Explicit State-Estimation Error Calculations for Flag Hidden Markov Models

Kyle Doty, *Student Member, IEEE*, Sandip Roy, *Member, IEEE*, and Thomas R. Fischer, *Fellow, IEEE*

***Abstract*—State estimation is studied for a special class of flag Hidden Markov Models (HMMs), which comprise 1) an arbitrary finite-state underlying Markov chain and 2) a structured observation process wherein a subset of states emit distinct flags with some probability while other states are unmeasured. For flag HMMs, an explicit computation of the probability of error for the maximum-likelihood filter and smoother is developed. Also, the form of the optimal filter is further characterized in terms of the time since the last flag, and this result is used to further simplify the error-probability computation. Some preliminary graph-theoretic insights into the error probability and its computation are discussed. Finally, these algebraic and structural results are leveraged to address sensor placement in two examples, including one on activity-monitoring in a home environment that is drawn from field data. These examples indicate that low error-probability filtering and smoothing can be achieved with relatively few sensors.**

***Index Terms*—Hidden Markov Models, state estimation, maximum likelihood estimation, estimation error, smart homes, wireless sensor networks, infrared sensors.**

I. Introduction

HIDDEN Markov Models (HMMs) are used in diverse applications such as speech processing [1], sensor-network-based activity monitoring [2], and mechanical failure prediction [3]. These applications leverage classical algorithms for state and parameter estimation of HMMs, including the forward-backward algorithm (filtering and smoothing), Viterbi algorithm (state-sequence estimation), and Baum-Welch algorithm (parameter estimation). Surprisingly, although these estimation algorithms are used widely, basic questions about their structure and performance remain unanswered.

Many emerging applications of HMMs, such as activity monitoring for assisted living and event-tracking using social-media data, demand 1) efficient performance analysis of estimators and/or 2) simple tools for sparse sensor placement. These applications would thus benefit from formulaic characterizations or computationally-friendly approximations for the estimation error probability. As a further step, these applications would benefit from simple insights into the estimator's structure and performance to facilitate effective sensor placement with or without complete knowledge of the HMM graph and parameters.

Manuscript received November 04, 2015; revised March 24, 2016; accepted April 16, 2016. Date of publication May 12, 2016; date of current version July 21, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wenwu Wang. This work was supported by the NSF through the IGERT program and through Grant NSF-1035369, and by the CASAS program at WSU. This work was also supported in part by the DHS HSARPA Program.

The authors are with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA (e-mail: kdoty@eecs.wsu.edu; sroy@eecs.wsu.edu; fischer@eecs.wsu.edu).

Digital Object Identifier 10.1109/TSP.2016.2568167

In this article, we study the probability of error of the maximum likelihood state estimator (both filter and smoother), for a special class of **flag** HMMs. These flag HMMs may have an arbitrary underlying finite-state Markov chain, but are constrained to have a specially-structured observation process wherein a subset of states probabilistically emit distinct flag symbols while the remaining states are unmeasured. For these flag HMMs, which are descriptive of several emerging sensing applications, an explicit formula is derived for the estimation error probability. The explicit probability-of-error analysis is shown to give insight into the form of the maximum likelihood estimator, allow computationally-attractive computations of the error, and enable development of relationships between the Markov chain's eigenstructure and the estimator/error. Via examples, we also demonstrate that the simplifications in computing the estimation error facilitate sensor placement. Specifically, sensor placement is studied in a home-activity-monitoring example based on real-world data, and for larger-dimension randomly-generated Markov chains. These examples show that accurate state estimation can be achieved while only using a small number of sensors.

The research presented here builds on a number of results in the literature, which formally analyze the error probability for HMM state and parameter estimators for narrow model classes. Specifically, a couple of works ([4] and [5]) have addressed Markov chains with asymptotically weak weights. Several other studies ([6] and [7]) are focused on two-state HMMs and their generalizations. The article [6] computes the filtering error for the two state case when the output is continuous-valued, but gets quantized. In [7], the authors analyze the error for the two-state chain with random unobserved outputs (random packet losses). Complementary to the state-estimator error analysis, several studies have pursued formal performance analysis of other HMM inference problems. For instance, both [8] and [9] bound the probability of error in model classification from observed data. Specifically, [8] presents an efficient calculation of the bound on the probability of error. The article [9] extends this analysis to bound the maximum a posteriori (MAP) estimates for different candidate HMMs, and studies how the bound changes as the number of observations increases. Alternatively to an error analysis, information-theoretic constructs such as conditional entropy have been used as proxies for the error probability and structural results have been obtained for these measures ([10]). Also relevant to our study is an important body of work on identification of hidden Markov models from statistical data [11]. Relative to these efforts, our study addresses HMMs with general hidden chains but structured observations, and develops algebraic results directly for the estimator error probability.

Object tracking and activity monitoring using wireless sensor networks is often based on HMMs. Because states and sensors correspond to physical-world locations in these applications, the

flag HMM construct wherein a subset of states are flagged is often apt. Since sensor upkeep, and even the sensors themselves, can be expensive, effective pruning, scheduling, and power management of sensors is crucial. The probability of error analysis developed here directly supports sensor selection and placement for these applications. To contextualize the work, we mention a sample of the literature on sensor networks focused on HMM performance analysis. First, [12] provides algorithms for dynamic sensors scheduling and sensor management, while [13] addresses optimal data quantization in sensor networks. [14] further looks at power efficiency in sensor networks that use data quantization. [15] also addresses power efficiency in sensor networks, but with a focus on object tracking. Finally, [16] focuses again on sensor scheduling, but approaches the problem through a MAP state estimator framework.

The state estimation problem for HMMs can be approached in several different ways, which can enable characterization of various error measures. Of importance to our work, [17] approaches state estimation from a direct Bayesian perspective rather than via an iterative approach, and characterizes an error measure that is defined in terms of a Kullback-Liebler distance. Meanwhile, alternates of the standard Viterbi algorithm have been proposed (e.g. [18]), which optimize costs other than the error probability.

The error probability analysis described here also connects to a set of research on characterizing estimation error and designing sensor placements in network processes (e.g., [19], [20]). These studies are focused on linear systems with state and process noise, for which the algebraic characterization of the minimum estimation error is classical. The recent studies on network processes specifically study local measurement of a process defined on a graph (in analogy with the flag HMM considered here), and pursue graph-theoretic characterizations of the estimation error and, in turn, sensor placement to reduce the error. The algebraic analysis of the filtering/smoothing error for flag HMMs pursued here, which is more complicated than the classical analysis for linear systems because of the filter's nonlinearity, is a starting point toward similar graph-theoretic results.

The paper is organized as follows. In Section 2, the flag HMM model is described, and the maximum-likelihood smoothing problem is reviewed. In Section 3, an explicit formula for the filtering and smoothing probability of error is derived for perfect sensors. Then, the probability of error is derived for the filter and smoother when the sensors are imperfect. Further, the form of the optimal state estimator is characterized, and this result is used to further simplify the error-probability expression. Connections of the error probability and its computation to Markov chain's graph are briefly discussed. Finally, in Section 4, the probability-of-error analysis is applied to support sensor placement in examples, including one example developed using data from a real-world Smart Home.

II. Problem Formulation

This article is concerned with state estimation performance for a special class of hidden Markov models (HMMs), wherein observations are flags of a subset of the Markov chain's states. A discrete time HMM with a finite-state underlying Markov chain is considered. Formally, the state $s[k]$ of the chain at each time k ($k = 0, 1, 2, ...$) may take integer values between 1 and n, i.e., $s[k] \in \{1, ..., n\}$. The transition matrix of the Markov chain is denoted by $A = [a_{ij}]$, where a_{ij} is the probability of transitioning from state i to state j or $P(s[k+1] = j | s[k] = i)$. Here, A is assumed to be the transition matrix for an ergodic Markov chain, but otherwise may be arbitrary. The observation model is specially structured. The Markov chain is assumed to have a subset of states, denoted as *sensed states*, that are probed. Specifically, when the current state is one of the sensed states, a flag indicating the state is outputted with a certain probability, and no flag is outputted otherwise. For all other states, no flag is outputted. Formally, we define the set F to contain all of the sensed states, say m in total. The observation $y[k]$ at time k may either be a null symbol (denoted as 0), or may identify a sensed state in F. Specifically, if the current state $s[k] = j$ is a sensed state ($j \in F$), then the observation symbol is governed by the following probabilistic model:

$$P(y[k] = i | s[k] = j) = \begin{cases} q_j & \text{for } i = j \\ 1 - q_j & \text{for } i = 0 \\ 0, & o/w \end{cases}.$$

Meanwhile, if the current state $s[k] = j$ is not a sensed state, then the observation symbol is generated according to: $P(y[k] = i | s[k] = j) = \begin{cases} 1, & i = 0 \\ 0, & o/w \end{cases}$.

