

# Multiple Hypothesis Tracking Using Clustered Measurements

Michael T. Wolf  
Jet Propulsion Laboratory\*  
California Institute of Technology  
wolf@jpl.nasa.gov

Joel W. Burdick  
Mechanical Engineering  
California Institute of Technology  
jwb@robotics.caltech.edu

**Abstract**—This paper introduces an algorithm for tracking targets whose locations are inferred from *clusters* of observations. This method, which we call MHTC, expands the traditional multiple hypothesis tracking (MHT) hypothesis tree to include *model hypotheses*—possible ways the data can be clustered in each time step—as well as ways the measurements can be associated with existing targets across time steps. We present this new hypothesis framework and its probability expressions and demonstrate MHTC’s operation in a robotic solution to tracking neural signal sources.

## I. INTRODUCTION

Robots that must observe several dynamic objects in their environment often employ methods for multitarget tracking (MTT). These techniques maintain the identities of observed objects of interest by associating new sensor measurements with estimated or known targets and using these associations to update target state estimates. Existing MTT methods typically assume that each target can generate at most one observation at each time step. In this paper, we address the problem of tracking targets that are each represented by a multitude of statistically distributed observations in each time step, a matter that has not previously been addressed in an integrated, comprehensive manner. We propose a new procedure, termed *MHTC* for *multiple hypothesis tracking of clusters*, that combines Bayesian techniques from clustering and data association into a novel hypothesis framework to effectively and robustly track such targets.

Specifically, we consider the following scenario. At each time step  $k$ , a set of observations  $Y^k = \{y_i^k\}_{i=1}^N$  is obtained from an unknown number of current targets. Each target has generated many observations, which may be modeled as samples from a probability distribution representing that target. These observations must be broken up into sets (i.e., clustered), assigning each observation to a cluster  $C_g^k$  ( $g = 1, \dots, G^k$ ) and estimating the number of sources  $\hat{G}^k$ . Additionally, the current clusters must be matched (i.e., tracked) to the targets estimated to exist at time  $k - 1$  (indexed by  $j$ ) and used to update the estimate of the current target state  $x_j^k$ . Several factors may complicate this process, such as newly appearing or disappearing targets, temporary occlusions or missed detections, and false measurements (clutter).

\*This work was completed at the California Institute of Technology with support from the National Institutes of Health and the Rose Hills Foundation.

This tracking problem arises, for example, in the authors’ related work on a miniature robot for autonomously positioning electrodes in the brain to obtain high quality extracellular recordings [1]–[3]. In the main loop of this robot’s control algorithm, the electrode’s signal is periodically sampled for a brief interval and analyzed to determine if positional adjustments will improve signal quality of a given neuron. However, an electrode may record the spiking activity of several nearby neurons, each one generating many spikes over the short time interval. Thus, the detected spikes must first be sorted according to their generating neurons, and previously identified neurons must be re-identified in the current recording interval (“tracked”), despite possible changes in the amplitude, phase, and numbers of neuronal signals.

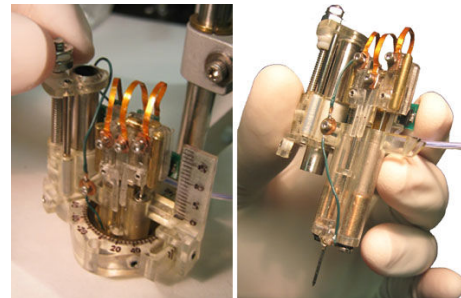


Fig. 1. Photographs of the Robotic Electrode Microdrive.

The problem statement described above may also arise in other robotics application areas. For example, in radar detection of pedestrians for robotically assisted driving, a single person might be represented by a *set* of distance measurements (due to variability in the range to different body parts and to reflections); these observations from each person must be grouped together, and the resulting pedestrians tracked over time [4]. In computer vision, objects of interest in each frame may be inferred from a grouping of features; each image must be segmented and each target object tracked over multiple frames [5].

In traditional MTT, the locations of several objects of interest (targets) are measured in sequential “scans” of an observation volume. Using these data, MTT solutions combine a filter for estimating the target states and a data association technique for assigning the current measurements to known

targets (see [6]–[9] for MTT overviews). Current MTT solutions differ primarily by their data association method, which governs which measurements are assigned to putative targets and so used to update the state estimates of those targets (typically through a Kalman Filter). An *exclusivity* principle is usually enforced, under which each target may generate at most one measurement and each measurement can represent only a single target. Under this assumption, a set of legal *data association hypotheses* may be defined, where each hypothesis assigns every measurement to an existing target (or possibly designates it as a new target or false measurement).

Multiple hypothesis tracking (MHT), attributed to Reid [10], is generally accepted as the preferred data association mechanism for modern MTT systems [11]—though other approaches, such as nearest neighbor and joint probabilistic data association (JPDA) [12], remain popular, largely because of MHT’s heavy computational demands. MHT maintains many possible data association hypotheses and propagates the corresponding target state estimates for each hypothesis, implicitly deferring decisions in anticipation that subsequent data measurements will resolve any ambiguity. A key recent development in MHT implementations is the use of an algorithm, originally due to Murty [13], to generate only the  $L$ -best hypotheses at each time step, avoiding the combinatorial explosion of hypotheses at a reasonable cost of sub-optimality [14].

The key differentiator of our problem (MTT for cluster-producing targets) versus traditional applications is the multitude of observations per target in each scan. The measurements of target location are therefore not received directly from the sensor but rather must be estimated from subsets of observations. Further, the uncertainty and difficulty inherent in the clustering problem greatly complicates the tracking task: Correctly assigning observations to their generating targets is essential for accurately estimating the location, and even the *number* of “measurements” (in this case, clusters) is unknown *a priori*. Some MTT solutions such as those using Finite Set Statistics [15] may allow for a many-to-one relationship of observations to targets, but to our knowledge no others fully combine advanced clustering and model selection techniques into the tracking solution. Given the significant ambiguities in neural data, such an integrated approach is critical for our application.

