Automatic Channel Selection and Neural Signal Estimation across Channels of Neural Probes

Olga Vysotska Barbara Frank Istvan Ulbert Oliver Paul Patrick Ruther Cyrill Stachniss Wolfram Burgard

***Abstract*— High-resolution microprobes are used to record single neuron activity in the brain. This technology is envisaged to be a central component for brain-controlled computers and robots. Current neural probes, however, allow for recording only a small number of the densely spaced electrodes simultaneously. Therefore, we address the problem of autonomously choosing, for a given number, the subset of electrodes with the corresponding size so as to extract as much information as possible. We first present an approach for predicting neural spikes across different channels of the probe. Our method employs nonparametric sparse Gaussian process regression to predict the signal of a channel given the signals recorded at neighboring sites. Second, we utilize the signal predictions for efficiently seeking for the subset of electrodes that minimizes the overall prediction error. In experiments carried out using real neural data, we demonstrate that our selection procedure provides highly accurate results. Furthermore, the solutions found in our experiments are close to the optimal solution.**

I. Introduction

The development of microprobes [5], [12], which can record neural activity with high resolution, opens new perspectives in brain research as they allow for recording the activity of single neurons and thus enable a better understanding of neural processes and interactions between brain regions. They are also envisaged to be central components for brain-controlled computers and robots. Further applications include smart energy-autonomous micronodes that provide closed-loop feedback—recording and stimulation—for neurological diseases such as epilepsy or Parkinson's disease.

Modern probes [1], [4], [5], [12] are often high-density probe arrays with a large number of available electrodes. The advantage of such probes is that they enable scanning of different brain layers for neural activity as illustrated in Fig. 1. Each electrode, also denoted as channel, of a probe can record the activity of neurons in its local vicinity. Due to hardware and wiring restrictions, probes typically can only transmit signals from a small subset of electrodes at the same time. In our probes, 188 closely neighboring electrodes are available from which we can choose a subset of eight electrodes for simultaneous recording. A typical recording experiment would first scan through all available electrodes for interesting signals and select the most informative ones for long-term observations. Thus, a central question in the design of an experiment is how to select the channels to be recorded. Due to the high resolution of the probes, channels typically measure redundant signals as the activity of one neuron is often also observable at different close-by electrodes. At the same time, the varying quality of the signals and the sheer amount of data recorded by the individual electrodes make the manual selection of recording channels a tedious task for a neuroscientist. Hence, our goal is to provide an efficient algorithm for selecting recording channels in an unsupervised fashion.

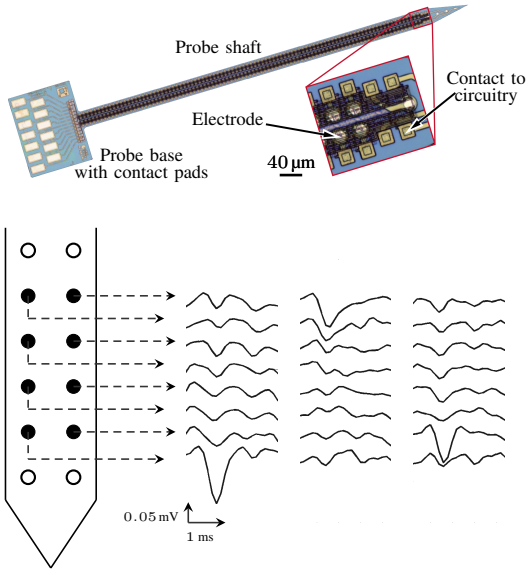

Fig. 1. Neural microprobe (top) and schematic sketch of the probe with different recording sites and typical recorded signals of neural activity (bottom).

This paper makes two main contributions in the context of recording and analyzing neural data. First, we present an approach for predicting the neural signal for a channel that is *not* recorded by considering the measured signals on neighboring channels. We address this task using sparse Gaussian process regression. Second, we present a method to identify, which of the channels to record in order to get the best possible neural data prediction on all channels from a global perspective. The computational complexity of our approach is $\mathcal{O}(n^2 \log n)$, where n is the number of channels on the probe. We furthermore present an experimental evaluation

This work has partially been partly supported by the German Research Foundation under grant number EXC 1086, by the Hungarian Scientific Research Fund OTKA K83251 and OTKA 81357, National Office for Research and Technology NKTH-ANR, Neurogen OMFB-00294/2010 and Multisca TET_10120110389, by the FP7 of the European Commission under grant NeuroSeeker No. 6000925 and TAMOP4.2.1.B11/2/KMR2011-0002.

All authors except I. Ulbert are with the University of Freiburg, Faculty of Engineering, 79110 Freiburg, DE. C. Stachniss is also with the University of Bonn, Inst. of Geodesy and Geoinformation, 53115 Bonn, DE.

I. Ulbert is with the Inst. of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Science, Budapest, Hungary and the Faculty of Information Technology and Bionics, Pter Pzmny Catholic University, Budapest, HU.

using real neural data, which demonstrates that our method is effective for recording and predicting neural signals. On our data, our method always selects near-optimal channels.