We refer to the model as a whole as a **flag** HMM. The flag HMM model is a specialization of the HMM model popularized by Rabiner [1], which is apt when a subset of network states are distinctly identified by (possibly imperfect) sensing capabilities. The model is appropriate for many emerging sensor-networking applications of HMMs, such as activity monitoring in complex spaces and intrusion tracking, wherein sensors only sparsely probe a measured space but allow precise identification of these states. The special case where the sensors detect their corresponding states without error ($q_j = 1$ for $j \in F$) is referred to as a **flag HMM with perfect sensors**.

A number of different estimation problems for HMMs have been studied in the literature. The focus here is on the *filtering* and the *smoothing* problems. Filtering refers to the estimation of the current state from the sequence of observations up to the current time, while smoothing is concerned with estimation of the state at a certain time from the sequence of observations up to and beyond the current time. Formally, in the smoothing problem, the state estimator or detector seeks to determine the current state $s[z]$ from the sequence of observations $y[0], ..., y[z + z_f]$, where z_f is the number of future data points that the smoother has access to, and $z_f \geq 0$. The filtering problem is a specialization wherein $z_f = 0$.

For this paper, a maximum likelihood (ML) detector is considered. The ML detector chooses as the time-z estimate the state that has highest probability given the sequence of observations: $i_d = \arg\max_{i=1,...,n} P(s[z] = i | y[z + z_f], ..., y[0])$, where i_d is the estimated state. The ML estimate is traditionally determined using a recursive computation of the conditional state-occupancy probabilities via the *forward and backward algorithms*, see e.g., [1]. The filtering problem is addressed in exactly the same way, however only the forward algorithm is needed.

Our interest here is in computing the probability of error of the estimate. An error event $e[z]$ is said to occur at time z if the detected state is not the current state.

i.e., $i_d \neq s[z]$. Conditioned on the observation sequence, the error probability is given by $P(e[z]|y[z+z_f],...,y[0]) = 1 - \max_{i=1,...,n} Pr(s[z]=i|y[z+z_f],...,y[0])$. As an aggregate measure of detection performance, it is natural to compute the average error probability over instantiations of the HMM (i.e., over possible observation sequences). This ensemble average in general may be time-dependent. We are specifically interested in the asymptotic value of this average error (equivalently, the average error assuming the Markov chain's initial state-occupancy probabilities are the asymptotic ones), or $P(e) = \lim_{z\to\infty} P(e[z])$. Alternately, this probability can also be viewed as the probability of error in detection at a randomly-selected time, which is sufficiently large to assume that the asymptotic model is in force. While the error probability $P(e)$ is defined as an ensemble average, it is also easy to check that the time average of the error probability for a single instantiation also converges to this value.

For a general Hidden Markov Model, the average probability of error $P(e)$ does not admit an analytical characterization. In this work, an algebraic expression for $P(e)$ is obtained for flag HMMs. This algebraic characterization enables calculation of the error with low computational effort and hence also supports sensor placement, as also developed here. Additionally, the analysis gives insight into the form or structure of the optimal estimator, and the dependence of the error on the network's spectrum and topology.

III. Algebraic Expressions For the Estimation Error

Explicit algebraic computations of the error probability are presented for the filtering and smoothing problems, for flag HMMs. For each problem, the perfect-sensor case is addressed first, and then leveraged to develop an analysis for the general (imperfect sensor) case. The filtering analysis for the perfect sensors case was developed in our recent conference paper [21], but is included here for completeness, and a further simplification of the algebraic result is obtained.

For ease of analysis, we find it convenient to relabel the Markov chain and rearrange the transition matrix so that the first set of states are the sensed states, and the rest are the unsensed states. Specifically, without loss of generality, the Markov chain states can be relabeled so that the sensed states $f_1,...,f_m \in F$ are identified as states $1,...,m$, while the remaining (unsensed states) are labeled $m+1,...,n$. The transition matrix $P = [p_{ij}]$ for the modified chain can be obtained via a permutation of the original transition matrix A, see e.g., [22].

Before presenting and proving the main results on estimation error probabilities, it is useful to overview the approach taken to develop the results. The error-probability computation centrally depends on a special property of flag HMMs, that the ML detector only needs to use the observations since the last flag and up to the next flag. This is true because the state of the underlying Markov chain is known exactly at the times of the last and next flag, hence the current state has no dependence on any other measurement data. Further, only the time elapsed since the last flag, the time until the next flag, and the values of the last and next flag are relevant to detection. These simplifications allow explicit characterizations of the probability of error.

The first result presented is for the filtering problem with perfect sensors. Specifically, this result characterizes a Maximum Likelihood Filter for a flag HMM with Perfect Sensors, which we refer to as MLF-PS in short. Next, the maximum-likelihood smoothing problem with perfect sensors, referred to as MLS-PS is addressed. Lastly, results are obtained for the maximum likelihood filter and smoother for the general imperfect-sensor case, which we refer to as MLF-IS and MLS-IS, respectively.

A. Filtering Problem with Perfect Sensors

An expression for the probability of error is obtained for the maximum likelihood filter for a flag HMM with perfect sensors, which we call MLF-PS.

Theorem 1: The MLF-PS has the following average probability of error:

$$P(e) = \sum_{f=1}^{m}\sum_{k=1}^{\infty}\left[e_f^T P\tilde{P}^{k-1}e_{\overline{F}} - \max_{m+1\leq i\leq n} e_f^T P\tilde{P}^{k-1}e_i\right]\left[e_f^T \mathrm{v}_1\right] \tag{1}$$

where $\tilde{P} = [\tilde{p}_{ij}]$, such that $\tilde{p}_{ij} = 0 \forall i = 1,...,m, \forall j = 1,...,n, \forall i \neq j$, $\tilde{p}_{ii} = 1 \forall i = 1,...,m$, $\tilde{p}_{ij} = p_{ij} \forall i = m+1, ...,n, \forall j = 1,...,n$; e_i is a column vector of all 0's except with a single 1 position, i; $e_{\overline{F}}$ is a column vector where $e_{\overline{F}}(i) = 0 \forall i = 1,...,m$ and $e_{\overline{F}}(i) = 1 \forall i = m+1,...,n$; and v_1 is the right eigenvector of the transpose of the transition matrix P associated with the eigenvalue at 1 (scaled to $\|\mathrm{v}_1\|_1 = 1$).

Proof: The asymptotic average probability of error can be computed as $P(e) = \lim_{z\to\infty} P(e[z])$. To do so, the probability of error for the ML detector at time z must be found. This can be done by finding the probability of error is for a specific output sequence, and then finding the expectation over all output sequences. For the flag HMM, this computation is equivalent to finding the error probability given that the last flag was f and was k steps back, and then averaging over the allowed flags f and time gaps k. We note here that the error probability is the probability of the current state not being the ML detector's state. Formally:

$$\begin{aligned}
P(e) &= \lim_{z\to\infty} P(e[z]) \\
&= \lim_{z\to\infty} \sum_{y[z],...,y[0]} [P(e[z]|y[z],...,y[0]) \times P(y[z],...,y[0])] \\
&= \sum_{f=1}^{m}\sum_{k=0}^{\infty} [P(s[z] \neq i_d | y[z] \\
&= 0,...,y[z-k+1] = 0, y[z-k] = f) \\
&\quad \times P(y[z] = 0,...,y[z-k+1] = 0 | y[z-k] = f) \\
&\quad P(y[z-k] = f)]
\end{aligned} \tag{2}$$

Next, the conditional probability in 2 can be written as

$$P(s[z] \neq i_d | y[z] = 0,...,y[z-k+1] = 0, y[z-k] = f) = 1 - \frac{P(s[z] = i_d, y[z] = 0,...,y[z-k+1] = 0 | y[z-k] = f)}{P(y[z] = 0,...,y[z-k+1] = 0 | y[z-k] = f)} \tag{3}$$