Our MHTC algorithm presented in this paper is a multiple hypothesis approach to combined clustering and tracking, propagating not only multiple data association hypotheses but also multiple hypotheses on how the data should be clustered. We detail this novel hypothesis framework as well as the explicit equations to calculate the probabilities of the hypotheses. Our previously presented Bayesian clustering algorithm [16], along with its inherent measures of cluster associations across time intervals, provides a key component of the MHTC algorithm. While this prior work included a simple “single hypothesis” nearest neighbor tracker, MHTC offers a significantly more robust solution, demanded in situations when, for instance, a target is temporarily occluded, targets

are entering or leaving the observation volume, a time step’s clustering result contains an error, or the signals of different targets are difficult to distinguish for a time. Additionally, MHTC utilizes a recursive filter for estimating the state of each target, explicitly identifies false clusters, and includes a more sophisticated model selection technique.

The remainder of the paper is structured as follows. Section II reviews our Bayesian clustering method [16], for it is integrated into the MHTC solution. Section III overviews the MHTC solution, while Section IV provides further mathematical detail. MHTC is demonstrated by experimental results from our neuron-tracking robot in Section V, and Section VI provides concluding remarks.

## II. REVIEW OF MAP CLUSTERING METHOD

Our clustering method [16] is founded on the optimization of a Gaussian mixture model via expectation–maximization (EM) [17]. The underlying assumption is that the observations  $y_i^k$  in different clusters  $\mathcal{C}_g^k$  can be modeled as samples from different multivariate statistical distributions, where each distribution represents a specific target. Let  $\mathcal{M}_m$  denote the  $m$ th mixture model class under consideration (the need for multiple model classes and the selection technique are discussed in Section III)—the model class dictates the model order  $G_m$  (i.e., the number of components/clusters), the form of the  $g$ th probability density  $f_g$  (typically Gaussian), and the form of the model parameters  $\Theta_m^k = \{\pi_g^k, \theta_g^k\}_{g=1}^{G_m}$ .  $\pi_g^k$  and  $\theta_g^k$  denote the mixture weight and parameters of the  $g$ th component, respectively (for a Gaussian distributions, the component parameters are the mean and covariance matrix:  $\theta_g^k = \{\mu_g^k, \Sigma_g^k\}$ ).

The goal of the clustering method is to optimize the mixture model parameters  $\Theta_m^k$  for each model class and assign each observation to its most probable mixture component. While most traditional clustering procedures use a maximum likelihood (ML) approach, we find the maximum *a posteriori* (MAP) parameters, incorporating Bayes’ Rule:

$$p(\Theta_m^k | Y^{1:k}, \mathcal{M}_m) \propto p(Y^k | \Theta_m^k, \mathcal{M}_m) p(\Theta_m^k | Y^{1:k-1}, \mathcal{M}_m), \quad (1)$$

where  $Y^{1:k} = \{Y^1, \dots, Y^k\}$  denotes all observations from the 1st through the  $k$ th time steps. The *mixture likelihood* of the model parameters given the data is

$$p(Y^k | \Theta_m^k, \mathcal{M}_m) = \prod_{n=1}^N \sum_{g=1}^{G_m} \pi_g^k f_g(y_n^k | \theta_g^k). \quad (2)$$

The novelty of our clustering approach lies in the way we incorporate a prior that aids in tracking targets. We have constructed an appropriate prior on the model parameters  $\Theta_m^k$ —focusing on the cluster “locations”  $\mu_g^k$ —based on the predicted target locations,  $\{\hat{\mu}_j^{k|k-1}\}_{j=1}^J$ , of all  $J$  targets hypothesized to exist from time  $k-1$ . This mixture prior on a cluster mean is

$$p(\mu_g^k | Y^{1:k-1}, \mathcal{M}_m) = \frac{\omega_g^k}{V} + \sum_{j=1}^J \omega_j^k f_{\mathcal{N}}(\mu_g^k | \hat{\mu}_j^{k|k-1}, S_j^{k-1}), \quad (3)$$

where  $\omega_j^k$  denotes the mixture weight;  $f_{\mathcal{N}}$  the Gaussian PDF;  $V$  the observation volume; and  $S_j^k$  the innovation covariance. The zeroth component is uniform over the observation volume  $V$  to capture new targets and the remaining components are Gaussian distributions to match existing targets near their predicted locations. Using uniform priors for the mixture weights and covariance matrix elements and letting  $\psi_j^{k|k-1}$  denote the parameters needed for the  $j$ th mixture prior component,

$$p(\Theta_m^k | Y^{1:k-1}, \mathcal{M}_m) \propto \prod_{g=1}^{G_m} \sum_{j=0}^J \omega_j^k f_j(\mu_g^k | \psi_j^{k|k-1}). \quad (4)$$

Given the form of (2) and (4), an analytical solution for the parameters that maximize (1) cannot be found. EM is thus used to estimate these parameter values. To apply this technique, we contrive *cluster association indicators*  $\mathcal{Z} = \{\zeta_{gj}\}$ , hidden data that specify whether the cluster  $\mathcal{C}_g^k$  is related to the  $j$ th target, or, ideally,

$$\zeta_{gj} = \begin{cases} 1 & \text{if } \mu_g^k \text{ is the new position of the } j\text{th target} \\ 0 & \text{otherwise.} \end{cases}$$

The EM algorithm iterates between an E-step to calculate the conditional expectation of the hidden data using the current parameter estimates,

$$\hat{\zeta}_{gj} = E[\zeta_{gj} | Y^{1:k}, \Theta_m] = \frac{\omega_j^k f_j(\hat{\mu}_g^k | \psi_j^{k|k-1})}{\sum_{l=0}^{G^{k-1}} \omega_l^k f_l(\hat{\mu}_g^k | \psi_l^{k|k-1})}, \quad (5)$$

and an M-step to find the parameter estimates  $\hat{\Theta}$  that maximize a modified form of (1) given  $\hat{\zeta}_{gj}$ , until some convergence threshold is reached. (The algorithm requires an initial guess or “seed clusters” for initialization.) See [16] or [18] for further details.

