R, Waters J (2020) Superficial bound of the depth limit of 2-photon imaging in mouse brain. *eNeuro* 7(1):ENEURO.0255–19.2019.
- [16] Jun JJ, et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551(7679):nature24636.
- [17] Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD (2019) Distributed coding of choice, action and engagement across the mouse brain. *Nature* pp. 1–8.
- [18] Siegle JH, et al. (2021) Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* pp. 1–7.
- [19] Chen S, et al. (2024) Brain-wide neural activity underlying memory-guided movement. *Cell* 187(3):676–691.e16.
- [20] Bondy AG, et al. (2024) Coordinated cross-brain activity during accumulation of sensory evidence and decision commitment. *bioRxiv*.
- [21] Paulk AC, et al. (2021) Large-scale neural recordings with single-cell resolution in human cortex using high-density neuropixels probes. *bioRxiv*.
- [22] Leonard MK, et al. (2023) Large-scale single-neuron speech sound encoding across the depth of human cortex. *Nature* pp. 1–10.

- [23] Trainito C, Nicolai Cv, Miller EK, Siegel M (2019) Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex. *Current Biology*.
- [24] Carr N, et al. (2025) High-resolution laminar recordings reveal structure-function relationships in monkey V1. *bioRxiv*.
- [25] Lee EK, et al. (2021) Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *eLife* 10:e67490.
- [26] Sun SH, Killian NJ, Pezaris JS (2024) More than expected: extracellular waveforms and functional responses in monkey lgn. *bioRxiv*.
- [27] Latuske P, Toader O, Allen K (2015) Interspike intervals reveal functionally distinct cell populations in the medial entorhinal cortex. *Journal of Neuroscience* 35(31):10963–10976.
- [28] Schneider A, et al. (2023) Transcriptomic cell type structures in vivo neuronal activity across multiple timescales. *Cell Reports* 42(4):112318.
- [29] Abbaspoor S, Hoffman K (2023) State-dependent circuit dynamics of superficial and deep ca1 pyramidal cells in macaques. *bioRxiv*.
- [30] Beau M, et al. (2025) A deep learning strategy to identify cell types across species from high-density extracellular recordings. *Cell*.
- [31] Pinto L, Dan Y (2015) Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron* 87(2):437–450.
- [32] Yu J, Hu H, Agmon A, Svoboda K (2019) Recruitment of GABAergic Interneurons in the Barrel Cortex during Active Tactile Behavior. *Neuron* 104(2):412–427.e4.
- [33] Ramadan M, et al. (2022) A Standardized Nonvisual Behavioral Event Is Broadcasted Homogeneously across Cortical Visual Areas without Modulating Visual Responses. *eNeuro* 9(5):ENEURO.0491–21.2022.
- [34] Cembrowski MS, Menon V (2018) Continuous variation within cell types of the nervous system. *Trends in Neurosciences* 41(6):337–348.
- [35] Hao Y, et al. (2021) Integrated analysis of multimodal single-cell data. *Cell* 184(13):3573–3587.e29.
- [36] Baltrusaitis T, Ahuja C, Morency LP (2019) Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41(2):423–443.
- [37] Zhu C, Preissl S, Ren B (2020) Single-cell multimodal omics: the power of many. *Nature Methods* 17(1):11–14.
- [38] McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv*.
- [39] Krimer LS, et al. (2005) Cluster analysis-based physiological classification and morphological properties of inhibitory neurons in layers 2–3 of monkey dorsolateral prefrontal cortex. *Journal of neurophysiology* 94(5):3009–3022.
- [40] Zaitsev AV, et al. (2008) Interneuron Diversity in Layers 2–3 of Monkey Prefrontal Cortex. *Cerebral Cortex* 19(7):1597–1615.
- [41] Swanson E, et al. (2021) Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using tea-seq. *eLife* 10:e63632.
- [42] Lee K, Carr N, Perliss A, Chandrasekaran C (2023) Wavemap for identifying putative cell types from in vivo electrophysiology. *STAR Protocols* 4(2):102320.
- [43] Huang W, et al. (2024) On the Comparison between Multi-modal and Single-modal Contrastive Learning. *arXiv*.
- [44] Urban-Ciecko J, Barth AL (2016) Somatostatin-expressing neurons in cortical networks. *Nature Reviews Neuroscience* 17(7):401–409.