Substituting (3) into (2), we obtain the second equation,

$$P(e) = \sum_{f=1}^{m}\sum_{k=0}^{\infty} \left[\left(1 - \frac{P(s[z]=i_d, y[z]=0,...,y[z-k+1]=0|y[z-k]=f)}{P(y[z]=0,...,y[z-k+1]=0|y[z-k]=f)}\right)\times P(y[z]=0,...,y[z-k+1]=0|y[z-k]=f)\,P(y[z-k]=f)\right]$$

which simplifies to:

$$P(e) = \sum_{f=1}^{m}\sum_{k=0}^{\infty} P(y[z-k]=f)\times [P(y[z]=0,...,y[z-k+1]=0|y[z-k]=f) - P(s[z]=i_d, y[z]=0,...,y[z-k+1]=0|y[z-k]=f)] \quad (4)$$

It remains for us to express the probabilities in 4 in terms of the HMM's parameters. To begin, notice that the $k=0$ term in the sum falls out because $P(s[z]=f|y[z]=f)=1$. Now, let us consider the probabilities in (4) for $k\neq 0$. In (4), the first probability is the that of being in sensed state f at a single time $z-k$. The second probability is the probability that there has been no flag in the last k steps. The final (rightmost) probability is the joint probability of the Markov chain being in the detected state, i_d and there being k time steps since the last flag. This is equivalent to the probability of being in the most likely state or the one with the largest probability with the last flag.

Let us work from last to first. The last probability is $P(s[z]=i, y[z]=0,...,y[z-k+1]=0|y[z-k]=f)$. This equals the probability that the Markov chain is in state i after k steps from a flag, however with the additional restriction that the chain has not passed through a sensed state again in the meantime. This probability can be computed recursively via a modified state transition matrix which absorbs trajectories that enter the sensed states, which is precisely the matrix $\tilde{P}$ defined in the theorem. Specifically, the probability can be computed as $P(s[z]=i, y[z]=0,...,y[z-k+1]=0|y[z-k]=f) = P(s[z]=i\neq f) = e_f^T P\tilde{P}^{k-1}e_i$. This is true because the first part, $e_f^T P$, computes the one-step transition probabilities out of the sensed state. After multiplication by $\tilde{P}^{k-1}$, the final $n-m$ elements of the resulting vector yields the probabilities of being in each unsensed state while not transitioning through a sensed state during those k steps. Finally, multiplication by e_i selects the i^{th} state's probability. The estimated state is the state with the largest probability that is not a sensed state. This leads to the following equation: $i_d = \arg\max_{m+1\leq i\leq n} e_f^T P\tilde{P}^{k-1}e_i$. Thus the probability of interest can be computed as:

$$P(s[z]=i_d, y[z]=0,...,y[z-k+1]=0|y[z-k]=f) = \max_{m+1\leq i\leq n} e_f^T P\tilde{P}^{k-1}e_i \quad (5)$$

The middle probability in 4 is found similarly. Since there is no flag at the current time, the total probability of getting a no-flag output sequence is equal to the sum of all probabilities of being in each unsensed state with the no-flag output sequence. i.e., $P(y[z]=0,...,y[z-k+1]=0|y[z-k]=f) = \sum_{i=m+1}^{n} P(s[z]=i, y[z]=0,...,y[z-k+1]=0|y[z-k]=f) = e_f^T P\tilde{P}^{k-1}e_{\overline{F}}$. All together, the second probability is:

$$P(y[z]=0,...,y[z-k+1]=0|y[z-k]=f) = e_f^T P\tilde{P}^{k-1}e_{\overline{F}} \quad (6)$$

Finally, the first probability in (4) is the asymptotic probability of being in the sensed state, where the asymptotic value can be assumed since the limiting value of the error is being computed. The asymptotic value of $P(y[z-k]=f)$ is well known to equal $e_f^T \mathrm{v}_1$, where v_1 is the right eigenvector of P^T associated with its (non-repeated) unity eigenvalue (scaled to have unity 1-norm $\|\mathrm{v}_1\|_1 = 1$). All together, the average probability of error is:

$$P(e) = \sum_{f=1}^{m}\sum_{k=1}^{\infty} \left[e_f^T \mathrm{v}_1\right]\left[e_f^T P\tilde{P}^{k-1}e_{\overline{F}} - \max_{m+1\leq i\leq n} e_f^T P\tilde{P}^{k-1}e_i\right]$$

■

Remark: The matrix $\tilde{P}$ is related to the matrix P as follows. The rows of $\tilde{P}$ associated with the non-sensor states are identical to the corresponding rows of P. Meanwhile, the rows of $\tilde{P}$ associated with the sensor states have a single unity entry on the diagonal, and are otherwise set to 0. Thus, $\tilde{P}$ is the transition matrix of a Markov chain which has the same transition characteristics as the original chain for the non-sensor states, but for which the sensor states are instead absorbing states (hence further transitions to the non-sensor states are disallowed).

Remark: Our focus here has been on finding the ensemble-average filtering error probability, which is also the asymptotic time-average error probability. However, the argument used above also readily allows computation of the filtering error probability for a specified observation sequence. Specifically, given the observation sequence, the probability of error at a particular time z depends only on the time since the last flag (say k) and the value of the last flag (say f). The error probability can then be computed as

$$P(e[z]|y[z],...,y[0]) = e_f^T P\tilde{P}^{k-1}e_{\overline{F}} - \max_{m+1\leq i\leq n} e_f^T P\tilde{P}^{k-1}e_i$$

An interesting further question is how quickly the time-averaged error, for a single instantiation of the HMM, approaches the asymptotic estimate, which can be viewed as a *sample complexity* issue (how many samples are needed for the estimate to be apt) [23]. We notice that the asymptotic (or ensemble-average) probability of error is found as a weighed average of the conditional error over the joint probability mass of the time since the last flag, and the value of that last flag. The sample complexity required to find the error thus depends on the rate convergence of the sample joint probability mass to the actual. Since the sequence of flags obtained and the times between them are jointly Markov, the Kullback-Leibler distance between the sample and actual distributions and hence the convergence rate can be bounded using standard theory [24]. Details are omitted.

B. Smoothing Problem

Next, the maximum likelihood smoother for a flag HMM with perfect sensors, which we refer to as MLS-PS, is characterized. In this analysis, we only consider the case where a flag has occurred both before and after the current time, since this is the typical case where new sensors data is being used to improve detection. This analysis can easily be modified to the case that no flag has yet occurred between times z and $z+z_f$, we omit the details in the interest of space.

Theorem 2: The maximum likelihood smoother for a flag HMM with multiple perfect sensors (MLS-PS) has the following average probability of error:

$$P(e)=\sum_{f_1=1}^{m}\sum_{f_2=1}^{m}\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\left[e_{f_1}^T \mathrm{v}_1\right]\left\{e_{f_1}^T P\tilde{P}^{k_1+k_2-2}E_{\overline{F}}Pe_{f_2}-\max_{m+1\le i\le n}\left(\left[e_{f_1}^T P\tilde{P}^{k_1-1}e_i\right]\left[e_i^T\tilde{P}^{k_2-1}E_{\overline{F}}Pe_{f_2}\right]\right)\right\} \tag{7}$$

where $\tilde{P}=[\tilde{p}_{ii}]$, such that $\tilde{p}_{ij}=0\forall i=1,...,m,\forall j=1,...,n,\forall i\neq j$, $\tilde{p}_{ii}=1\forall i=1,...,m$, $\tilde{p}_{ij}=p_{ij}\forall i=m+1,...,n,\forall j=1,...,n$; e_i is a column vector of all 0's except with a single 1 position, i; $e_{\overline{F}}$ is a column vector where $e_{\overline{F}}(i)=0\forall i=1,...,m$ and $e_{\overline{F}}(i)=1\forall i=m+1,...,n$; $E_{\overline{F}}$ is a diagonal matrix with the entries of $e_{\overline{F}}$ on the main diagonal; and v_1 is the right eigenvector of the transpose of transition matrix for P associated with the eigenvalue at 1 (scaled to $\|\mathrm{v}_1\|_1=1$).