### III. MHTC FRAMEWORK

#### A. Definitions

1) *Hypothesis Terminology*: We define two types of hypotheses in MHTC: model hypotheses and data association hypotheses. A *model hypothesis* represents a possible clustering of the observations and is denoted by its corresponding mixture model class,  $\mathcal{M}_m$ . Each *data association hypotheses*,  $h_l = \{\tau_l, \nu_l, \phi_l\}$ , assigns each cluster in a given model hypothesis to a target (or marks it as spurious): The set  $\tau_l$  contains the assignments of the model’s clusters to known targets;  $\nu_l$  contains the indices of the model’s clusters that are identified as new neurons; and  $\phi_l$  holds the the indices of *false clusters* (spurious groupings of outliers or similar clustering errors) in the current model. Note that  $N_\tau$ ,  $N_\nu$ , and  $N_\phi$  are the respective cardinalities of these sets and  $G_m = N_\tau + N_\nu + N_\phi$ .

We call the combination of a data association hypothesis and its parent model hypothesis a particular *joint hypothesis* at time  $k$ ,  $H_l^k = \{\mathcal{M}_{m(l)}, h_l\}$ . The joint hypothesis  $H_l^k$  thus postulates a complete set of data associations for time  $k$ , including the observation–cluster associations in  $\mathcal{M}_{m(l)}$  and the cluster–target associations in  $h_l$ . A particular joint hypothesis is combined with its *parent hypothesis*  $H_{\rho(l)}^{1:k-1}$  to define a *global*

*hypothesis*,  $H_l^{1:k} = \{H_l^k, H_{\rho(l)}^{1:k-1}\}$ , which includes the full history of all model and data association hypotheses from time 1 through  $k$ .<sup>1</sup> Finally, it is convenient to define  $\Omega^k$  as the set of all surviving global hypotheses  $\{H_l^{1:k}\}_{l=1}^L$  and all data  $Y^{1:k}$ , which thus provides all relevant measured and hypothesized information about time  $k$ :  $\Omega^k = \{\{H_l^{1:k}\}_{l=1}^L, Y^{1:k}\}$ .

2) *Dynamical System Model*: After measurements are assigned under the hypothesis  $h_l$ , they are used to update the target’s *track*, its sequence of estimated states. The implementation presented in Section V utilizes a simple linear, discrete-time, Gauss–Markov system:

$$\begin{aligned} x_j^k &= F^{k-1} x_j^{k-1} + v_j^{k-1} & v_j^k &\sim \mathcal{N}(0, Q_j^k) \\ \hat{\mu}_j^k &= H^k x_j^k + w_j^k & w_j^k &\sim \mathcal{N}(0, R_j^k) \end{aligned}$$

and the Kalman Filter for state estimation. Note that other system models and filters are possible.

3) *Probability Models*: Given a set of targets in the parent hypothesis from  $H_{\rho(l)}^{1:k-1}$ , the probabilities of the existence and location of new measurements in interval  $k$  are modeled as follows. The occurrence that the  $j$ th existing target is *detected* (i.e., produces a cluster) is considered a Bernoulli trial with probability  $P_{d,j}$ . If the target is detected, the associated measurement is expected to appear near the target’s predicted location with a Gaussian distribution,  $f_{\mathcal{N}}(\mu_g^k | \hat{\mu}_j^{k|k-1}, S_j^k)$ , where the predicted mean and its covariance are provided by the Kalman filter. The numbers of new targets or false clusters appearing in a given time interval are each modeled by the Poisson distribution with respective rates  $\lambda_\nu$  and  $\lambda_\phi$ . If a measurement originates from a new target or false cluster, it may arise anywhere in the observation volume  $V$  with a uniform PDF. The parameters  $P_{d,j}$ ,  $\lambda_\nu$ , and  $\lambda_\phi$  are set by the user and may vary across sampling intervals.

#### B. Hypothesis Tree Structure

As shown in Figure 2, the MHTC algorithm extends the traditional MHT tree to include model hypotheses as well as data association hypotheses. If  $L$  global hypotheses exist at time  $(k-1)$  and we consider  $\bar{M}$  model classes for each parent hypothesis, then  $(L\bar{M})$  model hypotheses are formed at time  $k$ , each of which is optimized according to the MAP EM procedure of Section II. By use of Murty’s algorithm, only the  $L$  best data association hypotheses are generated from each parent model hypothesis. To end the hypothesis management at time  $k$ , the best  $L$  global hypotheses are selected from the  $(L^2\bar{M})$  that have been generated. Section III-C provides further detail on the above process.

#### C. Overview of the MHTC Process

This section walks through the MHTC process of the combined clustering and multiple hypothesis tracking, as illustrated in Figure 3. Steps 2–6 are similar to the procedure detailed in [16], but some of these steps require a reformulation in the context of the MHT framework.

<sup>1</sup>Note that the subscripts  $m(l)$  and  $\rho(l)$  are used to indicate the index of the model or global hypothesis, respectively, that is the parent of the  $l$ th data association hypothesis; similarly, in a slightly abusive notation,  $\rho(m)$  may also indicate the parent global hypothesis of the  $m$ th model hypothesis.