II. Related work

To the best of our knowledge, the problem of automatic channel selection for neural probes has received little attention thus far. Seidl et al. [14] propose a semi-supervised selection approach that computes the signal-to-noise ratio (SNR) at different recording sites and in this way supports the user in the manual decision process. Van Dijck et al. [3] present an approach to automated electrode selection and electronic depth control, that is, compensating probe shifts by shifting the recording channels. Similar to Seidl et al. [14], they compute a quality measure for each electrode. They iteratively select electrodes based on their SNR but penalize this measure by taking into account the similarity of a captured spike train to the spike trains of already selected electrodes. This leads to an overall better distribution of recording electrodes, as demonstrated on a simulated neuronal model with five neurons. In contrast to our work, they only take into account the co-occurrence of spikes at different channels, while our approach allows for a prediction of the spike signal at different channels and uses this information to perform a fully automatic channel selection.

Additionally, different machine learning techniques have been applied to analyze signals acquired by neural probes, for instance, to detect and classify observed spikes according to different neurons in the vicinity of a dense microelectrode array [7], [9]. Fraser et al. [6] describe an approach to track neurons over time. Recently, Van Dijck et al. [2] proposed an approach to recognize cerebellar cortical neurons across species using Gaussian process regression based on statistics of measured inter-spike intervals. In this paper, we apply sparse Gaussian processes for predicting signals of recording channels based on data recorded at other channels. We furthermore employ resulting Gaussian process models for selecting the most informative channels of the entire probe. Fig. 2 illustrates typical prediction results of our method.

III. Predicting neural signals across channels using GP regression

In the remainder of this paper, we first describe our approach to predict the signals of individual channels based on measurements at neighboring recording sites. Second, we present our selection procedure that makes use of these predictions and determines a subset of recording channels that minimize the overall prediction error.

We use a neural probe with in general n channels. The hardware restrictions of the neural probe, however, allow to record only b channels at the same time. Since we are interested in monitoring the neural activity on all channels, we aim at predicting the signal in the channels we cannot record based on the ones we can record.

Our first goal is to predict the signal of all neural probe channels that cannot be recorded. We refer to the channels that are recorded as *active* and the others as *passive* channels. Given neural measurements $y^{c}_{1:t}$ recorded on at least two channels $c', c_1, \ldots, c_b$ ($b > 0$) and acquired at time indices $1, \ldots, t$, we learn a predictive model $p(y^{c'}_{t':t'+t} \mid y^{c'}_{1:t}, y^{c_{1:b}}_{1:t}, y^{c_{1:b}}_{t':t'+t})$ for estimating the neural signals $y^{c'}_{t':t'+t}$ on the passive channel c' given the neural signals $y^{c_{1:b}}_{t':t'+t}$ from the active channels $c_1, \ldots, c_b$ at time $t', \ldots, t'+t$ with $t' > t$.

We approach this problem in a nonparametric way, i.e., not assuming a parametric form of the underlying function $f(\cdot)$ in $y^{c'}_i = f(y^{c_{1:b}}_i) + \epsilon$. In particular, we use the Gaussian process model [10], which is a Bayesian approach to the non-linear regression problem. In this approach, one places a prior $p(f)$ on the space of functions by assuming that a Gaussian process is a collection of random variables, any subset of which having a joint Gaussian distribution. More formally, if we assume that $\{(y^{c_{1:b}}_i, f_i)\}^t_{i=1}$ with $f_i = f(y^{c_{1:b}}_i)$ are samples from a Gaussian process and define $\mathbf{f} = (f_1, \ldots, f_t)^\top$, we have

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}) \ , \quad \boldsymbol{\mu} \in \mathbb{R}^t, \mathbf{K} \in \mathbb{R}^{t \times t}. \tag{1}$$

For simplicity of notation, we assume $\boldsymbol{\mu} = \mathbf{0}$, since the expectation is a linear operator and, thus, for any deterministic mean function $m(x)$, the Gaussian process over $f'(x) = f(x) - m(x)$ has zero mean.

The more interesting part of the model is indeed the covariance matrix $\mathbf{K}$. It is specified by $[\mathbf{K}]_{ij} = \mathrm{cov}(f_i, f_j) = k(y^{c_{1:b}}_i, y^{c_{1:b}}_j)$ using a *covariance function* k which defines the covariance of two function values. Intuitively, the covariance function specifies how similar the two function values are depending only on the corresponding inputs. The standard choice for k is the squared exponential covariance function

$$k_{SE}(y^{c_{1:b}}_i, y^{c_{1:b}}_j) = \sigma_f^2 \, \exp\left(-\frac{1}{2}\frac{|y^{c_{1:b}}_i - y^{c_{1:b}}_j|^2}{\ell^2}\right), \tag{2}$$

where the so-called *length-scale* parameter ℓ defines the global smoothness of the function f and σ_f^2 denotes the amplitude (or signal variance) parameter. These parameters, along with the global noise variance σ_n^2 that is assumed for the noise component, are known as the *hyperparameters* of the process. They are denoted as $\boldsymbol{\theta}_{SE} = \langle \sigma_f, \ell, \sigma_n \rangle$.