Proof: The proof for the smoothing error expression is similar to that for the filtering error, except that the optimal detector and hence error computation accounts for future observations. Because of the similarity, only key steps are presented here. Specifically, the probability of error at time z can be computed by conditioning on the time and value of the last flag before z as well as the time and value of the next flag after z, and then averaging over the allowed flags f_1 and f_2 and time gaps k_1 and k_2. That is,

$$\begin{aligned}P(e)&=\lim_{z\to\infty}P(e[z])\\&=\lim_{z\to\infty}\sum_{y[z_f],...,y[0]}[P(e[z]|y[z_f],...,y[0])P(y[z_f],...,y[0])]\\&=\sum_{f_1=1}^{m}\sum_{f_2=1}^{m}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}[P(s[z]\neq i_d|y[z+k_2]\\&=f_2,y[z+k_2-1]=0,...,\\&\quad y[z-k_1+1]=0,y[z-k_1]=f_1)\\&\quad\times P(y[z+k_2]=f_2,y[z+k_2-1]=0,...,\\&\quad y[z-k_1+1]=0|y[z-k_1]=f_1)]\times P(y[z-k_1]=f_1)\end{aligned} \tag{8}$$

Using standard algebraic simplications and conditional-probability concepts (see [25] for details), the error probability can be rewritten as:

$$\begin{aligned}P(e)&=\sum_{f_1=1}^{m}\sum_{f_2=1}^{m}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}P(y[z-k_1]=f_1)\\&\quad[P(y[z+k_2]=f_2,y[z+k_2-1]=0,...,.\\&\quad y[z-k_1+1]=0|y[z-k_1]=f_1)\\&\quad-P(s[z]=i_d,y[z]=0,...,y[z-k_1+1]\\&=0|y[z-k_1]=f_1)\\&\quad\times P(y[z+k_2]=f_2,y[z+k_2-1]=0,...,y[z+1]\\&=0|s[z]=i_d)]\end{aligned} \tag{9}$$

The various probabilities in (9) can be computed in a similar way to those in the filtering analysis, so the details are omitted to save space (see [25]). The error-probability expression in the theorem statement results. ■

C. Imperfect Sensors

The filtering and smoothing problems with imperfect sensors are addressed next. Specifically, the problems are solved by transforming them to the perfect-sensor problems for an augmented hidden Markov model. Our main effort in this section is to define the augmented hidden Markov model and then argue for the equivalence between the two models. Thereupon, the analysis for the perfect-sensor case can be applied directly to characterize the detector and its performance, as we formalize in a concluding theorem.

1) The Augmented Hidden Markov Model: An augmented hidden Markov model with $n+m$ states is defined. The augmented model is defined so that each sensed state i in the original model corresponds to two states in the augmented model (states i and $n+i$), with one of these states being sensed perfectly and the other being unsensed; the remaining states have a one-to-one correspondence between the two models. Precisely, the $(n+m)\times(n+m)$ transition matrix $P'=\left[p'_{ij}\right]$ for the augmented model is defined as follows:

$$\begin{aligned}p'_{ij}&=q_j p_{ij}\ \forall i=1,...,n,\forall j=1,...,m,\\p'_{ij}&=p_{ij}\ \forall i=1,...,n,\forall j=m+1,...,n,\\p'_{i,n+j}&=(1-q_j)\,p_{ij},\forall i=1,...,n,\forall j=1,...,m,\\p'_{n+i,j}&=p'_{ij}\ \forall i=1,...,m,\forall j=1,...,n+m.\end{aligned}$$

In the augmented hidden Markov model, perfect sensing of the first m states is assumed, in analogy with the original perfect-sensor case. In the augmented model, the pair of states i and $i+n$ for $i=1,...,m$ are viewed as the *aggregate state* i, which corresponds to the sensed state i in the original imperfectly-sensed HMM. The unsensed states $i=m+1,...,n$, are defined to be aggregate states by themselves. We note that the aggregate states also define a Markov process, since the transition probabilities from each state composing the aggregate are identical. It is also assumed that the initial state occupancy probabilities for the first n states of the augmented model are identical to the state occupancy probabilities of the original chain.

The augmented hidden Markov model has been constructed to be equivalent to the original hidden Markov model, in a statistical sense. Specifically, the output sequence of the

augmented hidden Markov model is statistically identical to that of the original hidden Markov model. Further, the underlying state transition model is identical to that of the original hidden Markov model, with only the difference that each sensed state is replaced with two states in the new model—one indicates that the sensed state is achieved and the flag is displayed, while the other indicates that the chain is in the sensed state but is not flagged. These probabilistic equivalences can be formally verified through a detailed analysis of the transitions and outputs of the augmented chain; however they are quite apparent from the construction of the chain so we omit this detailed argument here, see the report [25] for these details.

From the above equivalence, it follows immediately that the maximum likelihood state estimate for the original HMM is identical to the maximum likelihood estimate of the (aggregate) state for the augmented HMM. However, the augmented HMM is a flag HMM with perfect sensors. Thus, it follows immediately that the ML estimate of the state and the aggregate state of the augmented HMM, and associated probability of error, can be computed; in this way, an explicit computation for the imperfect-sensor case can be achieved. This result is formalized in the following theorem:

Theorem 3: The maximum likelihood filter for the imperfect-sensor case (MLF-IS) has the following average probability of error:

$$P(e) = \sum_{f=1}^{m} \sum_{k=1}^{\infty} \left[e_f'^T P' \tilde{P'}^{k-1} e'_{\overline{F}} - \max_{m+1 \le i \le n+m} e_f'^T P' \tilde{P'}^{k-1} e'_i \right] \left[e_f'^T \mathrm{v}'_1 \right] \tag{10}$$

where $P' = \left[p'_{ij}\right]$, $p'_{ij} = q_j p_{ij} \forall i = 1, ..., m$, $p'_{ij} = p_{ij} \forall j = m+1, ..., n$, $p'_{i(n+j)} = (1 - q_j) p_{ij}, \forall j = 1, ..., m$, $p'_{(n+i)j} = p'_{ij} \forall i = 1, ..., m, \forall j = 1, ..., n, ..., n+m$; where $\tilde{P'} = [\tilde{p'_{ij}}]$, $(\tilde{p'_{ij}} = 0 \forall i \neq j, \quad \tilde{p'_{ii}} = 1) \forall i = 1, ..., m, \quad \tilde{p'_{ij}} = p'_{ij} \forall i = m+1, ..., n$; where $e_f'^T$ is a row vector of all 0's except with a single 1 at the most recent sensor position, f; where $e'_{\overline{F}}$ is a row vector with $e'_{\overline{F}}(i) = 0 \forall i = 1, ..., m$ and $e'_{\overline{F}}(i) = 1 \forall i = m+1, ..., n$; where the index variable k represents the number of steps since the last sensor event; and where v'_1 is the eigenvector of the transpose of transition matrix for the Markov chain, P', associated with the eigenvalue of 1 (scaled to $\| \mathrm{v}'_1 \|_1 = 1$).

The maximum likelihood smoother and error probability for the imperfect-sensors case (MLS-IS) can be obtained in exactly the same way, and the result is that P is replaced with P' in the MLS-PS expression.

D. Detector Characterization and Simplification of the Error Analysis

The maximum-likelihood filter for a flag HMM turns out to have a special form, which allows some further simplification of the filtering algorithm and also the probability-of-error computation.

To identify the special filter structure, we first note that the maximum likelihood estimate of the current (time z) state depends only on: 1) the value f of the last flag and 2) the time k since the last flag. We claim that, as the time since the last flag becomes large, the detector converges to a single state under broad conditions.