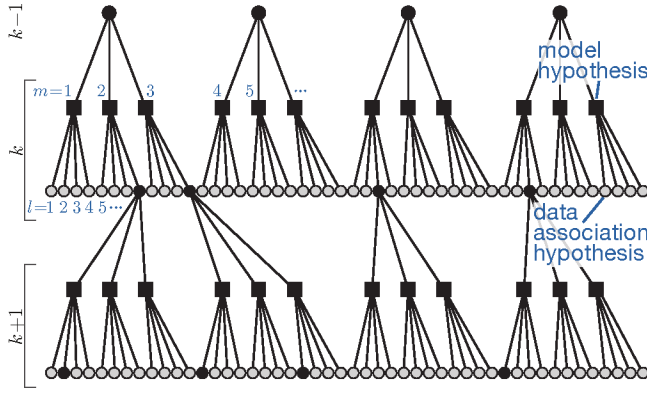


Fig. 2. MHTC hypothesis tree structure, illustrating the integration of model hypotheses into the traditional MHT framework, using  $L = 4$  and  $\bar{M} = 3$ . Squares represent model hypotheses (i.e., clustering output) and black circles represent surviving data association hypotheses at each time step.

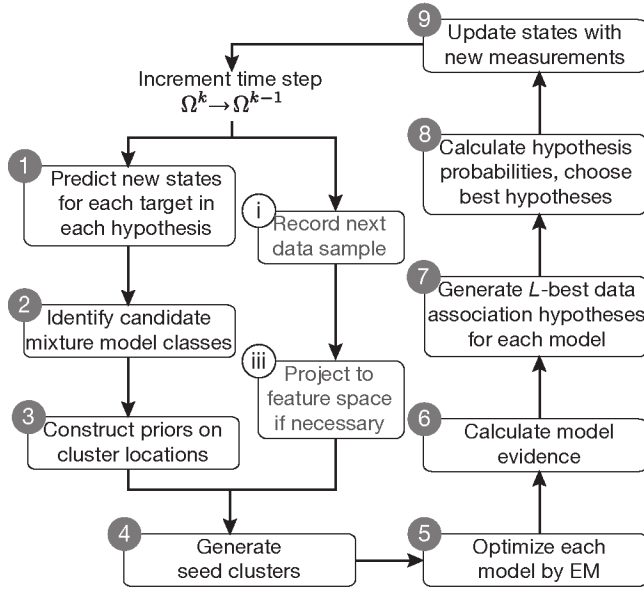


Fig. 3. MHTC process diagram. See text for description of each step. Steps 1-9 indicate core clustering and hypothesis tracking procedures, whereas steps i-iii are for data acquisition only.

**Step 1.** Given  $\Omega^{k-1}$  and the measurement update from time  $k-1$  (see Step 9), the first step is to predict the measured locations (means) of the targets. We use the Kalman Filter equations to calculate  $\hat{x}_j^{k|k-1}$  (state prediction),  $\Lambda_j^{k|k-1}$  (state prediction covariance), and  $\hat{\mu}_j^{k|k-1}$  (measurement prediction).

**Step 2.** For every parent hypothesis in  $\Omega^{k-1}$ , a set of mixture model classes  $\{\mathcal{M}_m\}$  is generated, which will be used to cluster the current data  $Y^k$ . A range of model classes are required primarily because the number of clusters  $G_m$  is unknown, and so various model orders are attempted and each resulting model hypothesis analyzed. However, to save the computation inherent in optimizing and evaluating “dead end” model classes, the probability of a model class arising from its parent hypothesis is calculated and then tested against

a threshold  $\beta$ :

$$P(\mathcal{M}_m | H_{\rho(m)}^{1:k-1}, Y^{1:k-1}) > \beta. \quad (6)$$

An expression for this probability, which depends on the probabilities of target detection, new targets, and false clusters, is provided in Table I. Model classes that do not pass this thresholding test are discarded; surviving model classes become the model hypotheses.

**Step 3-5.** These steps follow the clustering procedure of [16]. In each model hypothesis, the prior on the model parameters is constructed as in Eq. (3). (The innovation covariance  $S_j^k$  may be calculated using the Kalman Filter equations.) Seed clusters (the initial conditions for EM) are based on the targets’ predicted locations, and EM is implemented according to the MAP procedure as previously discussed.

**Steps i-iii.** These steps collectively acquire and prepare incoming data for the clustering and tracking procedures. In many applications (e.g. machine vision and neural data), it is necessary or helpful to extract features from the recorded data and use the feature vector in tracking operations.

**Step 6.** The evidence of each model hypothesis,  $p(Y^k | \mathcal{M}_m, H_{\rho(m)}^{1:k-1}, Y^{1:k-1})$ , may now be calculated, preferably via Laplace’s method [19]. If desired for computational saving, one could now prune extremely unlikely models before continuing, by setting a threshold on model evidence. However, this is only worthwhile for large numbers of targets or large numbers of hypotheses ( $L$ ).

**Step 7.** The core step in MHT generates the data association hypotheses,  $h_l = \{\tau_l, \nu_l, \phi_l\}$ . As noted earlier, it is desirable to utilize Murty’s  $L$ -best ranked linear assignment algorithm to produce only the best data association hypotheses from each parent cluster hypothesis, obviating the need for full enumeration of all possible data associations. This technique requires careful formulation of the probability calculations so that Murty’s algorithm can operate on a matrix of assignment likelihoods  $\mathbf{A}$ . Section IV derives an appropriate expression of  $\mathbf{A}$  and describes the hypothesis generation in more detail.

