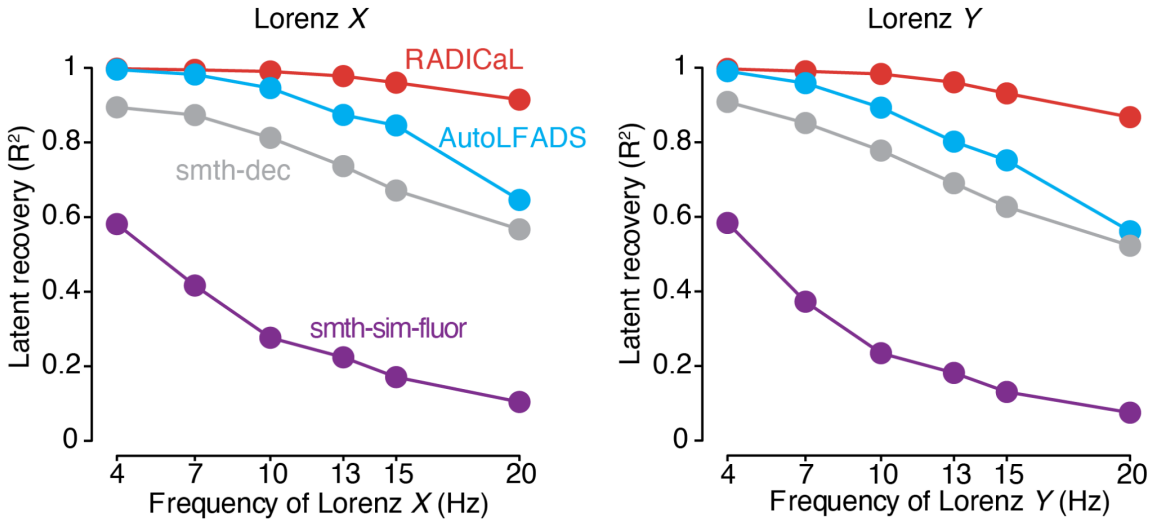
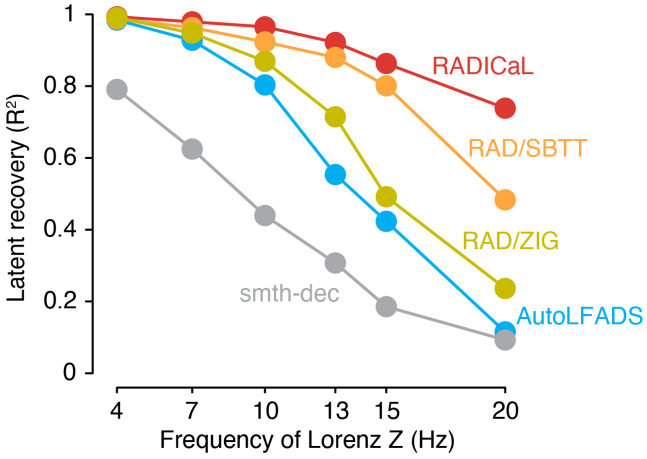
			LR	CD	DO	KL CO	KL IC	L2 Con	L2 Gen	Sig Prior
Defaults	RADICaL	Ranges	(1e-5, 5e-3)	(0.01, 0.99)	(0.3, 1.0)	(1e-6, 1e-4)	(1e-6, 1e-4)	(1e-5, 0.1)	(1e-5, 0.1)	(1.0, 100.0)
		Initial values	1e-3	0.5	uniform	loguniform	loguniform	loguniform	loguniform	20.0
		Explore weight	0.3	0.3	0.3	0.8	0.8	0.8	0.8	0.2
		Limit explore	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
	AutoLFADS	Ranges	(1e-5, 5e-3)	(0.01, 0.99)	(0.3, 1.0)	(1e-6, 1e-4)	(1e-6, 1e-4)	(1e-5, 0.1)	(1e-5, 0.1)	-
		Initial values	1e-3	0.5	uniform	loguniform	loguniform	loguniform	loguniform	-
		Explore weight	0.3	0.3	0.3	0.8	0.8	0.8	0.8	-
		Limit explore	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	-