- [45] Wu SJ, et al. (2023) Cortical somatostatin interneuron subtypes form cell-type-specific circuits. *Neuron*.
- [46] Yavorska I, Wehr M (2016) Somatostatin-Expressing Inhibitory Interneurons in Cortical Circuits. *Frontiers in Neural Circuits* 10:76.
- [47] Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience* 12(5):nn.2303.
- [48] Hostetler RE, Hu H, Agmon A (2023) Genetically defined subtypes of somatostatin-containing cortical interneurons. *eNeuro* 10(8).
- [49] Sun SH, et al. (2021) Analysis of extracellular spike waveforms and associated receptive fields of neurons in cat primary visual cortex. *The Journal of Physiology* 599(8):2211–2238.
- [50] Sundqvist M, Chiquet J, Rigai G (2020) Adjusting the adjusted rand index – a multinomial story. *arXiv: Methodology*.
- [51] Kvitsiani D, et al. (2013) Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498(7454):363–366.
- [52] Kim D, et al. (2016) Distinct Roles of Parvalbumin- and Somatostatin-Expressing Interneurons in Working Memory. *Neuron* 92(4):902–915.
- [53] Petersen PC, Hernandez M, Buzsáki G (2020) The Buzsáki Lab Databank - Public electrophysiological datasets from awake animals.
- [54] Sanzeni A, et al. (2020) Inhibition stabilization is a widespread property of cortical networks. *eLife* 9:e54875.
- [55] Ye Z, et al. (2025) Ultra-high-density neuropixels probes improve detection and identification in neuronal recordings. *Neuron* 113(23):3966–3982.
- [56] Herzfeld DJ, Hall NJ, Lisberger SG (2025) Strategies to decipher neuron identity from extracellular recordings in the cerebellum of behaving non-human primates. *bioRxiv*.
- [57] Laboratory IB, et al. (2024) Reproducibility of in vivo electrophysiological measurements in mice. *bioRxiv* p. 2022.05.09.491042.
- [58] Zhang Y, Patil P, Johnson WE, Parmigiani G (2021) Robustifying genomic classifiers to batch effects via ensemble learning. *Bioinformatics* 37(11):1521–1527.
- [59] Soneson C, Gerster S, Delorenzi M (2014) Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PloS one* 9(6):e100335.
- [60] Stuart T, et al. (2019) Comprehensive Integration of Single-Cell Data. *Cell* 177(7):1888–1902.e21.
- [61] Beau M, et al. (2021) Neuropixels: loading, processing and plotting neuropixels data in python.
- [62] AIBS (2022) Allen Brain Observatory: Visual Behavior Neuropixels Technical White Paper.
- [63] Zhu F, et al. (2025) Open-world machine learning: A review and new outlooks.
- [64] Sohal VS, Zhang F, Yizhar O, Deisseroth K (2009) Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* 459(7247):698–702.
- [65] Cardin JA, et al. (2009) Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459(7247):663–667.
- [66] Chen G, et al. (2017) Distinct inhibitory circuits orchestrate cortical beta and gamma band oscillations. *Neuron* 96(6):1403–1418.e6.
- [67] Ma Wp, et al. (2010) Visual Representations by Cortical Somatostatin Inhibitory Neurons—Selective But with Weak and Delayed Responses. *The Journal of Neuroscience* 30(43):14371–14379.
- [68] Gentet LJ, et al. (2012) Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nature Neuroscience* 15(4):607–612.
- [69] Luo L (2021) Architectures of neuronal circuits. *Science* 373(6559):eabg7285.
- [70] Bastos AM, et al. (2012) Canonical microcircuits for predictive coding. *Neuron* 76(4):695–711.