In our approach, we use a variant of the squared exponential covariance function that additionally performs an automatic relevance determination [8]. In contrast to k_{SE}, it uses a characteristic length-scale parameter for each of the b dimensions of the input space, i.e., the hyperparameters $\boldsymbol{\theta}_{SEARD} = \langle \sigma_f, \ell_{1:b}, \sigma_n \rangle$. The separate length-scale parameters are an important tool to identify which of the b input dimensions are important for the prediction. As a result of that, if we find largely irrelevant signals in one of the input channels, they will not impact the prediction.

Given a set $\mathcal{D} = \{(y^{c_{1:b}}_i, y^{c'}_i)\}^t_{i=1}$ of training data where $y^{c_{1:b}}_i$ are the b-dimensional inputs resulting from the b active channels and $y^{c'}_i$ are the targets, the goal in regression is to predict target values $y^{c'}_{t'}$ at a new input point $y^{c_{1:b}}_{t'}$ based

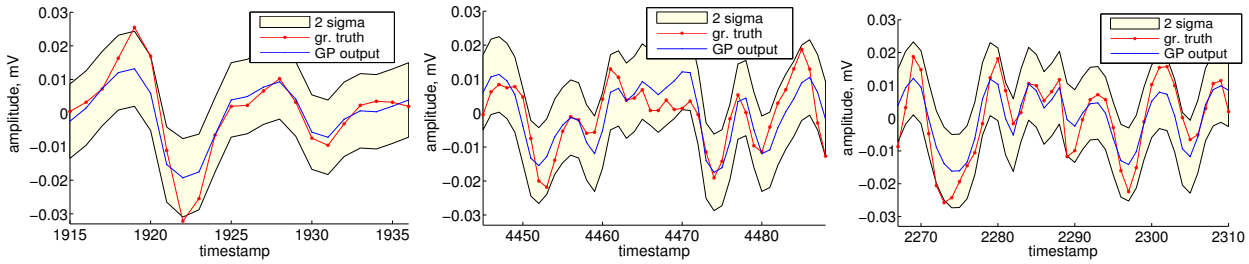

Fig. 2. Examples for the across-channel prediction of neural signals for different time series using Gaussian processes.

on the time index $t' > t$. Let $\mathbf{X} = [y_{t'}^{c_{1:b}}; \ldots; y_{t'+t-1}^{c_{1:b}}]^\top$ be the matrix of the inputs and $\mathbf{X}_*$ be defined analogously for multiple test data points. In the GP model, any finite set of samples is jointly Gaussian distributed

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{X}_*) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I} & k(\mathbf{X},\mathbf{X}_*) \\ k(\mathbf{X}_*,\mathbf{X}) & k(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix}\right),$$

where $k(\mathbf{X},\mathbf{X})$ refers to the covariance matrix built by evaluating the covariance function $k(\cdot,\cdot)$ for all pairs of all row vectors of $\mathbf{X}$. To make predictions at $\mathbf{X}_*$, we obtain the predictive mean

$$\begin{aligned} \bar{f}(\mathbf{X}_*) &= \mathbb{E}[f(\mathbf{X}_*)] \\ &= k(\mathbf{X}_*,\mathbf{X}) \left[k(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I}\right]^{-1} \mathbf{y} \end{aligned} \quad (3)$$

and the (noise-free) predictive variance

$$\begin{aligned} \mathbb{V}[f(\mathbf{X}_*)] &= [k(\mathbf{X}_*,\mathbf{X}_*) - k(\mathbf{X}_*,\mathbf{X})] \\ &\quad \left[k(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I}\right]^{-1} k(\mathbf{X},\mathbf{X}_*)\ , \end{aligned} \quad (4)$$

where $\mathbf{I}$ is the identity matrix. The corresponding (noisy) predictive variance for an observation can be obtained by adding the noise term σ_n^2 to the individual components of $\mathbb{V}[f(\mathbf{X}_*)]$.

One bottleneck of the standard GP model is its cubic complexity $\mathcal{O}(t^3)$ in the number of data points t. Different solutions have been proposed for lowering this upper bound such as dividing the input space into different regions and solving these problems individually or by deriving sparse approximations for the whole space. Sparse GPs [15], [16] use a reduced set of inputs to approximate the full GP model. This new set can be either a subset of the original inputs [15] or a set of r new pseudo-inputs [16] which are determined using an optimization procedure. This reduces the complexity from $\mathcal{O}(t^3)$ to $\mathcal{O}(tr^2)$ with $r \ll t$ and typically $r = \text{const}$. In practice, this results in a nearly linear complexity. In our work, we rely on sparse GPs with pseudo-inputs as proposed by [16] with $r = 20$ as it is the default setting.