**Step 8.** Suppose that a total of  $\bar{M}$  model hypotheses exist at this time, each of which has now spawned  $L$  data association hypotheses. From the  $(\bar{M}L)$  hypotheses that have been generated, the most probable  $L$  global hypotheses must be selected. Evaluating each model and data association hypothesis together with its parent hypothesis  $H_{\rho(l)}^{1:k-1}$ , the probability of each new global hypothesis  $P(H_l^{1:k} | Y^{1:k})$  can be calculated, as detailed in Section IV. This step provides the set of best global hypotheses in  $\Omega^k$ .

**Step 9.** Finally, for each  $H_l^{1:k} \in \Omega^k$ , the hypothesized data associations  $h_l$ , along with the optimized parameters  $\Theta_m^k$  of the corresponding model hypothesis, are used to update the Kalman Filter, calculating  $K_j^k$  (Kalman filter gain),  $\hat{x}_j^{k|k}$  (updated state estimate), and  $\Lambda_j^{k|k}$  (updated state covariance).

#### IV. MHTC PROBABILITIES

This section provides expressions for the key probabilities necessary for the MHTC algorithm and formulates the data

TABLE I  
FACTORS IN GLOBAL HYPOTHESIS PROBABILITY

Definition	Expression for Model in Sect. III-A	Interpretation
$\mathcal{P}_{1,l}$ $p(\{\hat{\mu}_g^k\}_{g=1}^{G_{m(l)}}   H_l^{1:k}, Y^{1:k-1}, \hat{\mu}_g^k \in \hat{\Theta}_{m(l)}^k)$	$\left[ \prod_{(g,j) \in \tau_l} f_{\mathcal{N}}(\hat{\mu}_g^k   \hat{\mu}_j^{k k-1}, S_j^k) \right] \left[ \prod_{g \in \nu_l} \frac{1}{\bar{V}} \right] \left[ \prod_{g \in \phi_l} \frac{1}{\bar{V}} \right]$	means' likelihood under $h_l$
$\mathcal{P}_{2,l}$ $p(H_l^k   H_{\rho(l)}^{1:k-1}, Y^{1:k-1})$	$A_m \left[ \prod_{j=1}^J f_{\mathcal{B}}(\delta_{j,l}   \mathbf{P}_{d,j}) \right] (\lambda_{\nu})^{N_{\nu}} (\lambda_{\phi})^{N_{\phi}}$	joint hypothesis prior <sup>a</sup>
$\mathcal{P}_3$ $P(\mathcal{M}_{m(l)}   H_{\rho(l)}^{1:k-1}, Y^{1:k-1})$	$\sum_{N_{\tau}=0}^J \left[ \frac{(\lambda_0)^{G_m - N_{\tau}} e^{-\lambda_0}}{(G_m - N_{\tau})!} \sum_{\delta \in \Upsilon} \prod_{j=1}^J f_{\mathcal{B}}(\delta_{j,l}   \mathbf{P}_{d,j}) \right]$	model hypothesis prior <sup>b</sup>
$\mathcal{P}_4$ $p(Y^k   \mathcal{M}_{m(l)}, H_{\rho(l)}^{1:k-1}, Y^{1:k-1})$	see [17] for Laplace's method (or other approximation)	model evidence
$\mathcal{P}_5$ $P(H_{\rho(l)}^{1:k-1}   Y^{1:k-1})$	same as Eq. (7), from previous time step	parent hypothesis probability

<sup>a</sup>  $A_m$  is a constant depending on the model class and does not require calculation.  $f_{\mathcal{B}}$  is the Bernoulli distribution, and  $\delta_{j,l}$  is an indicator variable of whether the  $j$ th target is tracked under the  $l$ th hypothesis ( $j \in \tau_l$ ).

<sup>b</sup>  $\lambda_0 = \lambda_{\nu} + \lambda_{\phi}$ .  $\delta$  is a vector of indicator variables  $\delta_j$ ,  $j = 1, \dots, J$ , and  $\Upsilon$  is the set of all possible  $\delta$  for a given  $J$ .

association problem for use of Murty's  $L$ -best assignment algorithm.

The key probability to be determined for MHTC is that of a global hypothesis given all collected data,  $P(H_l^{1:k} | Y^{1:k})$ , the basis of the final hypothesis selection for time  $k$  (in Step 9). The expression for this probability includes all relevant measures about the parent hypothesis, model hypothesis, and data association hypothesis. This *global hypothesis probability* may be expressed as

$$P(H_l^{1:k} | Y^{1:k}) \approx \frac{1}{\mathcal{C}} \frac{\mathcal{P}_{1,l} \mathcal{P}_{2,l}}{\sum_{n \in \Gamma} \mathcal{P}_{1,n} \mathcal{P}_{2,n}} \mathcal{P}_3 \mathcal{P}_4 \mathcal{P}_5, \quad (7)$$

where  $\mathcal{C}$  is a normalization constant,  $\Gamma$  is the set of indices of all legal data association hypotheses given the model hypothesis  $\mathcal{M}_{m(l)}$ . The comprising factors ( $\mathcal{P}_{1,l}$ ,  $\mathcal{P}_{2,l}$ , etc.) have natural interpretations for why they influence the global hypothesis probability and are described in Table I. A brief sketch of the proof for Eq. (7) is provided in the next paragraph. See [18] for further detail, along with a derivation for the expressions in Table I.

