Experimental technologies to record the activity of many neurons at the same time in different species and brain areas are rapidly advancing. These experimental advancements are paralleled by advances in theoretical and computational methods for analyzing the data accumulated using the recording technologies. Such theoretical methods usually take the form of probabilistic models that try to describe the multi-neuronal activity of the recorded neurons. With such probabilistic models one aims to address numerous issues: What correlations are important in describing the multi-neuronal pattern? How does the pattern of activity covary with external stimuli or experimental conditions? What dimensionality does the neural data live in and how is this related to the underlying network interactions? The probabilistic models can also be used to make predictions about the structure of the neural code, by studying the properties of the fitted model, or by generating synthetic data from it.

In general, despite the rapid advances in recording technology, the best experimental measurements of neuronal activity still only provide data from a small subset of neurons that comprise a neuronal network. A wealth of studies on building probabilistic models of neural data focuses on describing such subsets, ignoring the fact that the observed neurons is a small group in a much bigger set of hidden neurons. Some other studies do include hidden variables which, amongst other things, aim to model the global features of the unrecorded neurons, but we still lack a through understanding of how to include the role of hidden neurons in probabilistic models, how our inferences about the recorded activity would be affected by them, and what we can we say about the rest of the network by studying the heavily subsampled recordings. In this paper, we aim to address these questions in the case of a simple maximum entropy model, namely the homogeneous maximum entropy model.

The maximum entropy approach has been used in a variety of setting for building statistical models of complex systems and datasets, ranging from neuronal activity in the retina, in the cortex, protein sequences, gene regulatory networks and natural images. The general idea is, for a dataset, to write down the distribution that maximizes the entropy of state variables, given some low order statistics. Now given the fact that the recorded neurons are a fraction of the neurons in the network, several quantitative questions arise that we will address in this paper: given the data from the sampled neurons, can we build a maximum entropy model over the whole network? Once we build such a network level maximum entropy model, can we see features in the neural activity which cannot be directly seen from a model build from the sampled neurons? Since we can marginalize down the network level maximum entropy model to the sampled network, how does this margianzlied maximum entropy model match the sample level model?

All these questions can be answered in the case of the homogeneous maximum entropy model. First we show how to go from the sample level maximum entropy level to the network level, by assuming different sizes of the network and also by assuming an uninformative prior over the size of the network. This is done by inferring the statistics of correlation functions at the network level from those of the sample level by using simple counting arguments. We then find that, when applied to experimental recordings from the Medial Entorhinal Cortex of rats and the monkey visual cortex, this network level maximum entropy model may exhibit features that the sample level model does not predict. Specifically, we observed modes in the distribution of the activity in the network level model that do not show up in the sample level. We study how the assumed size of the full network affects the appearance of these modes and find that there is a minimum size of the full network for which such modes can be observed. We then compared the distribution found by marginalizing the full network maximum entropy model down to the sample level, and the distribution fit directly to the sample level. For the two datasets that we tested, we found that the two distributions match each other to a large degree but that there are also differences between them. We quantify how these differences also depend on the assumed size of the network and find that …(WE SHOULD TEST SHI) This predicts that for a large enough population (DO WE PREDICT THAT IF THE FULL NETWORK GETS BIGGER THE DIFFERENCE ALSO GET BIGGER)?…

The rest of the paper is organized as follows. We first describe how to go from the sample level maximum entropy model to the full network maximum entropy model. In section 2, we apply this to the two experimental datasets and study the effect of the assumed size of the network as well as the moments that we use for building the maximum entropy model. In section 3 we compare the distributions found from marginalizing the maximum entropy model down to the sample level and the original sample level model.