- [71] Fishell G, Kepecs A (2019) Interneuron Types as Attractors and Controllers. *Annual Review of Neuroscience* 43(1):1–30.
- [72] Lee S, Kruglikov I, Huang ZJ, Fishell G, Rudy B (2013) A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nature Neuroscience* 16(11):1662–1670.
- [73] Kobayashi R, et al. (2019) Reconstructing neuronal circuitry from parallel spike trains. *Nature Communications* 10(1):4468.
- [74] Campagnola L, et al. (2022) Local connectivity and synaptic dynamics in mouse and human neocortex. *Science* 375(6585):eabj5861.
- [75] Pi HJ, et al. (2013) Cortical interneurons that specialize in disinhibitory control. *Nature* 503(7477):521–524.
- [76] Moscovich A, Rosset S (2022) On the Cross-Validation Bias due to Unsupervised Preprocessing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84(4):1474–1502.
- [77] Wei Y, et al. (2023) Associations between in vitro, in vivo and in silico cell classes in mouse primary visual cortex. *Nature Communications* 14(1):2344.
- [78] Onorato I, et al. (2023) Distinct roles of PV and Sst interneurons in visually-induced gamma oscillations. *bioRxiv*.
- [79] Yu H, et al. (2025) In vivo cell-type and brain region classification via multimodal contrastive learning in *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24–28, 2025*. (OpenReview.net).
- [80] Li N, et al. (2019) Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *eLife* 8:e48622.
- [81] Prönneke A, et al. (2015) Characterizing VIP Neurons in the Barrel Cortex of VIP^{cre}/tdTomato Mice Reveals Layer-Specific Differences. *Cerebral Cortex* 25(12):4854–4868.
- [82] Buccino AP, Garcia S, Yger P (2022) Spike sorting: new trends and challenges of the era of high-density probes. *Progress in Biomedical Engineering* 4(2):022005.
- [83] Pachitariu M, Steinmetz N, Kadir S, Carandini M, D. HK (2016) Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* p. 061481.
- [84] Vincent JP, Economo MN (2023) Assessing cross-contamination in spike-sorted electrophysiology data. *bioRxiv*.
- [85] Trautmann EM, et al. (2019) Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron*.
- [86] Lakunina A, et al. (2025) Neuropixels opto: Combining high-resolution electrophysiology and optogenetics. *bioRxiv*.
- [87] Aharoni D, Hoogland TM (2019) Circuit Investigations With Open-Source Miniaturized Microscopes: Past, Present and Future. *Frontiers in Cellular Neuroscience* 13:141.
- [88] Peixoto D, et al. (2021) Decoding and perturbing decision states in real time. *Nature* pp. 1–6.
- [89] Boucher PO, et al. (2023) Initial conditions combine with sensory evidence to induce decision-related dynamics in premotor cortex. *Nature Communications* 14(1):6510.
- [90] Esparza J, Sebastián ER, Prida LMdl (2023) From cell types to population dynamics: Making hippocampal manifolds physiologically interpretable. *Current Opinion in Neurobiology* 83:102800.
- [91] Ji N (2014) The Practical and Fundamental Limits of Optical Imaging in Mammalian Brains. *Neuron* 83(6):1242–1245.
- [92] Lecoq J, Orlova N, Grewe BF (2019) Wide. Fast. Deep. Recent Advances in Multi-Photon Microscopy of in vivo Neuronal Activity. *The Journal of Neuroscience* pp. 1527–18.
- [93] Brooks FP, Davis HC, Wong-Campos JD, Cohen AE (2024) Optical constraints on two-photon voltage imaging. *Neurophotonics* 11(3):035007–035007.