First, Bayes' Rule provides the decomposition

$$P(H_l^{1:k} | Y^{1:k}) = \frac{1}{\mathcal{C}} p(Y^k | H_l^{1:k}, Y^{1:k-1}) P(H_l^{1:k} | Y^{1:k-1}), \quad (8)$$

where  $\mathcal{C} = p(Y^k | Y^{1:k-1})$  is independent of a particular hypothesis. The last factor on the right-hand side is broken down via the chain rule to:

$$P(H_l^{1:k} | Y^{1:k-1}) = P(h_l | \mathcal{M}_{m(l)}, H_{\rho(l)}^{1:k-1}, Y^{1:k-1}) \mathcal{P}_3 \mathcal{P}_5. \quad (9)$$

The other factor of (8) is expanded using a combination of Bayes' Rule and Laplace's method for approximating integrals.

$$p(Y^k | H_l^{1:k}, Y^{1:k-1}) \approx \frac{\mathcal{P}_{1,l} \mathcal{P}_{2,l}}{\sum_{n \in \Gamma} \mathcal{P}_{1,n} \mathcal{P}_{2,n}} \frac{\mathcal{P}_4}{P(h_l | \mathcal{M}_{m(l)}, H_{\rho(l)}^{1:k-1}, Y^{1:k-1})}. \quad (10)$$

Substituting (9) and (10) into (8) gives the expression in (7).

Next, we consider the calculations required for generating the  $L$ -best data association hypotheses  $\{h_l\}$  from each model hypothesis  $\mathcal{M}_m$  in Step 7 of the MHTC algorithm. For this

step, only the product  $(\mathcal{P}_{1,l} \mathcal{P}_{2,l})$  needs to be examined, as all other factors in (7) are identical for a given model hypothesis. Thus, we refer to this product as the *data association hypothesis plausibility*—it is proportional to the (posterior) probability but is technically neither a likelihood nor a normalized probability. To formulate the data association problem such that Murty's algorithm may be applied, we construct a cost matrix for the corresponding linear assignment problem of mapping current measurements to known targets (including the notions of new targets and false clusters), where the total cost of an assignment hypothesis is equivalent to using  $(\mathcal{P}_{1,l} \mathcal{P}_{2,l})$ .

Let  $\mathbf{A} \in \mathbb{R}^{G_m \times J+2G_m}$  be the *data association matrix*, where the rows are the  $G_m$  current measurements (cluster means) and the columns represent the  $J$  existing targets,  $G_m$  possible new targets, and  $G_m$  possible false clusters<sup>2</sup>. The elements of this matrix,  $[a_{gj}]$ , essentially define the likelihood of assigning the  $g$ th measurement to the  $j$ th target. For equivalence to  $(\mathcal{P}_{1,l} \mathcal{P}_{2,l})$ , the data association matrix can be defined:

$$\mathbf{A} \equiv \begin{bmatrix} a_{11} & \dots & a_{1J} & & \\ \vdots & \ddots & \vdots & \text{diag}(\alpha^{\nu}) & \text{diag}(\alpha^{\phi}) \\ a_{G_m 1} & \dots & a_{G_m J} & & \end{bmatrix}, \quad (11)$$

where  $\alpha^{\nu} = [\alpha_1^{\nu}, \dots, \alpha_{G_m}^{\nu}]^T$ ,  $\alpha^{\phi} = [\alpha_1^{\phi}, \dots, \alpha_{G_m}^{\phi}]^T$ , and

$$\begin{aligned} a_{gj} &= \frac{1}{1 - \mathbf{P}_{d,j}} \hat{\zeta}_{gj} & g = 1, \dots, G_m; j = 1, \dots, J \\ \alpha_g^{\nu} &= \frac{\lambda_{\nu}}{\lambda_{\nu} + \lambda_{\phi}} \hat{\zeta}_{g0} & g = 1, \dots, G_m \\ \alpha_g^{\phi} &= \frac{\lambda_{\phi}}{\lambda_{\nu} + \lambda_{\phi}} \hat{\zeta}_{g0} & g = 1, \dots, G_m, \end{aligned}$$

where  $\hat{\zeta}_{gj}$  is calculated from Eq. (5). Then the probability of a legal data association hypothesis  $h_l$  is proportional to the product of the elements of  $\mathbf{A}$  assigned by  $h_l$ ; that is,

$$\mathcal{P}_{1,l} \mathcal{P}_{2,l} = \mathcal{D} \prod_{(g,j) \in \tilde{h}_l} a_{gj}, \quad (12)$$

<sup>2</sup>Since each measurement may be independently assigned as a new target or false clusters, and only unique assignments are allowed, new targets and false clusters each require a column for each measurement.



where  $\mathcal{D}$  is a constant for each model hypothesis and  $\tilde{h}_l$  is simply another way of labeling the assignments in  $h_l$ :

$$\tilde{h}_l \equiv \tau_l \cup \{(g, g + J) : g \in \nu_l\} \cup \{(g, g + J + G_m) : g \in \phi_l\}.$$

(See [18] for a proof of this equivalence.)

To employ Murty's algorithm, we define the linear assignment cost matrix  $\mathbf{A}^* = -[\log a_{gj}]$ , where the elements of  $\mathbf{A}$  that are zero are instead replaced by a suitably large upper bound. Murty's algorithm may then be applied to  $\mathbf{A}^*$  to generate the  $L$ -best data association hypotheses for every model hypothesis  $\mathcal{M}_m$ .<sup>3</sup>

## V. EXPERIMENTAL RESULTS

In this section we present results from applying the proposed MHTC algorithm to extracellular neural recordings obtained by inserting an electrode into cortical tissue. Our robotic device [2] adjusts the electrode position in small increments to optimize the signal; as the relative electrode–neuron position changes, each neuron's spike waveforms evolves. Results are analyzed in a two-dimensional principal component (PCA) feature space, a common practice in the neuroscience community. Although the ground truths for these data sets are unknowable, several details are called out that imply the results' veracity.

