<u>Motivation Letter</u>

Maria Kesa

Neuroscientists can now record circuit activity from tens of thousands of neurons simultaneously, both locally using two photon calcium imaging and in multiple regions through neuropixel probes. There are theories in the air, long-standing debates about how the brain works. If a theory is about how neurons collectively process information in a population code, you may need to record from a large fraction of the circuit of interest or across several brain areas simultaneously because recordings from single neurons or a small local patch of neurons may be biased and fail to generalize (Stringer, Pachitariu et al, 2019a, 2019b) . But even if you have managed to record from a large population you still face formidable challenges. This data is just a mass of numbers. How do you connect your ideas to objective reality, to data or develop new data-driven theories? There has to be a way to explore the data through data analysis and visualization algorithms, ideally using interactive user interfaces. A dataset without algorithms and systems to process it is like a door without a key. I am applying for a PhD in Neuroscience because I would like to discover new things about the brain by developing machine learning algorithms to quantify behavior and large-scale neural activity.

In fact, this type of work is exactly what I am doing in my current job. I work as a software consultant at HHMI Janelia. My previous jobs involved developing and programming machine learning algorithms for the analysis of high-dimensional data and probabilistic modeling of gene expression and MRI data, as well as deep learning to satellite images. As I had a profound interest for brain research my job at Janelia is ideal as it involves the analysis of massive recordings of tens of thousands of neurons from the mouse primary visual cortex along with associated behavioral movies. The dataset is perfect for applying machine learning techniques and benchmarking algorithms.

In my master's thesis my aim was to detect cell assemblies as patterns of co-activating cells from the recordings and use this information to decode the stimulus identity. I applied an algorithm from natural language processing to open calcium imaging data from Allen Brain Observatory. The open dataset that I work on now allowed me to apply the approach to neural big data. However, the algorithm that I used in my thesis was too slow for interactive visualizations, so we developed a simplified, approximate version that adds an L0 sparsity penalty to the columns of a matrix factorization algorithm. This forces the algorithm to learn compact groups of cells that show co-activation in their temporal dynamics in response to various stimuli. I've been working on this approach for the past seven months: coding up the algorithm in Python, applying it to data, comparing it to other algorithms, decoding stimuli from neural recordings, extracting receptive fields of groups of cells found by the algorithm, coming up with ways to make the algorithm faster, finding mappings between behavioral and neural data, studying how ensembles encode sensory and behavioral information and developing a graphical user interface to visualize the components learned by the algorithm. I presented an initial version of our work at the poster session at the Carnegie Mellon workshop on Statistical Analysis of Neural Data. The github repository of the project where the poster can also be found is located at https://github.com/MouseLand/kesa-et-al-2019 and https://github.com/MouseLand/EnsemblePursuit

The next logical step for me is to pursue a PhD degree where I can work on similar types of problems to push my skills and work to perfection. I am particularly interested in Alex and Mackenzie Mathis' laboratories. I wrote to Dr. Mackenzie and linked to my Github and poster. She said that I have the perfect profile for their laboratories. They are working on meso-scale recordings in the motor cortex to develop and test theories about sensory-motor loops. For example, they have developed a deep learning-based system for quantifying animal pose information, *DeepLabCut*. In this regard, I have experience in convolutional neural networks from an internship in satellite image analysis at Starlab Barcelona that could be useful to analyze behavioral image and video data. Further, I think that my experience at Janelia – where I worked with behavioral and neural data – is a stepping-stone and a great asset to continue developing useful algorithms and systems for the analysis of massive neural and behavioral data sets. Quantifying behavior opens up exciting possibilities for research, as it makes it possible to find statistical mappings between behavior and neural activity.

On the theoretical side, I have experience with simulations of neural systems. I did a project with professor Berry from Princeton on rate-based simulations of cortical circuits which got selected as an example project for future students of the summer school where I participated. I took an online course by professor Gerstner on Neuronal Dynamics. I enjoyed it so much that I also participated in an EPFL workshop “From Synaptic Plasticity to Motor Control” in 2017.

I have been very fortunate to have been selected to participate in several summer schools in neuroscience and machine learning with scholarships. For example, I got scholarships from Princeton, Berkeley, Okinawa Institute of Science and Technology, Tampere University and the Human Brain Project. Lately, I also participated in the Neural Data Challenge at Bristol University where I applied a Hidden Markov Model to mouse behavioral data. This could be a first step towards what I could be doing in my PhD. There are long-standing theories about animal behavior that roughly posit that complex behaviors are composed of movement syllables (Wiltschko et al, 2015). I could begin my PhD by applying HMM's to find evidence in the data for these kinds of structures and try to link the movement syllables from quantified behavior to neural activity. In particular, are representations of movement in the motor cortex continuous or discrete? Are movement syllables composed of neural syllables? I would like to base my work and research in the two complementary directions: going from theories to data and going from data to theories. But I would also like to build algorithm-based visualizations to observe patterns in the data and perhaps come up with my own data-driven theories. Pursuing a PhD at EPFL would help me mature as a scientist and conduct exciting work bridging data-driven neuroscience with machine learning. And that is my passion.