IV. Selecting the Most Relevant Channels

The goal of our algorithm is to find the best set of b channels so that knowing these b channels will allow us to predict the remaining ones with a minimal overall prediction error. To decide which channels to record (*active channels*) and which channels to predict (*passive channels*), we use a graph structure that encodes the *pairwise prediction errors* on a training set. The training set can be obtained by scanning the probe for pairwise recordings, which requires at most $n^2 - n$ recordings over small time windows. Note that using only pairwise predictions is an approximation but allows for an efficient solution to split up the channels in active and passive ones. To finally obtain a prediction for the selected passive channels, we use information from all active channels and combine them into a joint GP model. Using joint GP models for the final prediction step improves the accuracy by $3 - 25\,\%$ in comparison to the pairwise prediction.

In our selection method, we build a complete directed weighted graph $G = (V, E)$, in which V are the vertices and E the edges. Each vertex corresponds to a channel on the neural probe. A directed edge from the node v to w models the predictability of the channel w given the signal of v. The weight of the edge (v, w) corresponds to the prediction error that is obtained when predicting w from v using the training data set. Our algorithm operates on the adjacency matrix A of G. Each element $a_{v,w}$ of the $n \times n$ matrix A represents the error obtained when predicting channel w based on channel v. Building up A is a straightforward task using the GP approach on a pairwise basis and has a complexity of $\mathcal{O}(n^2\, t)$, where t is the size of the training set, i.e., the number of data points taken for each channel. Note that this approach can be realized more efficiently if we take into account that signals of electrodes far away from each other typically do not have correlated signals so that the corresponding prediction error is very likely to be high.

Our algorithm aims at dividing the channels into active channels, specified by the set R, and passive ones, specified by the set I so that the prediction error is minimized. Given our graph representation, we can solve this problem by computing the optimal graph partitioning. As this is an NP-hard problem, we derive a greedy approach with a time complexity of $\mathcal{O}(n^2 \log n)$. As we will demonstrate, our approach, despite its heuristic nature, leads to near-optimal solutions in all our experiments.

The assignment of channels to R or I is performed based on the elements of the adjacency matrix. In the process of making the assignments, we assume that each channel in R can predict multiple channels in I. This is reasonable, as each channel could measure several neurons that might be observed at different other channels as well. Furthermore, we assume that each channel in I is predicted based on a single channel in R. This is an approximation but it allows us to come up with an efficient algorithm. Initially, all nodes are labeled as undefined U. In each step, the algorithm assigns at

least one node to one of the two sets R or I. Using a greedy heuristic, it does so by searching the smallest element in the matrix A:

$$(i^*, j^*) \quad = \quad \underset{(i,j)}{\operatorname{argmin}} a_{ij}, \tag{5}$$

where i and j refer to the row and column of the matrix A, i.e., $a_{i^*j^*}$ is the smallest element in A. The channel i is labeled as R and j as I. The matrix A is updated by deleting the j-th row and column. This deletion prevents the channel j from being predicted by another channel $\neq i$ and assures that j will not predict any other channel. We also delete the i-th column from matrix A, since channel i is already labeled as R. This prevents channel i from being selected as I later on. We repeat this procedure until the set of U channels is empty.

In addition to the greedy selection heuristic, we consider two alternative heuristics in our experiments.

1) Average heuristic: The main objective of this heuristic is to find the subset of active channels that have the best average prediction capabilities. In each iteration, we choose the channel i with the smallest average prediction error over all other channels to be active R. Furthermore, the channel j predicted best by this channel is labeled as passive I:

$$j^* \quad = \quad \underset{j}{\operatorname{argmin}} a_{i^*j}, \tag{6}$$

$$i^* \quad = \quad \underset{i}{\operatorname{argmin}} \frac{1}{n} \sum_{j=1}^{n} a_{ij} \tag{7}$$

2) Average heuristic with multiple passive channels (MPC): As in the average heuristic explained above, the channel with the smallest averaged prediction error over all channels is chosen to be active. The channel predicted best by this channel becomes passive. Additionaly, all channels for which the chosen active channel leads to the minimum error are labeled as passive. Average (MPC) runs the same steps as the average heuristic, but it performs a verification for additional passive channels.

As stated above, we have a restriction on the number of channels $b < n$ that we can record simultaneously. Thus, to minimize the prediction error, the set R should eventually contain exactly b elements. During the assignment procedure, we can encounter two different situations:

1) The R set contains already b elements, but not all the vertices have been assigned yet. In this case, we select the best possible assignment given the predicting channel has already been selected, i.e.,

$$(i^*, j^*) \quad = \quad \underset{(i,j)}{\operatorname{argmin}} a_{ij}, \text{with } i \in R, j \in I. \tag{8}$$

2) All the vertices are assigned, but $|R| < b$, i.e., we record less channels than possible. The solution is to select the channels from the I set that have the highest prediction error and assign them to R.