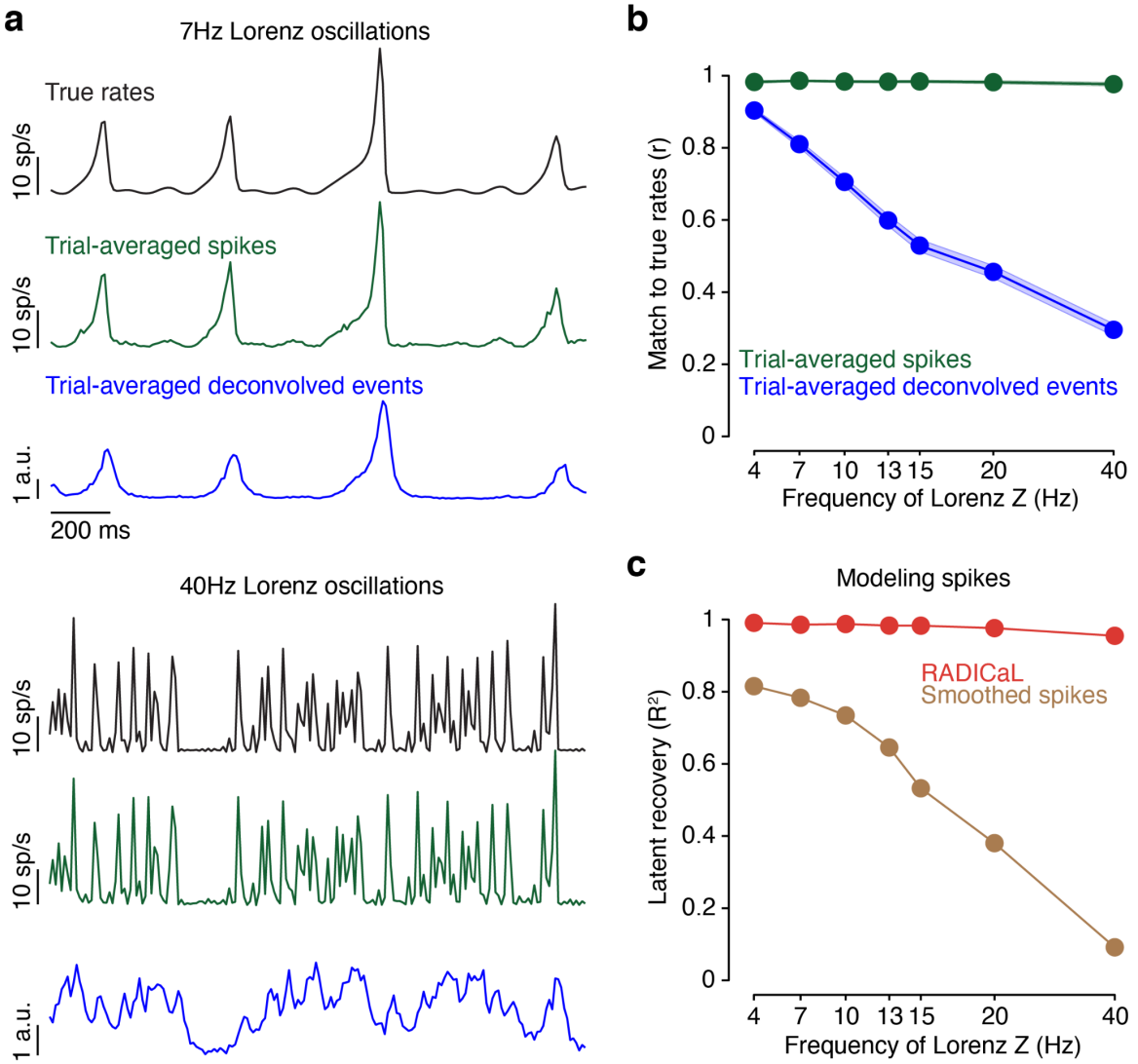
Supplementary Table 1. Hyperparameter ranges for RADICaL and AutoLFADS runs. Cells with a dash indicate "not applicable" for the method. LR is the learning rate. CD is the coordinated dropout rate (i.e., proportion of samples dropped at input). DO is the dropout probability for the RNN network. KL indicates the weight applied to the KL divergence of a posterior from its prior. CO indicates the controller output distributions and IC indicates the initial condition distributions. L2 indicates the weight applied to the Frobenius norm of the recurrent kernel of the GRU cell. Con indicates the controller GRU cell, Gen indicates the generator GRU cell, IC Enc indicates the initial condition encoder GRU cells. Sig Prior indicates the prior of the scaling factors applied to the sigmoid nonlinearity when mapping from factors to ZIG parameters.