- [94] Bliss-Moreau E, Costa VD, Baxter MG (2022) A pragmatic reevaluation of the efficacy of nonhuman primate optogenetics for psychiatry. *Oxford Open Neuroscience* 1:kvac006.
- [95] Lakunina AA, Nardoci MB, Ahmadian Y, Jaramillo S (2020) Somatostatin-Expressing Interneurons in the Auditory Cortex Mediate Sustained Suppression by Spectral Surround. *The Journal of Neuroscience* 40(18):3564–3575.
- [96] Ma Y, Hu H, Berrebi AS, Mathers PH, Agmon A (2006) Distinct Subtypes of Somatostatin-Containing Neocortical Interneurons Revealed in Transgenic Mice. *The Journal of Neuroscience* 26(19):5069–5082.
- [97] Hu H, Cavendish JZ, Agmon A (2013) Not all that glitters is gold: off-target recombination in the somatostatin-IRES-Cre mouse line labels a subset of fast-spiking interneurons. *Frontiers in Neural Circuits* 7:195.
- [98] Petersen PC, Siegle JH, Steinmetz NA, Mahallati S, Buzsáki G (2021) CellExplorer: A framework for visualizing and characterizing single neurons. *Neuron* 109(22):3594–3608.e2.
- [99] Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. (AAAI Press), pp. 226–231.
- [100] Traag VA, Waltman L, Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* 9(1):5233.
- [101] Kuhn M (2008) Building predictive models in r using the caret package. *Journal of Statistical Software* 28(5):1–26.
- [102] Vigneswaran G, Kraskov A, Lemon RN (2011) Large Identified Pyramidal Cells in Macaque Motor and Premotor Cortex Exhibit “Thin Spikes”: Implications for Cell Type Classification. *The Journal of Neuroscience* 31(40):14235–14242.
- [103] Kramer MA, Eden UT (2016) *Case studies in neural data analysis: A guide for the practicing neuroscientist*. (The MIT Press).
- [104] Aertsen AM, Gerstein GL (1985) Evaluation of neuronal connectivity: Sensitivity of cross-correlation. *Brain Research* 340(2):341–354.
- [105] Amarasingham A, Harrison MT, Hatsopoulos NG, Geman S (2012) Conditional modeling and the jitter method of spike resampling. *Journal of Neurophysiology* 107(2):517–531.
- [106] Xiao D, Forys BJ, Vanni MP, Murphy TH (2021) MesoNet allows automated scaling and segmentation of mouse mesoscale cortical maps using machine learning. *Nature Communications* 12(1):5992.

Acknowledgments

We are grateful to Dr. Karel Svoboda, Dr. Anna Lakunina, and Dr. Josh Siegle for the initial ideas of cell type classifiers for electrophysiology. We also thank Dr. Adam Smoulder, Dr. Yujin Han, Dr. Josh Siegle, Dr. Ching Fang, Pierre Boucher, Nicole Carr, Tian Wang, Vivan Moosmann, Tushar Arora, Mateo Umaguig, and Dr. Munib Hasnain for their thoughtful comments on our manuscript.

CC was supported by an NIH NINDS R00NS092972, R01NS121409, R21NS135361 and R01NS122969 award; the Moorman-Simon Interdisciplinary Career Development Professorship from Boston University; the Whitehall Foundation (2019-12-77); and the Young Investigator Award from the Brain and Behavior Research Foundation (27923). The auditory cortex dataset (collected by AL and SJ) was supported by an NIH NIDCD R01DC01553. SJ was also supported by an NIH NINDS RF1NS131993. EKL was supported by an NIH NINDS F31NS131018. The Neuropixels Ultras dataset was supported by NIH NINDS/NIMH U01NS113252 awarded to NS.

Author Contributions

EKL and CC jointly developed the multi-modal cell type classifier approach. AEG helped with derived waveform metric analysis and several control analyses. GH helped collect one of the open datasets and provided technical advice, especially with regard to the usage of quality metrics, and helped edit initial manuscript drafts. HY and CH provided many useful discussions, suggested analyses, and provided access to several datasets. AL and SJ provided technical support and also collected the S1 open datasets. AS, SO, and NAS provided technical support and collected the Ultras dataset. PFP provided guidance on various approaches for multi-modal integration and other technical advice. EKL curated datasets. CC wrote initial R code and performed some analyses. Code was further refined and augmented by EKL. Additional analyses were also written by EKL. EKL made figures and wrote initial drafts of the paper with CC. All authors edited the manuscript and provided feedback on presentation and clarity.

Competing Interests Statement

The authors declare no competing interests.