After termination, the algorithm provides the separation of all channels into R with $|R| = b$ and I with $|I| = n - b$. In addition to that, for all elements in I, we know which channel in R can predict the channel with the smallest prediction error, evaluated on a pairwise basis. Algorithm 1 summarizes the main steps of the assignment routine.

Algorithm 1 Greedy channel selection (V, A)

```
if (|V| ≤ b) then   // nothing needs to be done
   R = V;
   return;
endif
R = ∅
I = ∅
U = V
while (U ≠ ∅) do
   (i, j) = argmin_(i,j) a_ij,
   if (|R| = b and i ∉ R) then   // R reached max. size
      A(i, j) = ∞;
   else
      move element i from U to R;
      move element j from U to I;
      A(j, :) = ∞;    // set the j-th row to infinity
      A(:, j) = ∞;    // set the j-th column to infinity
      A(:, i) = ∞;    // set the i-th column to infinity
   endif
end while
while (b − |R| > 0)    // we can listen to more channels
   (i, j) = argmax_(i,j) a_ij   with   j ∈ I,
   move element j from I to R;
end while
```

V. Computational complexity

A straightforward approach to implement our algorithm is to make the inference directly on the adjacency matrix A. This leads to cubic complexity, i.e., $\mathcal{O}(n^3)$ where n is the overall number of channels. Two different processes contribute to this complexity. First, the algorithm must retrieve the smallest element in A, which leads to a complexity of $\mathcal{O}(n^2)$ and then delete the corresponding row and column from the matrix, which can be done in $\mathcal{O}(n)$. Thus, the complexity of one iteration is in $\mathcal{O}(n^2)$. Second, the maximum number of iterations is bounded by $n-1$ as in each iteration at least one element of U is removed. This leads to the overall complexity of $\mathcal{O}(n^3)$.

By exploiting more efficient data structures, however, we can reduce the complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2 \log n)$. We can turn the adjacency matrix A into a sorted list, in which each element stores the prediction error as well as the indices i and j of the corresponding entry in A. The list has then n^2 elements and sorting it takes $\mathcal{O}(n^2 \log(n^2)) = \mathcal{O}(n^2 \log n)$. After this initialization step, we have to perform up to $n-1$ iterations. In each iteration, we first need to remove the smallest element from the list, which is a constant time operation as the list is sorted. The remaining step is the delete operation, however, instead of deleting an element, we add the indices to delete from the list into a hash-table that allows for constant time adding and querying operations. Every time an element is removed from the list, it

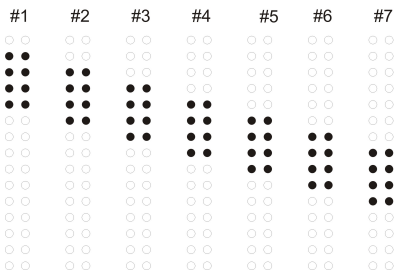

Fig. 3. Outline of a scanning experiment to collect training data of neighboring channels for our selection approach by switching electrode configurations.

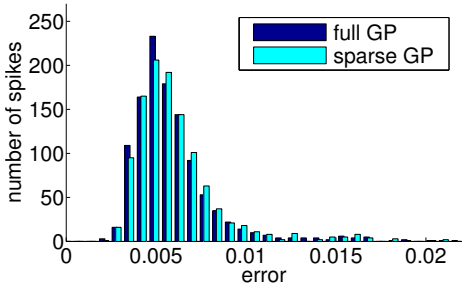

Fig. 4. Distribution of spike errors when using GPs vs. sparse GPs.

is discarded if its indices are stored in the hash-table. Over all iterations, this leads to a complexity of $\mathcal{O}(n^2)$. Thus, this procedure is dominated by the sort operation and has an overall complexity of $\mathcal{O}(n^2 \log n)$. For the average and average MPC heuristics, the same efficient data structure cannot be used, since the heuristics work directly with the matrix structure. Thus, the complexity of the channel selection algorithm in these cases is $\mathcal{O}(n^3)$.

VI. Experimental evaluation

After explaining our neural probe and the recorded neural data, we present our evaluation, which suggests that the combination of channel selection and GP-based prediction is well-suited to estimate neural signals across channels.

A. Neural probes and data acquisition

The neural probes used in this study to acquire in vivo data from anesthetized rats were developed in the framework of the EU project NeuroProbes [11]. The probes enable extracellular potentials to be acquired using densely packed recording sites arranged in a 2D array as illustrated in Fig. 1. These silicon-based probes are available as single-shaft probes and comb-like probe arrays [13]. They comprise a switch matrix realized in a 0.6 µm complementary-metal-oxide-semiconductor (CMOS)-technology used to flexibly select different sets of electrodes. This switch matrix enables the integration of a large number of recording sites despite tight space constraints driven by the request to minimize tissue reaction by reducing the overall probe dimensions [12].