Supplementary Figure 1 | Performance of estimating other Lorenz dimensions in the simulation experiments. Performance of estimating Lorenz *X* (*left*) and *Y* (*right*) dimensions as a function of simulation frequency was quantified by variance explained (R^2) for all 4 methods. Note that these variables are dominated by lower frequencies than the *Z* variable used in other figures, and therefore make for an easier challenge. We therefore used the *Z* variable for all other results.

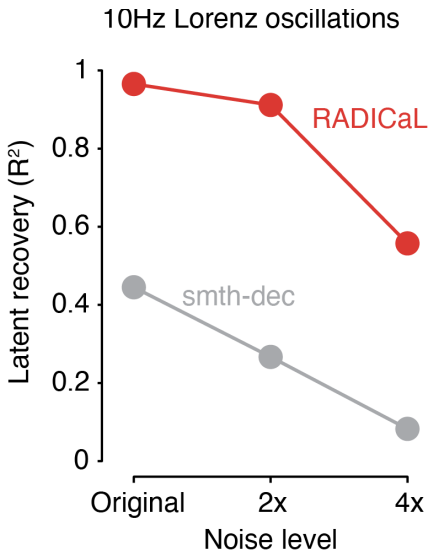


Supplementary Figure 2 | Both SBTT and ZIG improve latent recovery performance separately. To understand the contributions of ZIG and SBTT independently in RADICaL's performance in latent recovery, we fit RADICaL to different Lorenz oscillation frequencies with only the ZIG emission model enabled (no SBTT; "RAD/ZIG") or only SBTT enabled (no ZIG; "RAD/SBTT"). Performance in estimating the Lorenz *Z* dimension as a function of Lorenz oscillation frequency was quantified by variance explained (R^2). RAD/ZIG performed a little better than AutoLFADs, while RAD/SBTT performs substantially better, but combining both (the full RADICaL model) performed substantially better still.