To see why, notice that the maximum-likelihood state is given by $i_d = \max_i P(s[z] = i | y[z] = 0, ..., y[z-k+1] = 0, y[z-k] = f)$. This probability can be computed via a modified Markov chain, which tracks the probability of each state given that none of the sensed states has been reached (see proofs of Theorems 1 and 2). Precisely, this conditional Markov chain replicates the original chain, but: transitions to the sensed states $1, ..., m$ are disallowed, and the remaining transition probabilities are scaled up to equal the conditional probabilities given that no transitions to sensed states have occurred. The transition matrix for this modified Markov chain, denoted as $\breve{P}$, thus is defined as follows: $\breve{P}_{ij} = 0$ for $i = 1, ..., n, j = 1, ..., m$; and $\breve{P}_{ij} = \frac{p_{ij}}{\sum_{j=m+1}^{n} p_{ij}}$ for $i = 1, ..., n, j = m+1, ..., n$. For this modified Markov chain, the sensed states are by definition transient, while at least some of the unsensed states $(m+1, ..., n)$ are recurrent. Under the condition that the unsensed states form a single ergodic class, it follows immediately that the probabilities $P(s[z] = i | y[z] = 0, ..., y[z-k+1] = 0, y[z-k] = f)$ for $i = m+1, ..., n$ approach an asymptote as k is made large. Thus, it follows that the maximum-likelihood state i_d is fixed after some look-ahead horizon, i.e., for $k \geq k'$ for some positive integer k'. Further, this asymptotic estimate is the same for all possible previous flags, although the time required to reach the asymptote may differ from each. The result is formalized in the following theorem:

Theorem 4: Consider a flag HMM whose modified Markov chain with transition matrix $\breve{P}$ has a single recurrent class, which is ergodic. The detected state, i_d, converges as a function of the time since the last flag, k, i.e., it is constant for all $k > k'$. Further this asymptotic detected state is the same regardless of the value of the last flag.

The fact that the detected state converges to a fixed value (under the broad ergodicity conditions) has two interesting implications. First, it allows off-line finite-memory storage of the detector, which then allows detector deployment without any on-line computation. Specifically, it is sufficient to maintain a table which, for each possible previous flag value, stores the optimal state estimate for $0 \leq k \leq k'$. This then allows on-line detection of the state without any real-time computation: only look-up of the estimate in the table is required.

Second, the theorem allows simplification of the probability-of-error computation. The infinite sum in the expression can be eliminated, through an explicit summation of the terms for $k \geq k'$. This can be done via a spectral (Jordan) decomposition of the matrix $\tilde{P}$ in the probability-of-error expression. Specifically, we decompose $\tilde{P}$ as $\tilde{P} = \tilde{V} \tilde{\Lambda} \tilde{W}$ where $\tilde{V}$ is a matrix containing the right eigenvectors and generalized eigenvectors, $\tilde{W}$ is a matrix containing the left eigenvectors, and $\tilde{\Lambda}$ is the block-diagonal Jordan matrix. To simplify the presentation, from here on the Jordan blocks are all assumed to be size 1, i.e., $\tilde{P}$ is diagonalizable; the analysis can be easily repeated for the general case but the notation becomes cumbersome. Considering the internal summation from k' to infinity in (1), we get:

$$\sum_{k=k'}^{\infty} e_f^T P \tilde{P}^{k-1} (e_{\overline{F}} - e_i) = \sum_{k=k'}^{\infty} e_f^T P \tilde{V} \tilde{\Lambda}^{k-1} \tilde{W} (e_{\overline{F}} - e_i)$$

where i is the (unchanging) detected state after time k'.

We then define $\tilde{\mathrm{vw}}_r = [e_f^T P \tilde{\mathrm{v}}_r][\tilde{\mathrm{w}}_r^T (e_{\overline{F}} - e_i)]$. Using the facts that $\tilde{\mathrm{vw}}_1 = 0$ and $|\tilde{\lambda}_r| < 1 \forall r = 2, ..., n$, we find:

$$\sum_{k=k'+1}^{\infty} e_f^T P \tilde{V} \tilde{\Lambda}^k \tilde{W} (e_{\overline{F}} - e_i) = \sum_{k=k'+1}^{\infty} \left[\sum_{r=2}^{n} \tilde{\mathrm{vw}}_r \tilde{\lambda}_r^k \right]$$

$$= \sum_{r=2}^{n} \tilde{\mathrm{vw}}_r \left[\sum_{k=k'+1}^{\infty} \tilde{\lambda}_r^k \right] = \sum_{r=2}^{n} \tilde{\mathrm{vw}}_r \left[\frac{\tilde{\lambda}_r^{k'+1}}{1 - \tilde{\lambda}_r} \right]$$

Substituting this back into (1), the probability of error thus becomes:

$$P(e) = \sum_{f=1}^{m} \left\{ \left(\sum_{k=1}^{k'-1} \left[e_f^T P \tilde{P}^{k-1} e_{\overline{F}} - \max_{m+1 \le i \le n} e_f^T P \tilde{P}^{k-1} e_i \right] \right) + \left(\sum_{r=2}^{n} \tilde{\mathrm{vw}}_r \left[\frac{\tilde{\lambda}_r^{k'+1}}{1 - \tilde{\lambda}_r} \right] \right) \right\} \left[e_f^T \mathrm{v}_1 \right]$$

This analysis yields a probability of error expression without an infinite sum, which may thus be viewed as a closed form expression. The expression permits exact computation of the probability of error using a finite computation, however it may be cumbersome to use in practice because 1) the time k' after which the detected state is fixed is unknown *a priori* and 2) the computation requires finding the spectrum of $\tilde{P}$. However, in practice, good approximations can be obtained without undertaking the full spectral analysis. To see how, consider the case that the time k since the last flag is large compared to $\frac{1}{1-\lambda_2}$, where λ_2 is the subdominant eigenvalue of $\tilde{P}$. In this case, from standard linear systems theory, the detector output is fixed and also probability of occurrence (the probability that the $\tilde{P}$ chain has not been absorbed into a sensor state) is small [26]. (We note here that λ_2 can be guaranteed to be real and positive since it is the dominant eigenvalue of the submatrix of $\tilde{P}$ corresponding to the transient non-sensor states.) Thus, it suffices to simply terminate the summation for some $k >> \frac{1}{1-\lambda_2}$, since the remaining infinite sum is negligible. This simple approximation can be further refined, if desired, by including only the infinite-summation term corresponding to λ_2. If even computing the dominant eigenvalue is cumbersome, the summation can instead be terminated when $\tilde{P}^{k-1} e_{\overline{F}}$ is sufficiently small. We also point out several interesting iterative algorithms for finding increasingly-tight lower bounds on asymptotic state occupancy probabilities [27], which could be exploited in characterizing the time k' and computing the infinite sum without undertaking a full spectral analysis; details are omitted.

E. Discussion: Graph-Theoretic Analyses and Sensor Placement

The algebraic and spectral analyses of the estimation error, and the characterization of the detector form, are a starting point toward developing graph-theoretic results on flag HMMs. Graph-theoretic analyses of both the error-computation effort, and the error itself, are potentially of interest. While stopping short of presenting formal lemmas/proofs, we illustrate how algebraic graph theory ideas can be used to develop such analyses.

First, with regard to complexity, the computational effort to find the error probability is seen to scale inversely with the distance of the subdominant eigenvalue of $\tilde{P}$ from the unit circle (see previous subsection). However, notice that the subdominant eigenvalue of $\tilde{P}$ is also the dominant eigenvalue of the principal submatrix of P corresponding to the non-sensed states, which is a substochastic matrix. A number of graph-theoretic characterizations of this eigenvalue have been developed [28], [29], in the context of the analyzing substochastic matrices, characterizing absorption and recurrence times in Markov chains, or bounding eigenvalues of M matrices. Based on [28], it follows that the subdominant eigenvalue of $\tilde{P}$ can be close to 1 only if there is a non-sensed state which is isolated from the sensed states, in the sense that the product of the probabilities on any path to a sensed state is small. Formal bounds on the subdominant eigenvalue and hence the computational effort can be obtained in terms of such path products, but we do not pursue this further here. Relevant to our effort here, this result suggests that placing at least one flag sensor in each strongly subnetwork of the Markov chain eases computation.