The silicon-based probe shanks are 180 µm wide and measure 100 µm in thickness. The inter-electrode spacing of ca. 40 µm allows 188 electrodes to be integrated along the 4-mm-long probe shanks [12]. Using the CMOS-based switch matrix, eight channels are simultaneously selectable to monitor neighboring cells, which typically have a size of 20 to 50 µm. Experiments have shown that the measured signal decreases rapidly depending on the distance to the neuron. Up to a distance of 50 µm, single neurons can be perceived with a sufficiently large signal-to-noise ratio.

The data sets of neural activity have been recorded at the Institute for Psychology of the Hungarian Academy of Sciences, Budapest, Hungary, according to the respective animal care regulations. The data was recorded in vivo in the neocortex of Wistar rats as described in detail by Seidl et al. [13] and Dombovári et al. [4]. Extracellular potentials were recorded and sampled at 20 kHz over a period of several minutes in each probe configuration. The data is bandpass filtered in the range of 300 Hz to 3 kHz. We detected spikes by amplitude thresholding [13], and the final data contains small time windows around each spike.

We evaluated our selection algorithm on a data set recorded simultaneously at eight neighboring channels. The data was recorded over several minutes, which allows for a cross-validation of our algorithm. On this data set, we evaluated the task of selecting two to six out of eight channels. This setup allows us to compare the prediction of our selection to ground truth data. In principle, to obtain training data for the entire probe with 188 electrodes, it is required to record all pairwise combinations of recording channels in order to determine the adjacency matrix. This could be time-consuming, given that each combination needs to be observed for several seconds or even minutes to gather enough training data. Possibly, this could be organized in a more efficient manner by exploiting the fact that typically only spatially close electrodes measure correlated signals. Thus, instead of considering all possible combinations, only neighboring configurations could be observed as outlined in Fig. 3. This, however, requires further investigations and an analysis of correlations in the obtained data.

B. Prediction error

To evaluate the quality of our across-channel prediction, we compute the prediction error of the signals based on cross-validation. We consider the error over all passive, i.e., predicted, channels I defined as

$$\mathrm{error}(I) = \sqrt{\frac{1}{N_c}\sum_{i=1}^{N_c}\left(\frac{1}{N_s}\sum_{j=1}^{N_s}\mathrm{spikeError}(i,j)\right)}. \quad (9)$$

Here, N_s is the number of spikes observed in channel i during recording and N_c is the number of channels in I. The $\mathrm{spikeError}(i,j)$ is the root mean square error between the predicted (GP-output) and the recorded (ground truth) j-th spike in the i-th channel in the test data set.

C. Signal prediction using sparse GPs

The first experiment illustrates that GPs are a well-suited tool for making predictions across channels. Fig. 2 depicts three small sequences of predicted neural spikes, together

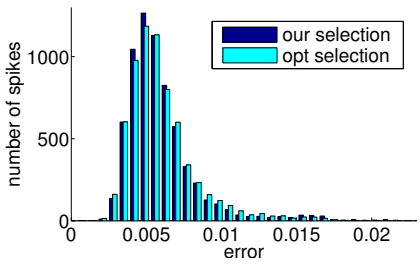

Fig. 5. Distribution of spike errors for our selection method vs. the optimal selection, which is computed by considering all possible assignments.

with the ground truth. We obtained similar results over all our data sets with average errors of approx. 6.5×10^{-3} mV. In these experiments, four active channels were used to predict signals of the remaining four passive channels.

We furthermore analyze the prediction performance of the sparse GP approach. In contrast to standard GPs, sparse GPs have a smaller computational complexity due to a reduced set of inputs. Therefore, we compare the prediction results of GPs versus sparse GPs. Fig. 4 depicts the distribution of the individual spike errors in a data set using GPs (blue) and sparse GPs (cyan). As can be seen, both approaches yield comparable results. The overall prediction errors for the GP prediction were around 6.45×10^{-3} mV and the difference between the full and sparse GP is around 8×10^{-6} mV. Thus, using sparse GPs instead of GPs results in a substantial reduction of computation time (cubic vs. linear complexity), while having only a small impact on the accuracy of the results.

D. Channel selection

The next experiments are designed to evaluate the performance of our method for selecting the active and passive set of the channels. In order to obtain training data for building the adjacency matrix, we use recordings for pairs of channels. As discussed above, the channels can be switched from active to passive mode in vivo to collect the required data.

In our data set, we use a set of $n = 8$ neighboring channels organized in two rows. We assume that only $b = 4$ are active at the same time. Thus, our approach has to select four out of eight channels for predicting the remaining four. As our probe has more channels and can actually record up to eight channels simultaneously, we were able to record all channels to obtain a ground truth data set. By evaluating all possible combinations of active and passive channels in an offline setup, we can always identify the optimal solution and compare it to our result.