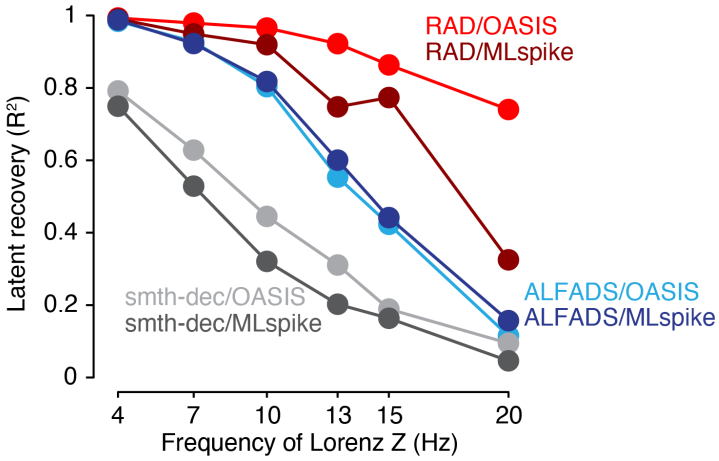


Supplementary Figure 3 | Deconvolution places an upper bound on RADICaL's performance recovering higher-frequency features. To understand how deconvolution performs across Lorenz oscillation frequencies, we measured how well trial-averaged deconvolved events captured the true underlying rates for individual (simulated) neurons. Averaging deconvolved events across the noisy repeated trials that have the same true underlying rates is a straightforward way to test, on average, whether deconvolution irreversibly loses information about the underlying rates. There were three main steps in the rates-to-events generation process: Poisson sampling of spikes from the underlying rates, fluorescence generation and sub-sampling, and deconvolution (detailed in *Methods*). To specifically isolate the effect of fluorescence generation and deconvolution on rate recovery, we also tested recovery with those steps omitted, i.e., spikes generated in the rates-to-events process were sub-sampled from a sampling frequency of 100 Hz to 33.3 Hz as was done for the fluorescence traces, and were averaged across trials to quantify how well they captured the true underlying rates. (a) Example ground truth firing rates, averaged spikes across trials (3000 trials), and averaged deconvolved calcium events across trials (3000 trials), for Lorenz oscillation frequencies of 7Hz (*top*) and 40Hz (*bottom*). (b) Performance in capturing the ground truth firing rates as a function of Lorenz oscillation frequency was quantified by correlation coefficient (r) between the trial-averaged spikes or deconvolved events and the true rates. Error bars indicate the variability across simulated neurons. The correlation between the trial-averaged deconvolved events and the true rates dropped as the Lorenz oscillation frequency increased, suggesting that deconvolution fails at higher Lorenz oscillation frequencies. The correlation between the trial-averaged spikes and the true rates did not drop as the Lorenz oscillation frequency increased, suggesting that the drop seen for the deconvolved events was mainly due to deconvolution and not the Poisson sampling and sub-sampling steps. (c) To determine whether RADICaL's performance loss for high-frequency signals was purely due to deconvolution failure or might involve limitations of the model itself, we eliminated the fluorescence generation/deconvolution step and applied RADICaL directly to the sub-sampled spiking activity. In this test, we did not use RADICAL's ZIG observation model, but kept the SBTT approach and used a Poisson observation model. Performance in using ground truth spikes to estimate the Lorenz Z dimension as a function of Lorenz oscillation frequency was quantified by variance explained (R^2) for smoothing and RADICaL. RADICaL retained high performance in latent recovery across Lorenz oscillation frequencies from 4Hz to 40Hz, whereas smoothing showed a much faster degradation of latent recovery performance. Together, these

analyses demonstrate that the degradation in RADICaL's performance at higher Lorenz oscillation frequencies is mainly due to inaccuracies in deconvolution, and not due to the model itself.