The graph topology of the Markov chain, and the locations of the sensed (flagged) states in the chain, also modulate the error probability itself. Understanding this dependence can assist in developing simple rubrics for sensor placement, and determining how pervasively a chain has to be sensed for effective detection. The dependence of the error probability on the graph and sensor locations is rather subtle. Based on the algebraic expression, we see that low-error filtering and smoothing are achieved if: 1) the sensed (flagged) states are entered with high frequency (since no error is incurred during these time steps), and 2) in between flags the trajectory of the Markov chain is highly deterministic (i.e., a particular sequence of states is highly likely, so that the chance of deviating from the most-likely state and hence incurring an error is low). The frequency with which the Markov chain enters each sensed state is determined by the asymptotic state occupancy probabilities, which are contained in the left eigenvector of P associated with its dominant unity eigenvalue; the dominant eigenvector is exactly characterized in the reversible case, and numerous graph-theoretic results are available in the general case [30], [31]. The uncertainty in the between-flags trajectory is more difficult to characterize. At its essence, this uncertainty has to do with the rate of mixing in the non-sensed part of the Markov chain: if the mixing rate is high, then the trajectory is uncertain and the error probability is large (and vice versa). Densely connected chains are fast mixing (see e.g., [32]), hence the probability of error is higher for such chains compared to sparsely connected ones. A majorization argument for the error also could be used to derive graph-theoretic results; details are again omitted.

The algebraic formula for the error, as well as the resulting spectral and graph-theoretic results, can in turn facilitate sensor-placement design. First, the algebraic expression for the error allows very rapid computation of the error probability, thus allowing optimization over possible sensor configurations whether via a combinatorial search or a greedy or other heuristic. In the examples developed in the next section, we have undertaken sensor selection via a greedy heuristic and via an exhaustive combinatorial search for smaller examples. This approach takes advantage of the simplicity of the error-probability computation. The examples also make clear that effective sensor

placement is closely tied to the graph structure of the Markov chain. The above preliminary graph-theoretic analysis suggests that rubrics for placement based on the graph topology, including degree and centrality measures, partition structure, and connectivity, may achieve good sensor placements.

IV. Examples

Two numerical examples are presented. The first example, which is based on the data gathered from a real-world *smart home*, demonstrates the applicability of the error analysis in activity monitoring in a home environment and also gives some insight into sensor placement for activity monitor. The second example addresses sensor placement in larger (30–100 state) chains, and explores the dependence of detector performance on chain size and connectivity.

A. Case Study–Smart Home Activity Monitoring

The error-probability analysis has been used to design sensor placements for monitoring elderly patients in a smart home environment. Smart home environments are living spaces with pervasive sensing (and perhaps actuation) capabilities, that can systematically monitor and respond to residents' needs. Smart home technologies may be particularly valuable for the elderly, as a means for providing cost-effective around-the-clock care while allowing individuals to remain independent and in their own homes [15] . However, implementing smart home technologies entails significant costs in terms of communication, sensing, computation, maintenance, etc. Thus algorithms are needed that allow effective monitoring with limited cost and overhead. Here, we study the problem of using sensor data on a smart-home resident's movement to estimate their activity pattern. Because of cost constraints (including implementation, maintenance, and communication costs, [2]) and the need for reliability, it is often preferable that sensors are used at only a subset of activity locations for persistent monitoring.

Using pervasive sensing data from an experimental smart home testbed, a Markov chain model with 12 states has been constructed for patient movement among activity states in the home. The data used is from the Aruba testbed in the CASAS project at Washington State University [33]. The following are the activities captured as states: 1) Other Activity, 2) Sleeping, 3) Bed to Toilet Transition, 4) Meal Preparation, 5) Relaxation, 6) Housekeeping, 7) Eating, 8) Washing Dishes, 9) Leave Home, 10) Enter Home, 11) Work, and 12) Respirate. The Markov chain is generated by only considering the activity transitions.

Our aim in this study was to evaluate activity monitoring from sparse sensing of the activity states. Specifically, ML filtering and smoothing of the activity state using perfect or imperfect sensing of a subset of the states was considered. The techniques developed in this paper were used to characterize the detectors' error probabilities, and hence to inform sensor placement. First, the MLF-PS algorithms were used to design and evaluate sensor selections of different cardinalities, using a greedy heuristic. These sets of sensors and the associated probabilities of error for filtering are shown in Table I. Since this is a relatively small Markov chain, the optimal (minimum probability of error) sensor selection for each desired cardinality was also found through a combinatorial search. In this case, the global optimum matched exactly the results found using the greedy search.

TABLE I
Simulated and Theoretical $P(e)$ for Different Sensor Placements in Smart Home Example for the MLF-PS

Number of Sensors	Optimum Sensed States	Theoretical $P(e)$	Simulation $P(e)$
1	5	0.1865	0.1861
3	5, 4, 10	0.1018	0.1015
5	5, 4, 10, 2, 7	0.0424	0.0424
7	5, 4, 10, 2, 7, 3, 9	0.0124	0.0124
9	5, 4, 10, 2, 7, 3, 9, 8, 6	0.0004	0.0004

TABLE II
Simulated and Theoretical $P(e)$ for Different Sensor Placements in Smart Home Example for the MLS-PS

Number of Sensors	Optimum Sensed States	Theoretical $P(e)$	Simulation $P(e)$
1	5	0.1712	0.1714
3	5, 4, 9	0.0898	0.0899
5	5, 4, 9, 2, 7	0.0358	0.0356
7	5, 4, 9, 2, 7, 10, 3	0.0112	0.0112
9	5, 4, 9, 2, 7, 10, 3, 8, 6	0.0004	0.0004

For comparison, the probability of error was also estimated using a simulation of the hidden Markov model and detector over 10^6 time steps. These simulation results are also shown in Table I and closely match the theoretical results. The same analysis was conducted for the MLS-PS case as well, except that the global optimum was not determined. The results for the smoother are shown in Table II. The optimal probabilities of state-estimation error for the filtering and smoothing cases are shown as a function of the number of sensors in Fig. 2.

The results show that small error probabilities can be achieved with relatively sparse sensing. A 10% estimation error rate is achieved using only 3 sensors, and only 7 sensors are required to achieve a 1% error rate. Once there are 9 sensors, the error rate is practically 0, and finally at 10 sensors, it is exactly 0. This case study demonstrates that a relatively small number of sensors can potentially allow effective activity monitoring in a smart home, and hence shows a realizable solution for reducing the number of sensors that collect/communicate Activity of Daily Living (ADLs) data in a smart home. The case study also verifies that the formal analysis of error probabilities matches error probabilities determined via Monte Carlo simulation. Finally, from the graph of the Markov chain (Fig. 1), it becomes clear that the optimal sensor placement is closely tied to the graph structure of the chain: in good designs, sensors are placed in different weakly-connected partitions of the Markov chain's graph. It is also interesting to note that the best sensor sets for the filtering and smoothing cases are similar but not identical.

We also studied how the error rate changes with the sensor fidelity. We focused on the case where all sensors have the same fidelity, i.e., $q_i = q \forall i \in F$. The dependences of the error rate on the sensor fidelity q for the filtering and smoothing problems are shown in Fig. 3. In both cases, 5 sensors were used. As expected, the performance degrades with decreasing sensor fidelity. Also, the "best" set of sensors (best greedy choice) were found for each sensor fidelity level and compared to the "best" set if the sensors

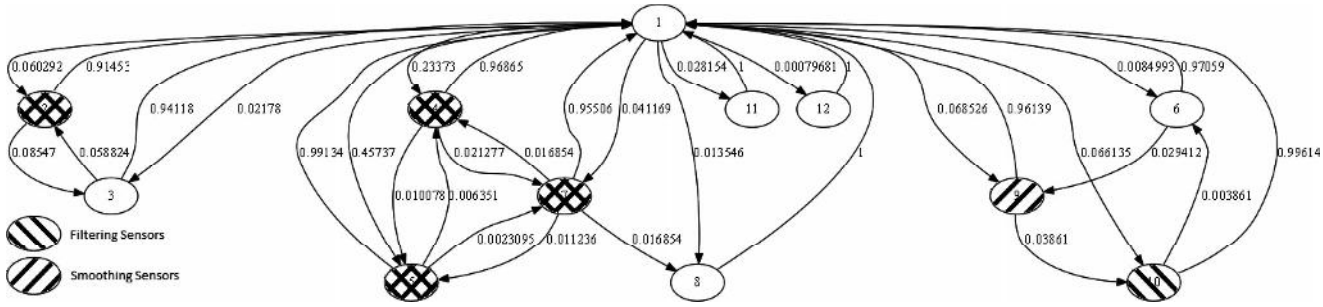

Fig. 1. Smart Home Markov Chain with 5 Sensors.