Study	n	Used modalities	Labeled cell types	Brain areas	Citation
Yu <i>et al.</i> 2019	246	Waveform shape, ISI dist., PSTH	PV ⁺ -4, PV ⁺ -5, SOM ⁺ , E-4, E-5	S1	https://doi.org/10.25378/janelia.8869115
Lakunina <i>et al.</i> 2020	373	Waveform shape, ISI dist.	PV ⁺ , SOM ⁺ , E	S1	https://github.com/sjara/jarapubs/tree/main/2018acsup/shared
Petersen <i>et al.</i> 2020 (Allen Institute Visual Coding and Buzsaki Lab)	430	Waveform shape, ISI dist., ACG, spike metrics	PV ⁺ , SOM ⁺ , E, Axo-axonic (PV ⁺), VGAT ⁺ (pan-inhib.), VIP ⁺	Visual cortex and CA1	https://cellexplorer.org/publicdata/ground-truth-data/ & https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html
Ye and Shelton <i>et al.</i> 2024	8,953	Waveform shape, 3D-ACG	PV ⁺ , SOM ⁺ , VIP ⁺	Visual cortex	https://www.biorxiv.org/content/10.1101/2023.08.23.554527v2
AIBS: Visual Behavior (Image set G and H)	5,582	Waveform shape, 3D-ACG	SOM ⁺ , VIP ⁺	Visual cortices	https://allensdk.readthedocs.io/en/latest/visual_behavior_neuropixels.html
Beau <i>et al.</i> 2025	3,301	Waveform shape, 3D-ACG	PkC _{SS} , PkC _{CS} , MLI, MFB, GoC	Cerebellum	https://www.c4-database.com/

Table 1: The seven datasets analyzed with the modalities used and the cell types identified in each. Two datasets are contained in⁹⁸.

Figure 1: Multimodal combination of physiological properties leads to structured representations that align with cell types. (A) UMAP on normalized average waveform shapes of neurons ($n = 246$) recorded *in vivo* juxtacellularly from mouse primary somatosensory cortex³². Each unit is colored according to one of eight ground truth cell types. Each modality was combined in different proportion according to the sum of each modality's per-unit contribution to the total; this total contribution is listed as a percentage below each scatter plot. (B) UMAP on ISI distributions of the same neurons in (A). (C) Each neuron in PhysMAP's WNN representation weighs each modality differently. These modal contributions are the sum of nearest neighbor edge modality weights (β_{Modality} in Fig. S1J) and are shown in their proportion of the per-unit WNN weight (pie charts) or across all units (inset histogram). (D) PhysMAP's combined multimodal representation obtained using the WNN approach³⁵. Again, each neuron is colored according to their ground truth cell type and layer. (E) Both modalities were concatenated into a single data vector per unit and passed into UMAP. This represents each modality in unweighted combination (as opposed to the WNN approach). Brain schematic in (A) adapted from¹⁰⁶ which is itself from the Allen Mouse Brain Atlas (<http://mouse.brain-map.org/>). Source data are provided as a Source Data file.

Figure 2: PhysMAP identifies cell types from juxtacellular recordings better than any modality alone. (A) Juxtacellular single unit recordings ($n = 246$) from mouse primary somatosensory cortex were collected under two modalities: juxtacellular waveform shape and inter-spike interval distribution in response to whisker deflection. Shown are two-dimensional PhysMAP visualizations of these units. Each unit is colored according to its ground truth cell type as deduced by optogenetic tagging in addition to layer information. The spike width (time from trough to peak), peak-to-trough ratio (ratio of trough and peak absolute amplitude), and spiking onset latency (trial-averaged time to half-peak in firing rate) for each neuron. Marker sizes shown are proportional in area to the log of the scalar value. (B) The PhysMAP projection of neurons with their ground truth cell type and layer (**left**) next to an example Leiden clustering¹⁰⁰ with resolution parameter set to 2 (**right**). (C) Leiden clustering is applied to the UMAP graphs of each modality alone (waveform [WF] or ISI dist. [ISI]), unweighted combination (concat.), or in weighted (PhysMAP) and shown with the associated modified adjusted Rand index (MARI;⁵⁰) across a range of resolution parameter values. Arrow marker indicates the clustering on PhysMAP used above in (B). Waveform metrics and "raw data" (i.e., without constructing a UMAP graph) were omitted because these do not yield a high-dimensional UMAP graph. (D) Each of the Leiden clusters located in (B) are shown broken down into their constituent cell types. Outer ring color indicates the weight of each constituent modality and inner pie chart denotes the relative proportions of underlying cell types within said cluster. Number at top right of each indicates the cluster in (B, right) each corresponds to. (E) Each of the Leiden clusters in (B, right) and (D) are shown with their respective normalized and smoothed peri-stimulus time histograms aligned to a whisker deflection ("touch" stimulus). Also labeled are the firing rate peak times of several Leiden clusters relative to the stimulus (dashed lines). (F) A gradient boosted tree model (GBM) classifier was trained on this multimodal data with 5-fold cross-validation on each modality's UMAP graph individually or on the multimodal WNN graph. The same classifier was also trained directly on the "raw data" (that is, without constructing a UMAP graph and projecting it; black line). The balanced accuracy performance of this classifier (mean \pm S.E.M.) on held-out data for each modality or the combined modalities for the five cell type classes with over 10 samples (units) each. This contained $n = 19, 58, 69, 44,$ and 19 cells of PV⁺-4, E-4, SOM⁺, E-5, and PV⁺-5 respectively. Brain schematic in (A) adapted from¹⁰⁶ which is itself from the Allen Mouse Brain Atlas (<http://mouse.brain-map.org/>). Source data are provided as a Source Data file.