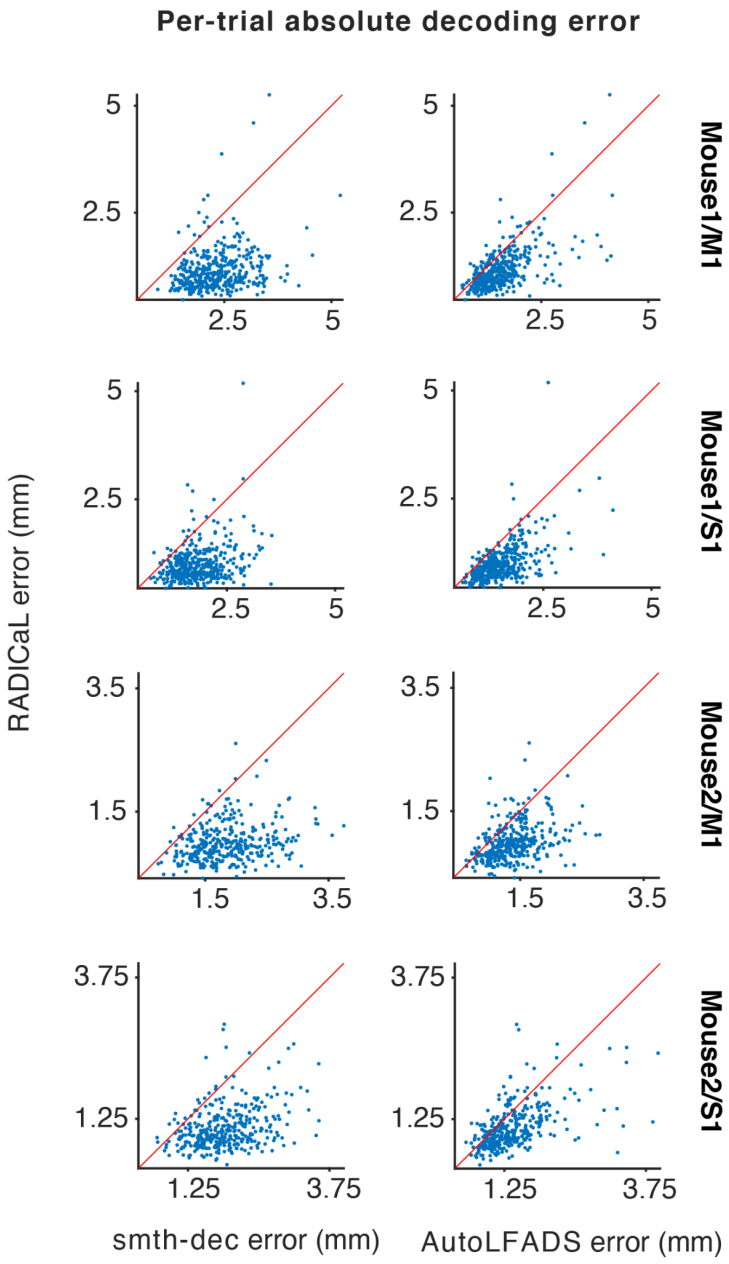


Supplementary Figure 4 | Model tolerance to spike inference noise. In our simulations, we chose parameters so that the resulting signal-to-noise regime produced similar correlations between real and inferred spike trains as observed in a recent benchmarking study[13] (see *Methods*). However, the spike inference noise can vary in real experiments and could affect RADICaL's performance. To test how larger spike inference noise affects the performance of RADICaL and smth-dec, we raised the level of the Gaussian noise used in generating simulated fluorescence traces by 2x or 4x. Performance in estimating the Lorenz *Z* dimension as a function of the level of spike inference noise was quantified by variance explained (R^2) for RADICaL and smth-dec. Performance declined for both methods as the noise level increased. However, RADICaL retained high performance at the 2x noise level (R^2=0.91) and reasonable performance at the 4x noise level (R^2=0.56). Smth-dec had low performance across the board (R^2=0.27 and 0.08 for 2x and 4x noise, respectively). Notably, RADICaL performed better at the 4x noise level than smth-dec at the original noise level.