Figure 4 displays the “tracks” of the best global hypothesis estimated by MHTC for a recording session lasting over 100 minutes. Under this hypothesis, two neurons, labeled A and B, are tracked over nearly the entire session, even through frequent changes of electrode position and difficult clustering challenges. (Track C, with few actual detections, may result from spurious groupings of outliers, or a temporarily detectable low-SNR neuron.) Note that neuron B is not initially recorded but MHTC here demonstrates its ability to recognize the appearance of new targets at  $k = 5$ .

These targets cannot be consistently tracked using ML clusters and nearest neighbor data association, as the bottom row in Figure 4 demonstrates. The ML technique is unable to maintain neuron identities over intervals of large movement, missed detections, or clustering mistakes—instead, it results in a multitude of false neuron identities and tracks that last only a short time. One reason MHTC can overcome these challenges is its ability to propagate several hypotheses; the instantaneous rank of the global hypothesis (whose tracks are displayed) is plotted in the fourth row of Figure 4. Note that during some steps, this global hypothesis that ended up being the most probable was only the second or fourth most probable at a given time, indicating that later data supported this hypothesis to “come back” and be deemed the best.

<sup>3</sup>Eq. (12) also implies that the global hypothesis probability (7) may be rewritten:

$$P(H_l^{1:k} | Y^{1:k}) \approx \frac{1}{c} \frac{\prod_{(g,j) \in \tilde{h}_l} a_{gj}}{\sum_{n \in \Gamma} \prod_{(g,j) \in \tilde{h}_n} a_{gj}} \mathcal{P}_3 \mathcal{P}_4 \mathcal{P}_5,$$

so that the assignment costs calculated during the data association hypothesis generation step are used directly to evaluate the global hypothesis probability. Note that the constant  $\mathcal{D}$  never requires calculation, as it cancels from the numerator and denominator.

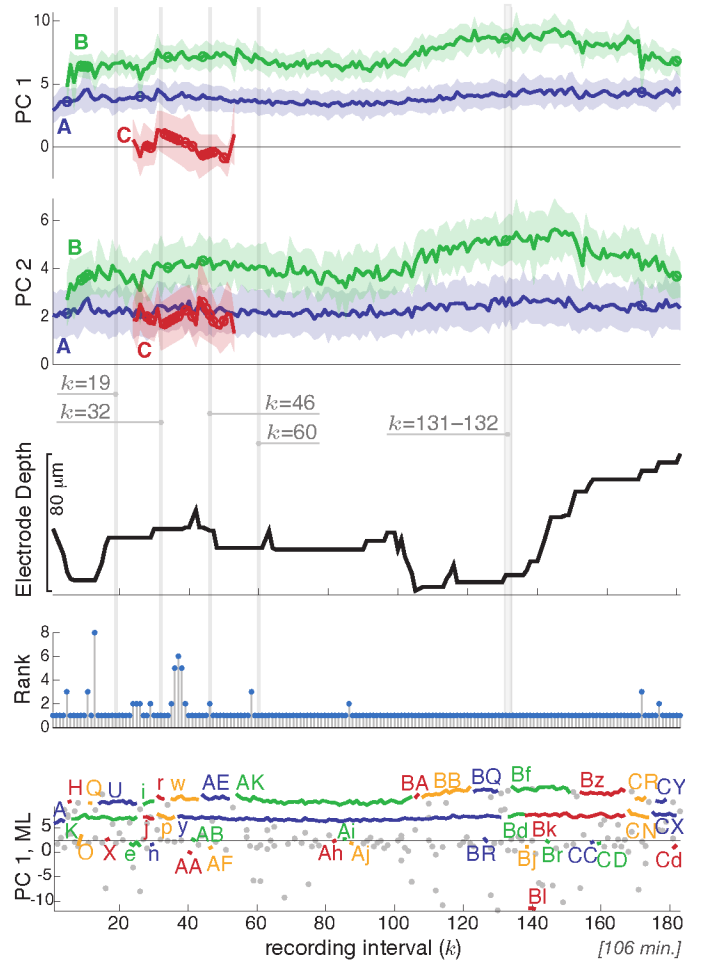


Fig. 4. MHTC results from a 106 min. session of neural data. The **top two rows** show the MHTC tracks from the best global hypothesis, in the first and second principal components, respectively. (For representation, these coordinates were obtained by PCA on the set of spike waveforms over the *entire* recording session (basis  $\mathcal{B}^{1:k}$ ), whereas the MHTC algorithm operates in the PCA coordinates derived from the spike waveforms in each interval,  $\mathcal{B}^k$ .) Each track is represented by its mean waveform (thick line), with a filled area showing two standard deviations of its assigned spikes, and is labeled by an ID (capital letter). Circles along the track line indicate a time step with a missed detection. Scale is shown on these  $y$ -axes for reference but has no clear physical interpretation due to the PCA projection. The **third row** plots the position of the electrode over the recording session. The **fourth row** plots the rank histories of this best global hypotheses for the session. The labeled time steps are further examined in Figure 5. In the **fifth row**, the tracks resulting from a maximum likelihood (ML) clustering technique with nearest neighbor data association are shown in the first principal component, for comparison to the top row. Only tracks lasting at least two time steps are colored and labeled (singleton tracks are designated by gray dots).