Tab. I summarizes the results of our algorithm on six different datasets. For each of these datasets, we computed the assignment with our algorithm and compare it to the optimal one. In the table, the last column lists the rank of the solution in the sorted list of all solutions. The rank is equal to 1 for the optimal solution and equal to 2 for the second best solution. With $n = 8$ and $b = 4$, 70 is the worst one as there are 70 sets of 4 out of 8 elements. As Tab. I indicates, our approach in most of the cases finds a solution that is

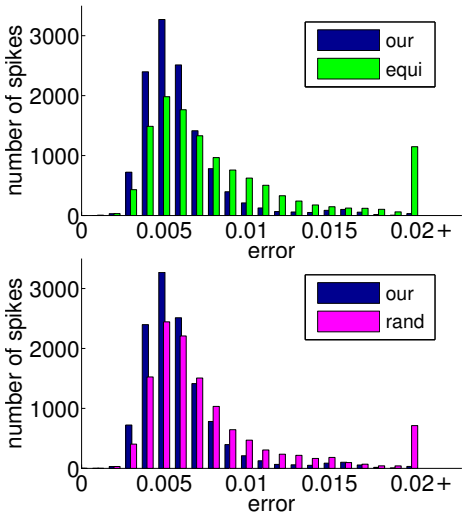

Fig. 6. Distribution of spike errors for our selection method vs. an equidistant (top) and a random (bottom) electrode selection.

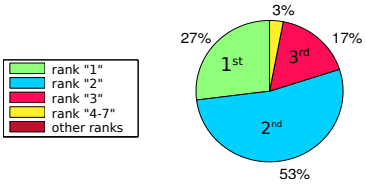

Fig. 7. Evaluation of how often our channel selection algorithm found a solution with a certain rank compared to the optimal solution.

among the best three ones. In such cases, the corresponding prediction errors are also close to those of the optimal set of active and passive channels. As Fig. 5 shows, there are only minor differences in the distribution of the spike prediction errors for our approach compared to the optimal selection.

We furthermore compared our selection approach to an equidistant and a random selection of electrodes. While the error for our approach is 6.40×10^{-3} mV, the error for the equidistant selection is 10.8×10^{-3} mV with a maximum amplitude of 0.12 mV. As can be seen in Fig. 6, the equidistant and random selections more often lead to larger individual spike errors.

To provide a more quantitative evaluation, we performed the same evaluation on 40 datasets. In our experiments, the solution of our channel selection algorithm is within the best three solutions in 97 % of all cases as depicted in Fig. 7. In 3% of the cases, the solution was worse but never larger than rank 13 out of 70. This indicates that—at least on our type of data—this is an effective selection strategy which brings down the computational complexity to $\mathcal{O}(n^2 \log n)$. Averaged over all 40 data sets, the difference in error between our approach and the optimal selection is 6.47×10^{-5} mV. We also evaluated the influence of the different heuristics on the selection. Results are summarized in Fig. 8. Here, the average and average MPC heuristics more often find the optimal solution. Thus, it can be beneficial to combine these heuristics and select the subset of channels that leads to the smallest overall prediction error.

TABLE I

CHANNEL PREDICTION RESULTS FOR DIFFERENT TRAINING DATA SETS.

Spikes		Error $\times 10^{-3}$			
train	test	our	opt	diff	rank
20	282	6.490	6.490	0	1
20	280	6.400	6.400	0	1
25	293	6.370	6.370	0	1
30	285	6.344	6.344	0	1
40	277	6.398	6.372	0.0261	2
20	280	6.621	6.559	0.0621	3

TABLE II

CHANNEL PREDICTION RESULTS FOR A VARYING NUMBER OF ACTIVE CHANNELS IN A 10-FOLD CROSS-VALIDATION.

# active	Opt. sel	Our approach		Equidist.	Rand.
channels	err. 10^{-3}	err. 10^{-3}	rank	err. 10^{-3}	err. 10^{-3}
2	7.91	10	7.72	10.5	10.26
3	6.941	6.945	1.1	9.19	9.68
4	6.427	6.49	1.81	9.23	8.932
5	5.666	6.047	5.27	6.48	8.95
6	5.432	5.475	2.9	6.48	8.844

In an additional experiment, we evaluated the influence of the number of active channels on the prediction result. This is designed to more closely reflect real-world situations in which a small subset must be chosen to represent a large set of electrodes. Tab. II summarizes the results for selecting two to six out of eight channels, averaged over 10 runs. As can be seen, our method is in most cases close to the optimal solution, and always outperforms the equidistant and the random selection strategies in terms of the prediction errors.

In summary, the experiments indicate that our selection method yields highly accurate results. Furthermore, they demonstrate that the higher resolution obtained by the current probe technology indeed provides better information. Otherwise, the equidistant strategy, which corresponds to probes with a coarser resolution, would perform equally well as the high-resolution probes paired with our algorithm.