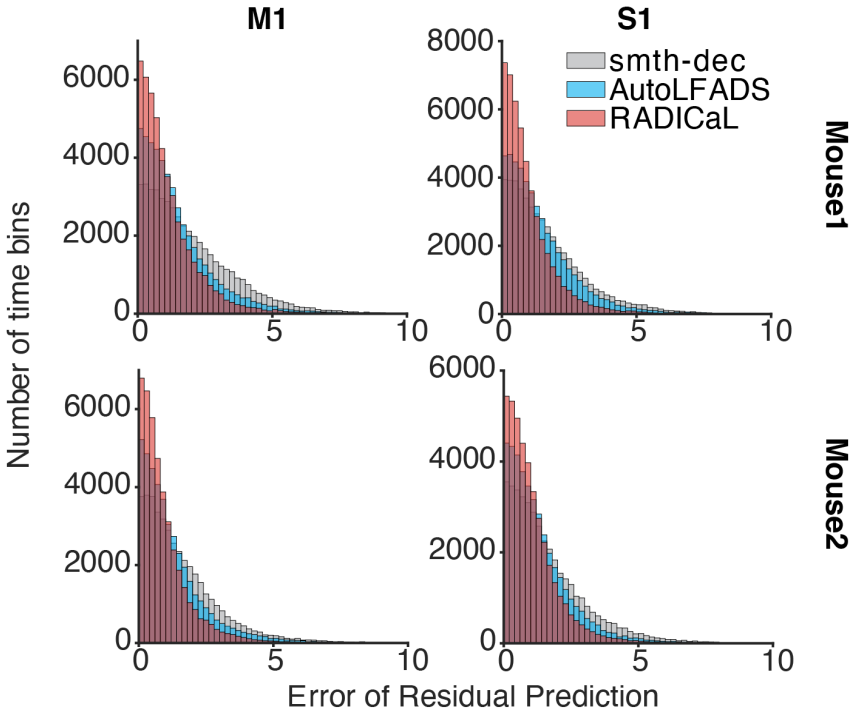


Supplementary Figure 5 | RADICAL (with SBTT) improves latent recovery when using spikes inferred by MLspike, but does not perform as well as when using OASIS for deconvolution. To test whether RADICaL could be effective with deconvolution algorithms that infer spike times instead of event rates, we analyzed simulated data that had spike inference performed with MLspike[27]. Performance in estimating the Lorenz *Z* dimension as a function of Lorenz oscillation frequency was quantified by variance explained (R^2) for six methods. These included three methods in which the inputs were deconvolved events from OASIS: RADICaL ("RAD/OASIS"), AutoLFADS ("ALFADS/OASIS") and smoothing ("smth-dec/OASIS"); and three methods in which inputs were spikes inferred with MLspike: RADICaL ("RAD/MLspike"), AutoLFADS ("ALFADS/MLspike") and smoothing ("smth-dec/MLspike"). When pairing RADICaL with deconvolution methods that produce spike times as output, we can use a Poisson observation model (as one would use for spikes measured via electrophysiology) instead of ZIG, while retaining the SBTT approach for sub-frame sampling. RADICaL with a Poisson observation model (RAD/MLspike) was able to model MLspike output, and substantially outperformed AutoLFADS and smth-dec (ALFADS/MLspike and smth-dec/MLspike), but did not perform as well as RADICaL applied to OASIS-

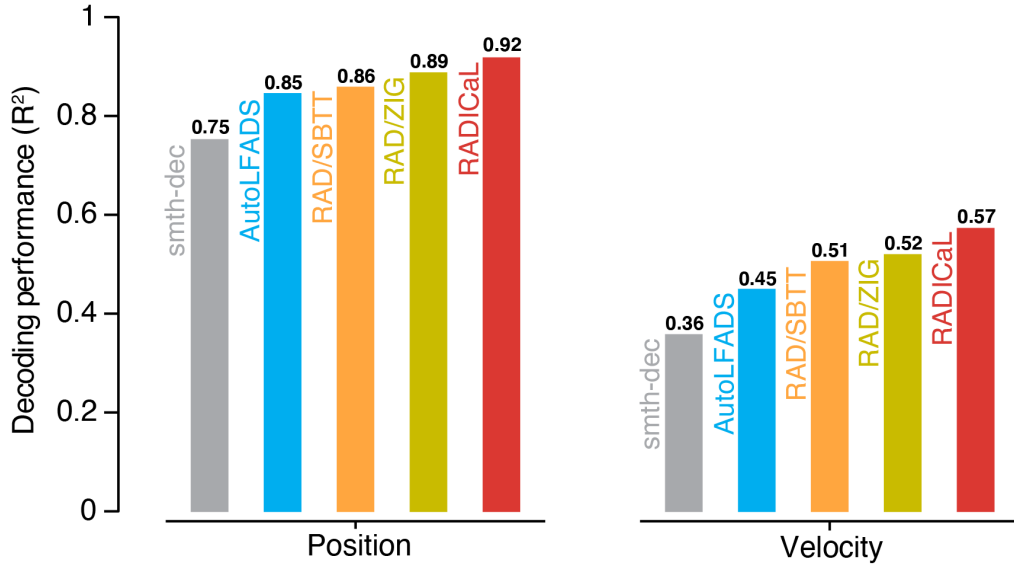
deconvolved events (RAD/OASIS). In addition, the parameter tuning required for MLspike is more involved and requires more expertise than OASIS (see *Methods*). Therefore, we recommend using OASIS as the deconvolution method for RADICaL.



