

RESEARCH ARTICLE

A Belief Propagation Based State Estimator for Semi-Intrusive Load Monitoring System

QIHONG DUAN¹, FENG LI¹, AND JUNRONG LIU^{ID}²

¹School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

²School of Mathematics, Northwest University, Xi'an, Shaanxi 710069, China

Corresponding author: Junrong Liu (jrly@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11991023 and Grant 12271433; and in part by the Research Achievement Cultivation Fund of Northwestern University, Fundamental Research Funds for the Central Universities.

ABSTRACT For industrial application of load monitoring techniques, it is important to establish high-performance state estimators of low-cost and low frequency smart meters (SM) and sensors in a power system, which can run under resource-constrained computing units. Because household electronic appliances often tap power from fixed sockets, a finite state table for the corresponding sensors is suitable and convenient. However, SM in the main line may have an enormous state table. In this study, we propose a belief propagation (BP) algorithm to calculate the power consumption of electronic appliances in a semi-intrusive load monitoring (SILM) system whose SM and sensors have state tables with sizes varying largely. The novelty of the proposed method lies in a continuous approximation to a large state table and a switching scheme between discrete and continuous parts of the SILM system. With datasets from numerical simulations and a real-world experimental SILM system in a set of high-density school buildings within a secondary distribution network, the proposed BP algorithm is compared with relevant state-of-the-art algorithms. The results show that the proposed algorithm achieves a percentage of error (8%), which outperforms the percentage achieved by the other methods, a linear state estimation of 99%, a hidden Markov model of 21%, and a full-discrete BP algorithm of 11%. In addition, the complexity of the proposed algorithm is the least of all methods, and the proposed algorithm can run by SoC on concentrators.

INDEX TERMS Smart meters, factor graph, BP algorithm, reference probability.

NOMENCLATURE

$\hat{x}_{i,j}$	MAP estimator for the state of SM or sensor indicated by i, j .
$p_i, i \neq 0$	probability function about state of the i -th SM and its sensors' states.
$q_{i,j}$	probability density function (pdf) of the measurement error at the SM or sensor indicated by i, j .
V	state table of the system as a whole.
$v_i, i \neq 0$	constraint about the i -th SM and its sensors.
H	true value of the system as a whole.
U	measurement error of the system as a whole.
X	state of the system as a whole.

Z	measurement value of the system as a whole.
$\mathcal{M}_{M_0,i \rightarrow C_0}$	message from the i -th SM to the root.
$\mathcal{M}_{C_0 \rightarrow M_0,i}$	message from the root to the i -th SM.
\mathcal{P}	real-world probability .
\mathcal{P}^\dagger	reference probability under which SMs and sensors are independent.
E^\dagger	expectation corresponding to the reference probability.

I. INTRODUCTION

The development of technology and the decline in prices make smart meters (SM) attractive, and hence there is a significant proliferation of them in electric power distribution systems [1]. Intrusive load monitoring (ILM) techniques install a sensor inside each household electronic appliance in the customers' environment. Although they can present more

The associate editor coordinating the review of this manuscript and approving it for publication was Haris Pervaiz ^{ID}.

accurate measurements than other techniques, people often worry about health hazards with radio waves from sensors and leaks of personal privacy by them [2]. An SM based on non-intrusive load monitoring (NILM) receives information at a single point of measurement outside a building and calculates the individual consumption of each electronic appliance in the building by a disaggregation algorithm. NILM techniques do not install a sensor inside end customers' electronic appliances, and have less violation to customers. However, the accuracy of measurement is a challenge for NILM. Semi-intrusive load monitoring (SILM) is a compromise of ILM and NILM in both accuracy of data and user experience, which uses SM outside the building and multiple cheap plug-and-play monitoring sensors attached to circuit breakers or sockets inside the building [3].