Figure 3: **PhysMAP identifies cell types in extracellular recordings.** (A) PhysMAP applied to putative single unit ($n = 373$) extracellular recordings from mouse auditory cortex⁹⁵ using silicon multi-channel probes. We used waveform shape and ISI distribution as our modalities and neurons are colored according to cell type whose identities were obtained via optogenetic tagging. Marker size is set according to the increase in spike rate from optogenetic stimulation relative to baseline. Highlighted are a population of wide-spiking SOM^+ cells (**top**) and narrow-spiking SOM^+ cells (**right**). Excitatory cells are shown as orange x's because optogenetic modulation was negative and too small to see by size. (B) A GBM classifier with five-fold cross-validation was trained on the 10-dimensional graph of each modality or PhysMAP. The classifier was also trained directly on the raw (un-dimensionality reduced) multimodal data ($n = 373$) or simply waveform spike width to identify each optotagged cell type or untagged label ($n = 8, 32, 45,$ and 288 of each excitatory, PV^+ , SOM^+ , and untagged respectively). The balanced classification accuracy is shown (mean \pm S.E.M.) with many error bars smaller than marker size. (C) Normalized average waveforms in this dataset are passed into UMAP and their projection shown with optotagged cells colored and degree of optogenetic modulation (average change in spikes/s during stimulation epochs) shown by marker size. (D) Similarly, ISI distributions for these same neurons are also passed into UMAP and their projection shown. (E) Gaussian kernel probability density estimates for the distribution of each optotagged type across a range of spike widths is shown both as a histogram and with a kernel density estimator (solid line and shaded region). (F) Additionally, spike width is shown for all cells, including untagged ones. Source data are provided as a Source Data file.

Figure 4: **PhysMAP outperforms single-modality approaches and deep learning methods for inhibitory cell type identification.** (A) UMAP visualizations of mouse V1 neurons ($n = 8,953$) recorded with Neuropixels Ultra probes, shown using different dimensionality reduction methods: PhysMAP multimodal integration (**top left**), UMAP on 3D-autocorrelograms alone (**top right**), and UMAP on waveforms alone (WaveMAP, **bottom**). Cells are color-coded by optogenetic tagging: PV^+ (teal), SOM^+ (purple), VIP^+ (red), and untagged cells (gray). (B) Confusion matrices comparing classification performance of a GBM (raw accuracy) trained on the high-dimensional representations (all matched at 60 dims.) across three approaches: PhysMAP (**top left**), 3D-ACG UMAP (**top right**), and waveform UMAP/WaveMAP (**bottom left**). A multi-layer perceptron was used to classify cell types from the latent embedding of a pair of β -variational autoencoders (**bottom right**). Values represent the percentage of cells from each true cell type (rows) assigned to each predicted type (columns). Color intensity corresponds to classification accuracy. Highest true positive percentages per cell type across all approaches are highlighted by embossing. Brain schematic in (A) adapted from¹⁰⁶ which is itself from the Allen Mouse Brain Atlas (<http://mouse.brain-map.org/>). Source data are provided as a Source Data file.