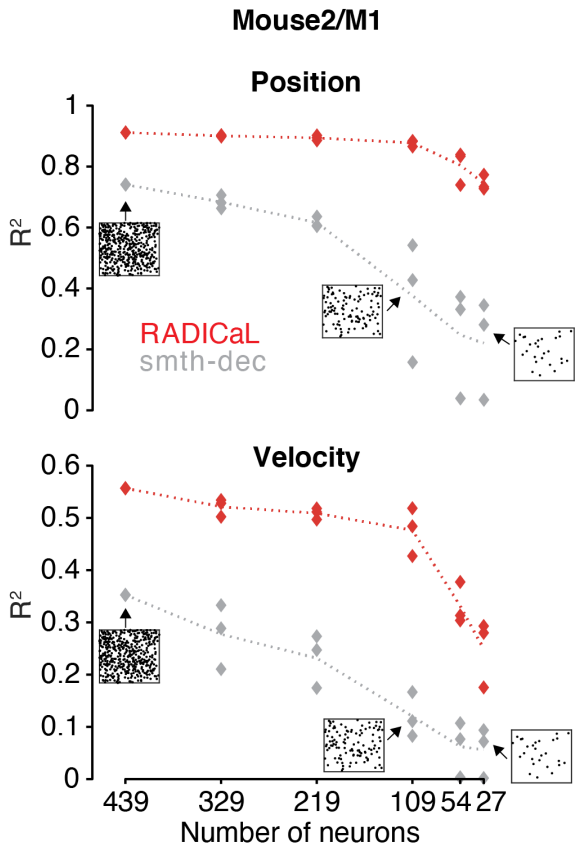
Supplementary Figure 6 | RADICaL reduces decoding errors on the vast majority of single trials for all datasets. Single-trial decoding error was quantified by measuring the absolute difference between the true and decoded hand position for each individual trial. Each point represents an individual trial. Error was greatly reduced compared with both smth-dec (*left*) and AutoLFADS (*right*).



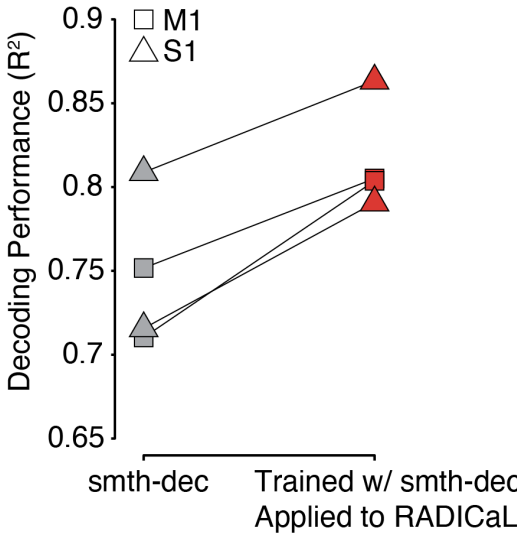
Supplementary Figure 7 | RADICaL improves prediction of single-trial deviations from the mean of hand positions. This figure demonstrates that RADICaL does not simply learn a 'typical' trajectory for left-reach trials and another for right-reach trials, but instead reflects small deviations from the condition average better than other methods. The residuals of hand positions (i.e., single-trial deviations from the mean) were computed by subtracting the left-reach or right-reach trial-averaged hand positions from the single trials. Error of residual prediction was computed by taking the absolute value of the difference between true and predicted residuals of hand positions.