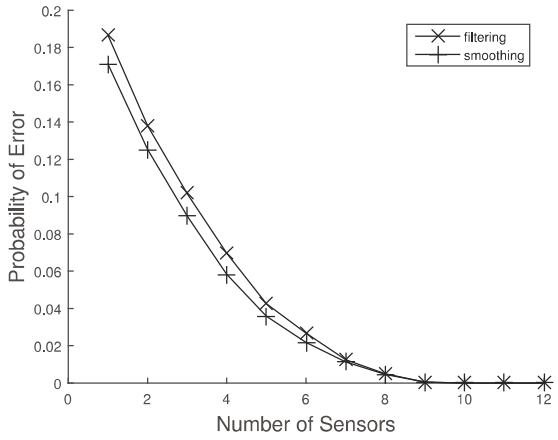

Fig. 2. Smart Home: Perfect Sensing Probability of Error vs. the Number of Sensors for the Filter and Smoother.

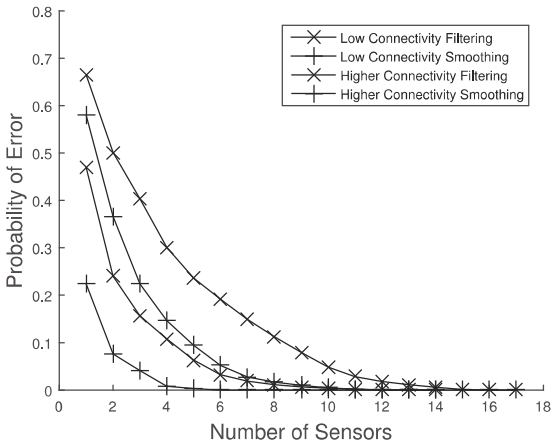

Fig. 4. 30-State Markov Chains: Perfect Sensing Probability of Error vs. the Number of Sensors for the Filter and Smoother.

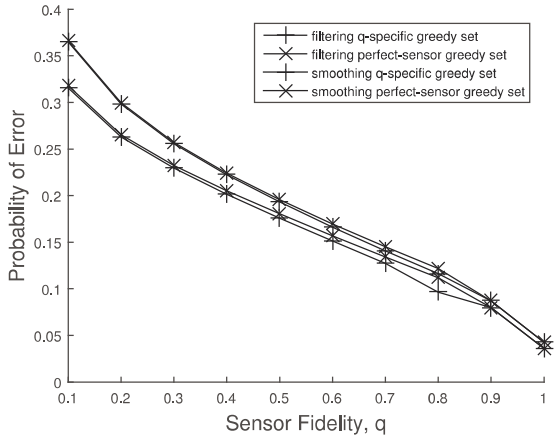

Fig. 3. Smart Home: Probability of Error vs. Sensor Fidelity for 5 Imperfect Sensors.

were perfect. For both filtering and smoothing, the q-specific set and the perfect-sensors set achieved similar performance. Thus, if the sensor fidelity is not known, there is not much performance loss in using the perfect-sensor set even when the sensors are imperfect. As a whole, detection with imperfect sensors was surprisingly effective. Even with sensor fidelities of only 10%, the probability of error for 5 sensors was around 36.5% for the filtering case and 31.5% for the smoothing case. This means that the estimated state was only wrong about a third of the time even when the sensors were only 10% accurate. Again, the Monte Carlo simulation matched the theoretical analysis.

B. Larger Examples

The flag HMM algorithms were applied to higher-dimensional (30 and 100 state) randomly-generated Markov chains, to gain further insight into sensor placement and to explore how chain characteristics impact detector performance. For these examples, the Markov chains' transition graphs were constructed by assigning bi-directional edges between each pair of states with some probability; the edge weights were also assigned stochastically (but subject to the constraint that the transition matrix is a stochastic matrix). Specifically, two 30-state chains – a sparse chain with 55 edges and a denser chain with 83 edges–were considered, as was a larger 100-state chain with 229 edges (which has a similar edge density to the denser 30-state chain). As in the smart home example, a greedy heuristic was used to select sensors for low-error filtering and smoothing, for a specified number of sensors. The performance of the MLF-PS and MLS-PS algorithms are shown as a function of the number of sensors in Fig. 4, for both the sparse and dense 30-state chains. It is seen that low probabilities of error can be achieved with a small number of sensors. The higher density chain does incur a larger filtering and smoothing error, however effective detection is still possible with fairly sparse sensing.

The performance of the MLF-IS algorithm was also analyzed in a similar fashion to that of the smart home example. Specifi-

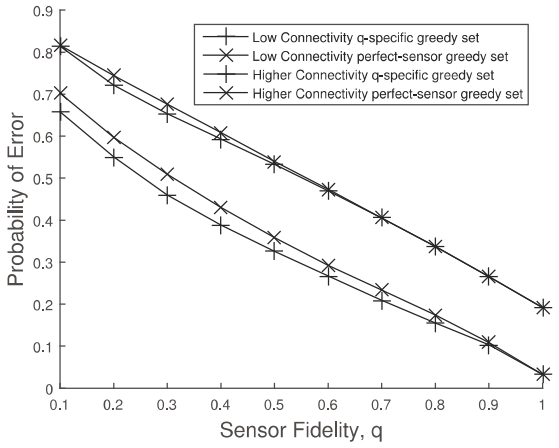

Fig. 5. 30-State Markov Chains: Probability of Error vs. Sensor Fidelity for 6 Imperfect Sensors for the Filter.

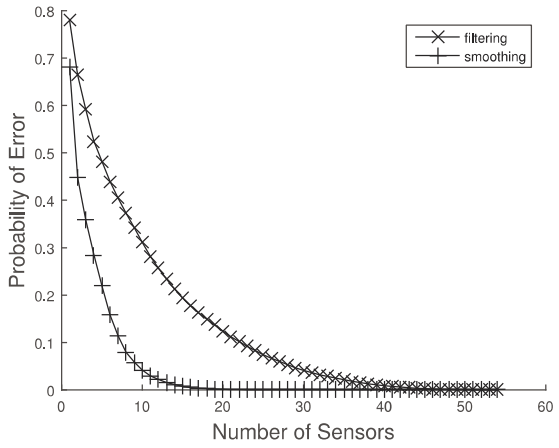

Fig. 6. 100-State Markov Chain: Perfect Sensing Probability of Error vs. the Number of Sensors for the Filter and Smoother.

cally, the greedy design was obtained for a set of 6 sensors, all with the same fidelity (see Fig. 5). As in the smart home example, the probability of error is surprisingly small with a limited number of sensors and low sensor fidelity, for both the sparser and denser chains. Also, for the denser chain, there is not much difference in error between the q-specific set of sensors (found by a greedy heuristic) and the set found in the perfect-sensor case. This leads us to believe that the difference between the q-specific set and the perfect-sensor set depends on the structure and connectivity of the Markov chain. We leave it to future work to show how the structure of the graph affects this difference.

The flag HMM error computations were also undertaken for the larger (100-state) Markov chain, to study scalability of the error-computation algorithm as well as sensor placement. The filtering and smoothing performance for greedy sensor placements, assuming perfect sensors, are shown in Fig. 6. Again, effective filtering and smoothing can be achieved with a relatively small number of sensors (with the fraction of states which must be sensed to achieve a given probability of error on par with the denser 30-state chain). The larger example also demonstrates that the probability of error and sensor placement is still relatively easy to compute.

V. Future Work

An important direction of future develop graph-theoretic and structural results on the probability of error, to support sensor placement. While some initial ideas toward a graph-theoretic analysis have been advanced in this article, much remains to be done. Specifically, simple bounds are needed on the error probability in terms of the transition matrix and flag locations, or alternately on the. Building on these analyses, algebraic graph theory tools can be brought to bear to obtain graph-theoretic results. Another direction of future work is to determine whether the error probability is submodular with respect to the sensor placement [20], with the aim of determining optimality or bounding the performance of the greedy algorithm. In the examples considered, submodularity of the error with regard to sensor selection has always been maintained (i.e., no counterexamples have been found).

Acknowledgment

The authors would like to thank Professor Diane J. Cook for fruitful discussions on the Smart Home application described in this paper.

References

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[2] D. J. Cook and L. B. Holder, "Sensor selection to support practical use of health-monitoring smart environments," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 4, pp. 339–351, 2011.