A detailed look at the data of certain time steps from this session is shown in Figure 5, displaying the spike waveforms recorded during that interval and their projections to the PCA space, grouped and colored according to MHTC's identified clusters. Note the lack of separation of the clusters at some time steps, demonstrating the difficulty of the clustering. Note that at  $k = 131$ , MHTC returns only one identified cluster since neuron B's spikes had insufficient evidence to make their own cluster; however, the MHTC algorithm recovers from this

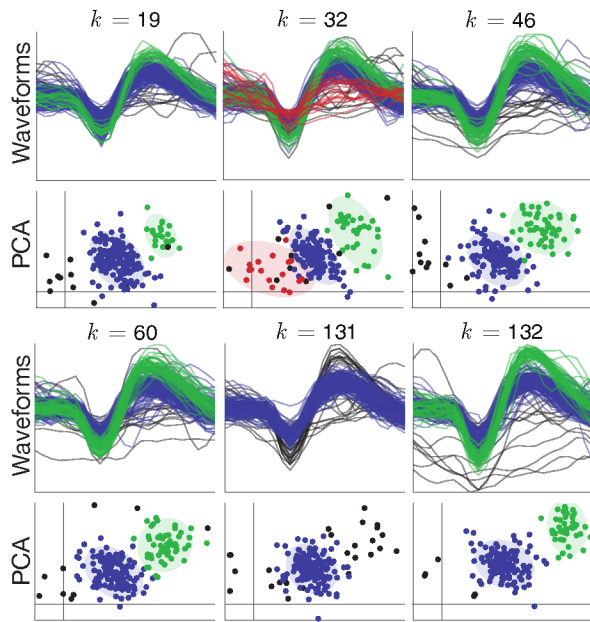


Fig. 5. MHTC clusters, showing detail of certain time steps of the session from Figure 4. Plot pairs include the full spike waveforms and their projection to the PCA feature space. Data are colored according to MHTC clustering results in both representations. PCA clusters include filled 2-sigma ellipses; black points indicate classification as “outliers.”

issue at the next interval ( $k = 132$ ) and correctly identifies neurons A and B.

## VI. CONCLUSION

The results above demonstrate the usefulness of the MHTC algorithm in a combined clustering/tracking application. MHTC is able to maintain the identities of individual neuron targets despite the signal variability over the recording session, including the introduction of a new neuron target and temporarily “silent” neurons (resulting in missed detections). The hypothesis rank plot in Figure 4 indicates how a hypothesis that may seem less likely in the current interval may be supported by future data and thus eventually prove to be the best hypothesis overall.

In conclusion, we have detailed a class of problems that entails both clustering and multitarget tracking and proposed a solution that considers these tasks in an integrated fashion. Like traditional MHT, MHTC propagates several possible solutions so that future data can help resolve current ambiguities, but, most importantly, we have expanded the classic hypothesis tree to include model hypotheses as well as data association hypotheses and have derived the expressions for the resulting probabilities. Thus, MHTC blends Bayesian methods into a robust solution for this multitarget tracking problem.

## REFERENCES

- [1] M. T. Wolf, J. G. Cham, E. A. Branchaud, and J. W. Burdick, “A miniature robot for isolating and tracking neurons in extracellular cortical recordings,” in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2008.
- [2] M. T. Wolf, J. G. Cham, E. A. Branchaud, G. H. Mulliken, R. A. Andersen, and J. W. Burdick, “A robotic neural interface for autonomous positioning of extracellular recording electrodes,” *Int. J. of Robotics Research*, 2008.
- [3] E. A. Branchaud, “An algorithm for the autonomous isolation of neurons in extracellular recordings,” Ph.D. dissertation, California Institute of Technology, Jun. 2006.
- [4] S. Chang, C. Assad, N. Mitsumoto, and J. W. Burdick, “Tracking humans with ultra-wideband radar,” (to be submitted IEEE Radar Conf 2009).
- [5] M. Andreetto, L. Zelnik-Manor, and P. Perona, “Non-parametric probabilistic image segmentation,” in *Int. Conf. on Computer Vision (ICCV)*, 2007.
- [6] S. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA: Artech House, 1999.
- [7] L. D. Stone, C. A. Barlow, and T. L. Corwin, *Bayesian Multiple Target Tracking*. Boston, MA: Artech House, 1999.
- [8] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. Orlando, FL: Academic Press, Inc., 1988.
- [9] G. W. Pulford, “Taxonomy of multiple target tracking methods,” *IEEE Proceedings — Radar, Sonar & Navigation*, vol. 152, no. 5, pp. 291–304, 2005.
- [10] D. B. Reid, “An algorithm for tracking multiple targets,” *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.
- [11] S. S. Blackman, “Multiple hypothesis tracking for multiple target tracking,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.
- [12] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, “Sonar tracking of multiple targets using joint probabilistic data association,” *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–184, Jul. 1983.
- [13] K. G. Murty, “An algorithm for ranking all the assignments in order of increasing cost,” *Operations Research*, vol. 16, pp. 682–687, 1968.
- [14] R. Danchick and G. E. Newnam, “Reformulating Reid’s MHT method with generalised Murty K-best ranked linear assignment algorithm,” *IEEE Proceedings — Radar, Sonar & Navigation*, vol. 153, no. 1, pp. 13–22, 2006.
- [15] R. P. S. Mahler, “Multitarget bayes filtering via first-order multitarget moments,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 39, no. 4, pp. 1152–1178, 2003.
- [16] M. T. Wolf and J. W. Burdick, “Bayesian clustering and tracking of neuronal signals for autonomous neural interfaces,” in *IEEE Conf. on Decision and Control (CDC)*, 2008.
- [17] G. McLachlan and D. Peel, *Finite Mixture Models*. Wiley Interscience, 2000.
- [18] M. T. Wolf, “Target tracking using clustered measurements, with applications to autonomous brain-machine interfaces,” Ph.D. dissertation, California Institute of Technology, 2008.
- [19] J. L. Beck and K.-V. Yuen, “Model selection using response measurements: Bayesian probabilistic approach,” *Journal of Engineering Mechanics*, vol. 130, no. 2, pp. 192–203, 2004.