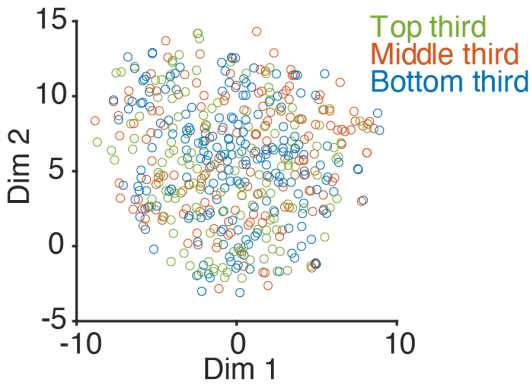
Supplementary Figure 8 | Performance of ZIG-only (A-ZIG) and SBTT-only (A-SBTT) on decoding hand kinematics. To test whether the innovations of RADICaL contributed separately to the improved decoding performance, we performed an ablation study where we enabled solely the ZIG emissions model (RAD/ZIG) or SBTT (RAD/SBTT). Decoding accuracy was quantified by measuring variance explained (R^2) between the true and decoded position (*left*) and velocity (*right*) across all trials, for RAD/ZIG, RAD/SBTT and other techniques. Analyzed data are from Mouse2/M1. Note that in this test, RAD/ZIG outperformed RAD/SBTT, which is the opposite of the results from synthetic data shown in **Supp. Fig. 2**. The discrepancy could potentially be due to different properties of the datasets, such as the frequency of the underlying features or noise properties. However, for both simulated and real data, either innovation (SBTT or ZIG) helps improve performance and combining them yields the highest performance.



Supplementary Figure 9 | RADICaL is robust to the random seed used in selecting subsets of neurons in a neuron downsampling experiment. This figure is related to Figure 5. Decoding performance measured using variance explained (R^2) as a function of the number of neurons used in each technique (*top*: Position; *bottom*: Velocity). For a given number of neurons (except the full population of 439 neurons), 3 random seeds were used, and each data point represents an individual random seed. The dotted line represents the mean performance across the three random seeds for each method. Data from Mouse2/M1. Figure insets indicate the selected neurons in the FOV for experiments of the full population and example subsets of the population.



Supplementary Figure 10 | RADICaL improves decoding performance using decoders trained with smth-dec rates. This analysis demonstrates that the decoding performance benefit due to RADICaL cannot be due to training a better decoder alone, but results from better denoising of the trajectories themselves. Decoding performance was quantified by measuring variance explained (R^2) between the true and decoded hand position across each of the 4 datasets (2 mice for M1, denoted by squares, and 2 mice for S1, denoted by triangles), for decoders trained with smth-dec rates and applied to smth-dec rates (gray) or applied to RADICaL rates (*red*).



Supplementary Figure 11 | Visualization of transformation from factors to neurons. This analysis demonstrates that the different bands of the image use the same factors and not segregated ones, despite being divided up into separate sub-bins for improving temporal resolution with SBTT. The plot shows a 2-dimensional t-SNE space representation of weights mapping from RADICaL factors to ZIG parameters for Mouse1/M1. Each point represents an individual neuron (510 neurons total). Neurons are color coded based on the neurons' position within the field of view (i.e., top, middle, and bottom). The interspersal of the points shows that neurons do not have systematically different relationships with the factors in RADICAL based on which band they are in.