VII. CONCLUSION

The ability to analyze neural signals is an important basis for brain computer interfaces and brain-controlled robots. This paper considers the problem of effectively obtaining data from neural probes that have limitations in the number of channels they can record simultaneously. It makes two contributions. First, it presents an approach using sparse Gaussian processes for predicting the signal of a channel based on the information of other channels. Second, it introduces a novel selection technique to identify, which of a given maximum number of channels to record so as to minimize the overall prediction error for the remaining channels. We implemented and tested our approach on data obtained with a real neural probe. As our experiments illustrate, our method can effectively select the channels that produce low prediction errors. We furthermore found that it often finds a solution that is close to the optimal set of channels. Directions for future work include the application of the approach to large scale data sets with a limited number of simultaneously observable channels and efficient exploration strategies for the acquisition of the required training data.

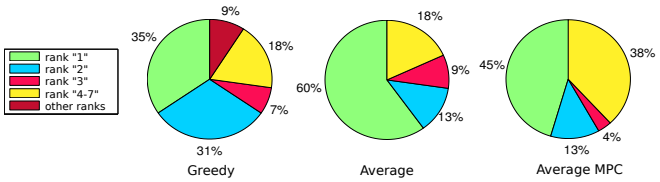

Fig. 8. Evaluation of how often our channel selection algorithm found a solution with a certain rank compared to the optimal solution (bottom) for greedy, average and average MPC heuristics.

REFERENCES

[1] T. J. Blanche, M. A. Spacek, J. F. Hetke, and N. V. Swindale. Polytrodes: High-density silicon electrode arrays for large-scale multiunit recording. *Journal of Neurophysiology*, 93(2987–3000), 2005.

[2] G. Van Dijck, M. M. Van Hulle, S. A. Heiney, P. M. Blazquez, H. Meng, D. E. Angelaki, A. Arenz, T. W. Margrie, A. Mostofi, S. Edgley, F. Bengtsson, C.-F. Ekerot, H. Jörntell, J. W. Dalley, and T. Holtzmann. Probabilistic identification of cerebellar cortical neurones across species. *PLoS ONE*, 8(3), 2013.

[3] G. Van Dijck, K. Seidl, O. Paul, P. Ruther, M. M. Van Hulle, and R. Maex. Enhancing the yield of high-density electrode arrays through automated electrode selection. *International Journal of Neural Systems*, 22(1), 2012.

[4] B. Dombovári, R. Fiáth, B. P. Kerekes, E. Tóth, L. Wittner, D. Horváth, K. Seidl, S. Herwik, T. Torfs, O. Paul, P. Ruther, H. Neves, and I. Ulbert. In vivo validation of the electronic depth control probes. *Biomedizinische Technik. Biomedical engineering*, 1–7, 2013.

[5] J. Du, T. J. Blanche, R. R. Harrison, H. A. Lester, and S. C. Masmanidis. Multiplexed, high density electrophysiology with nanofabricated neural probes. *PLoS ONE, Public Library of Science*, 6(10), 2011.

[6] G. W. Fraser and A. B. Schwartz. Recording from the same neurons chronically in motor cortex. *Journal of Neurophysiology*, 107, 2012.

[7] K. D. Harris, D. A. Henze, J. Csicsvari, H. Hirase, and G. Buzsaki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of Neurophysiology*, 84:401–414, 2000.

[8] R. M. Neal. Bayesian learning for neural networks. *Lecture Notes in Statistics*, 118, 1996.

[9] R. Quian Quiroga, Z. Nadasdy, and Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16:1661–1687, 2004.

[10] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[11] P. Ruther, S. Herwik, S. Kisban, K. Seidl, and O. Paul. Recent progress in neural probes using silicon MEMS technology. *IEEJ Transactions on Electrical and Electronic Engineering*, 5:505–515, 2010.

[12] K. Seidl, H. Herwig, T. Torfs, H. P. Neves, O. Paul, and P. Ruther. CMOS-based high-density silicon microprobe array for electronic depth control in intracortical neural recording. *Journal of Microelectromechanical Systems*, 20(6):1439–1448, 2011.

[13] K. Seidl, M. Schwärzle, I. Ulbert, H. P. Neves, O. Paul, and P. Ruther. CMOS-based high-density silicon microprobe arrays for electronic depth control in intracortical neural recording – characterization and application. *Journal of Microelectromechanical Systems*, 21(6):1426–1435, 2012.

[14] K. Seidl, T. Torfs, P. A. de Maizire, G. Van Dijck, R. Csercsa, B. Dombovari, Y. Nurcahyo, H. Ramirez, M. M. van Hulle, G. A. Orban, O. Paul, I. Ulbert, H. Neves, and P. Ruther. Control and data acuqisition software for high-density CMOS-based microprobe arrays implementing electronic depth control. *Biomedizinische Technik. Biomedical Engineering*, 55:183–191, 2010.

[15] A. J. Smola and P. L. Bartlett. Sparse greedy Gaussian process regression. In *Proc. of the Conf. on Neural Information Processing Systems (NIPS)*, pages 619–625, 2000.

[16] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. In *Advances in Neural Information Processing Systems 18*, pages 1259–1266, 2006.