[3] L. Heck and J. McClellan, "Mechanical system monitoring using hidden Markov models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP-91)*, 1991, pp. 1697–1700.

[4] L. Shue, B. D. Anderson, and F. De Bruyne, "Asymptotic smoothing errors for hidden Markov models," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3289–3302, Dec. 2000.

[5] R. Khasminskii and O. Zeitouni, "Asymptotic filtering for finite state Markov chains," *Stochastic Process. Their Appl.*, vol. 63, no. 1, pp. 1–10, 1996.

[6] L. Shue, S. Dey, B. Anderson, and F. De Bruyne, "On state-estimation of a two-state hidden Markov model with quantization," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 202–208, Jan. 2001.

[7] A. S. Leong, S. Dey, and J. S. Evans, "Probability of error analysis for hidden Markov model filtering with random packet loss," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 809–821, Mar. 2007.

[8] E. Athanasopoulou and C. N. Hadjicostis, "Probability of error bounds for failure diagnosis and classification in hidden Markov models," in *Proc. 47th IEEE Conf. Decision Contr. (CDC 2008)*, 2008, pp. 1477–1482.

[9] C. Keroglou and C. N. Hadjicostis, "Bounds on the probability of misclassification among hidden Markov models," in *Proc. 2011 50th IEEE Conf. Decision Contr. Eur. Contr. Conf. (CDC-ECC)*, 2011, pp. 385–390.

[10] D. Hernando, V. Crespi, and G. Cybenko, "Efficient computation of the hidden Markov model entropy for a given observation sequence," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2681–2685, Jul. 2005.

[11] M. Vidyasagar, "The complete realization problem for hidden Markov models: A survey and some new results," *Math. Contr., Signals, Syst.*, vol. 23, no. 1–3, pp. 1–65, 2011.

[12] V. Krishnamurthy, "Algorithms for optimal scheduling and management of hidden Markov model sensors," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1382–1397, Jun. 2002.

[13] M. Huang and S. Dey, "Distributed state estimation for hidden Markov models by sensor networks with dynamic quantization," in *Proc. Intell. Sensors, Sensor Netw., Inf. Process. Conf.*, 2004, pp. 355–360.

[14] N. Ghasemi and S. Dey, "Power-efficient dynamic quantization for multisensor HMM state estimation over fading channels," in *Proc. 3rd Int. Symp. Commun., Contr. Signal Process. (ISCCSP 2008)*, 2008, pp. 1553–1558.
[15] H. Kang, X. Li, and P. J. Moran, "Power-aware Markov chain based tracking approach for wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC 2007)*, 2007, pp. 4209–4214.
[16] D. Jun, D. M. Cohen, and D. L. Jones, "A direct algorithm for joint optimal sensor scheduling and MAP state estimation for hidden Markov models," in *Proc. 2013 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 4212–4215.
[17] K. Yamazaki, "An asymptotic analysis of Bayesian state estimation in hidden Markov models," in *Proc. 2011 IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP)*, 2011, pp. 1–6.
[18] I. Zeljkovic, "Decoding optimal state sequence with smooth state likelihoods," in *Proc. 1996 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP-96)*, 1996, vol. 1, pp. 129–132.
[19] M. Xue, W. Wang, and S. Roy, "Security concepts for the dynamics of autonomous vehicle networks," *Automatica*, vol. 50, no. 3, pp. 852–857, 2014.
[20] T. H. Summers and J. Lygeros, "Optimal sensor and actuator placement in complex dynamical networks," 2013, arXiv Preprint arXiv:1306.2491.
[21] K. Doty, S. Roy, D. Sahabandu, and R. Saeedi, "Explicit estimation-error-probability computation and sensor design for flag hidden Markov models," in *Proc. 2015 49th Annu. Conf. Inf. Sci. Syst. CISS)*, 2015, pp. 1–6.
[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[23] M. A. Dahleh, T. V. Theodosopoulos, and J. N. Tsitsiklis, "The sample complexity of worst-case identification of FIR linear systems," in *Proc. 32nd IEEE Conf. Decision Contr.*, 1993, pp. 2082–2086.
[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.
[25] K. Doty, S. Roy, and T. R. Fischer, "Explicit state-estimation error calculations for flag hidden Markov models," [Online]. Available: http://www.eecs.wsu.edu/∼sroy, to be included in K. Doty's thesis (in preparation).
[26] W. J. Rugh, *Linear System Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1996, vol. 2.
[27] A. Busic and J.-M. Fourneau, "Iterative component-wise bounds for the steady-state distribution of a Markov chain," *Numer. Linear Algebra Appl.*, vol. 18, no. 6, pp. 1031–1049, 2011.
[28] D. Hartfiel and C. D. Meyer, "On the structure of stochastic matrices with a subdominant eigenvalue near 1," *Linear Algebra Appl.*, vol. 272, no. 1, pp. 193–203, 1998.
[29] A. Berman and R. J.Plemmons, "Nonnegative matrices," *Math. Sci., Classics Appl. Math.*, vol. 9, 1979.
[30] R. G. Gallager, *Discrete Stochastic Processes*. Berlin, Germany: Springer Verlag, 2012, vol. 321.
[31] R. Grassi, S. Stefani, and A. Torriero, "Some new results on the eigenvector centrality," *Math. Sociol.*, vol. 31, no. 3, pp. 237–248, 2007.
[32] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing markov chain on a graph," *SIAM Rev.*, vol. 46, no. 4, pp. 667–689, 2004.
[33] *WSU CASAS Dataset*, [Online]. Available: http://ailab.wsu.edu/casas/datasets.html
[34] K. Doty, S. Roy, and T. R. Fischer, "State estimation for flag hidden markov models with imperfect sensors," in *Proc. Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2016, pp. 360–365.
[35] D. Cook, M. Schmitter-Edgecombe, A. Crandall, C. Sanders, and B. Thomas, "Collecting and disseminating smart home sensor data in the CASAS project," in *Proc. CHI Workshop Develop. Shared Home Beha. Datasets Adv. HCI Ubiquitous Comput. Res.*, 2009, pp. 1–7.
[36] J. S. Evans and V. Krishnamurthy, "Hidden Markov model state estimation with randomly delayed observations," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2157–2166, Aug. 1999.
[37] O. Cappe, E. Moulines, and T. Ryden, "Inference in Hidden Markov Models," 2009.
[38] J. G. Proakis, M. Salehi, N. Zhou, and X. Li, *Communication Systems Engineering*. Upper Saddle River, NJ, USA: Prentice Hall, 1994, vol. 94.
[39] R. Merris, "Laplacian graph eigenvectors," *Linear Algebra Appl.*, vol. 278, no. 1, pp. 221–236, 1998.

Kyle Doty (S'08) graduated with a B.S. degree in electrical engineering magna cum laude from Washington State University in 2011. He is currently working towards a Ph.D. degree in electrical engineering in the areas of signal processing and controls. His research focus has been on state estimation of hidden Markov models and satellite communications.

Sandip Roy (M'11) is a Professor and Associate Director in the School of Electrical Engineering and Computer Science at Washington State University. His research is concerned with developing strategic decision-support capabilities for large-scale infrastructures by developing and using network-control-theory techniques. He holds a B.S. degree from the University of Illinois at Urbana Champaign, and S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering.

Thomas R. Fischer (S'76–M'79–M'88–F'96) received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts, Amherst, and the Sc.B. degree magna cum laude from Brown University. From June 1975 until August 1976, he was a Staff Engineer at the Charles Stark Draper Laboratory, Cambridge, MA. From 1979 until 1988 he was with the Department of Electrical Engineering, Texas A&M University. Since January 1989 he has been a Professor in the School of Electrical Engineering and Computer Science at Washington State University. From August 2000 through June 2004 he was Director of the School of EECS. His current research interests include data compression, image and video coding, joint source/channel coding, digital communications, and digital signal processing. From 1989 to 1992 Professor Fischer was Associate Editor for Source Coding for the IEEE TRANSACTIONS ON INFORMATION THEORY. From 1995 to 1997 he was a member of the Information Theory Society Board of Governors. He has served on the Program Committee for several Workshops, Symposia, and Conferences. He was a co-recipient of the IEEE Signal Processing Society's 1993 Senior Award in the Speech Processing Area. In 1996 he was selected Fellow of the IEEE.