Figure 5: **PhysMAP provides batch effect detection within and between modalities (A)** Two mouse visual cortex datasets visualized using PhysMAP: the Visual Behavior dataset (VB, green, recorded with Neuropixels 1.0; $n = 5,582$) and the Ultras dataset (Ultras, yellow, recorded with Neuropixels Ultra; $n = 8,953$). Diagrams show probe placements for each experiment. Waveform shape and 3D-autocorrelogram (3D-ACG) are combined in PhysMAP and experiments are shown in the same visualization. **(B)** Constituent modality-specific UMAP projections with waveform shape (**top**) and 3D-ACG modality (**bottom**). Units are again colored according to originating experiment (green for Ultras; yellow for VB). Horizontal dashed line is just to make clear that top and bottom are part of different modalities. **(C)** Waveform shapes and 3D-autocorrelograms from two subsets (different animals) of VB experiments, one containing an image set H (in red) and another an image set G (in blue). **(D)** The VB multimodal representation is split into its constituent modalities. This consists of waveform shape (**left**) and 3D-ACG (**right**). Brain schematics in (A) adapted from¹⁰⁶ which is itself from the Allen Mouse Brain Atlas (<http://mouse.brain-map.org/>). Source data are provided as a Source Data file.

Figure 6: **PhysMAP enables cross-dataset cell type identification through modality assessment and label transfer (A)** Cell type ground truth labels in the combined (VB+Ultras) dataset PhysMAP of waveform shape. Optotagged inhibitory interneurons are color-coded: PV^+ (teal), SOM^+ (purple), and VIP^+ (red) cells, with untagged cells in gray. **(B)** Inferred SOM^+ cells in the VB dataset (pink circles) based on nearest-neighbor label transfer with ground truth SOM^+ cells from the Ultras dataset with waveform shapes shown (pink in inset) on a background of all waveforms in the VB dataset (gray in inset). **(C)** Inferred PV^+ cells in the VB dataset (green circles) using the same approach also with waveform shapes shown (green in inset) on a background of all waveforms in the VB dataset (gray in inset). Source data are provided as a Source Data file.

Figure 7: **Label transfer allows inference of PV^+ and SOM^+ cell types in unannotated datasets. (A)** Simultaneous raster plots in response to a dark full-field flash (**top**), smoothed trial-averaged mean firing rates (spikes/s with SEM and 100 ms time bin Gaussian smoothing; **middle**), and autocorrelograms (ACG; **bottom**) for simultaneously recorded inferred PV^+ cell in visual areas VISal (green). Black overbar or vertical dashed lines indicate stimulus onset and offset times. Each raster's row is a simultaneous trial shared by all cells. **(B)** Simultaneously inferred SOM^+ cell in VISpm (pink, contemporaneously recorded with the PV^+ neuron in Fig. 7A) also with rasters, smoothed trial-averaged firing rates, and ACGs (**top**, **middle** [mean spikes/s with SEM and 100 ms time bin Gaussian smoothing], and **bottom** respectively). **(C)** PSTH's (mean normalized spikes/s with SEM) for all inferred PV^+ units (across experiments). **(D)** PSTH's (mean normalized spikes/s with SEM) for inferred SOM^+ neurons. Each trace consists of binned spike times (1 ms, non-overlapping bins) and is normalized by the maximum trial-averaged firing rate for said unit. Source data are provided as a Source Data file.

Figure 8: **Simultaneous monitoring of three interneuron cell types with PhysMAP reveals microcircuits during behavior.** (A) A simultaneously recorded optotagged VIP⁺ cell (red) during visual stimulus (black full-field flash) presentation with spike raster (**top**) and trial-averaged PSTH (mean spikes/s with SEM; **bottom**). (B) Cross-correlogram with the VIP⁺ cell spike times (target) onto a SOM⁺ (second SOM⁺ cell in A) cell's spike times (reference). Bins are 0.1 ms. Dashed line to indicate time bin zero. (C, **top**) Cross-correlograms for a putative PV⁺ cell onto the SOM⁺ cell in F (2 ms bin width). Dashed line indicates 0 ms bin. (C, **bottom**) Cross-correlograms for a putative PV⁺ cell onto the VIP⁺ cell in F (2 ms bin width). Dashed line indicates 0 ms lag. (D, **top**) Trial-averaged PSTH aligned to onset of black full-field flash (mean spikes/s with SEM). (D, **middle**) Average spike waveform of the putative PV⁺ neuron inhibiting both the SOM⁺ and VIP⁺ cells. (D, **bottom**) 2D-autocorrelogram of the putative PV⁺ cell with peak in firing rate at 67 Hz (2 ms time bins). Source data are provided as a Source Data file.