Identifying electronic appliances is one of the important issues of NILM and SILM, and there are proven solutions for the issue [4], [5], [6], [7]. State estimation, which calculates power consumption for each electronic appliance, is another important issue. Solutions [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], based on different ideas, offered kinds of state estimation. However, solutions in the literature today are with expensive measurement equipment working at high frequency for high accuracy measurement of electrical parameters. Low-cost and high-performance state estimators are important for the industrial application of NILM and SILM [21], [22]. Few low-cost state estimators utilize features of the system and keep a balance between their performance and the accuracy of measurement equipment [23], [24], [25]. Today, such NILM and SILM install low frequency sensors [24] to measure variables in the system, use mini netbook [23] or microcontroller unit [25] for computation.

This study addresses low-cost and high-performance state estimation for a power distribution system with SILM. Based on full noised data not only from a target sensor but also from other sensors and SMs, the proposed belief propagation (BP) algorithm obtains a high-performance state estimator for the target with low-cost resource-constrained computing units. The performance of the proposed algorithm is based on reasonable state tables for different kinds of sensors and SMs in the system, a continuous approximation to a large state table, and a switching scheme between discrete and continuous parts of the SILM system. Concretely, we study an experimental SILM system in a set of high-density school buildings within a secondary distribution network. The electronic appliances in the power systems may change sometimes, but we have good apriori information about them to ensure their identification. Because of capitalized cost of the SILM system, there are sensors with low accuracy and low frequency in the branches, and we need to evaluate observed variables with the noise of measurement. In the system, a sensor may connect to few kinds of electronic appliances and it has a discrete state table. An SM in the main line connects to several sensors in different branches, and the SM reports a current which is the total of the branch currents.

Therefore, the measuring value at the SM is assumed to be continuous. Keeping the consistency of the algorithms' underlying mechanism, we establish a new form of the BP algorithm by which nodes in a network exchange conditional expectations of involved variables. As details of computation of conditional expectations, which have different forms for discrete and continuous variables, are hidden for nodes, the method can present consistent state estimation for every part of a mixed system with discrete and continuous variables.

The rest of the paper is organized as follows. Section II reviews related work and shows gaps related to high-performance state estimation data followed by the proposed low-cost objective. Section III describes the experimental SILM system and its model. Section IV presents a consistent BP algorithm for discrete and continuous parts under general assumptions. Simulation results to test the performance of the proposed method and discussions are in Section V. Section VI discusses the result of the experimental SILM system and Section VII concludes the paper and details of calculations and proofs are in Appendixes.

II. RELATED WORK AND NOMENCLATURE

To illustrate the role of state estimation, let us scan the main parts in ILM, NILM, and SILM. Identifying electronic appliances by disaggregation algorithms is the first step of ILM, NILM, and SILM [5]. Disaggregation algorithms include feature extractions as well as load signatures routines, which require apriori information of different equipment and depend highly on the training data [4], [6]. With the information provided by sensors inside the building, identifying an electronic appliance is relatively easier for SILM than NILM when the appliance taps power from wires or sockets with fixed positions [7]. Calculating power consumption, or state estimation, for each electronic appliance is another important issue of ILM, NILM, and SILM. Because there are many sensors and SMs in a power distribution system and their data are related, good state estimation can derive measurements of higher accuracy for an appliance than its raw value with random noise. A classical and efficient state estimation, linear state estimation with weighted least square, is fit for large distribution systems with continuous measurement values [8]. The method assumes that the variables in the system, such as current, phase angle, frequency, active power, reactive power, and apparent power, are real or complex numbers with Gaussian noise.

Spite that linear state estimation is state-of-the-art, there are many other modern methods for state estimation in today's electric power systems. A state estimation by neural networks is investigated by [9], and [10] proposed a random forest algorithm of state estimation for appliances in office buildings. The Bagging classification algorithm and fuzzy analysis have good performances in state estimation and load monitoring, as reported by [11] and [12]. As an important artificial intelligence (AI) method, the hidden Markov model (HMM) makes the state estimation of an individual load by comparing different steady-state power levels of

appliances. A modified HMM based on [13] and [14] shows high accuracy after a long burning time in our test, which is consistent with the theoretical result [15]. Different from linear state estimation which handles continuous measurement values only, the Bayesian network and its BP algorithms are suited for estimating discrete random variables [16], [17]. The method assumes that a variable in such a power distribution system has a value chosen from several candidates, uses a factor graph to describe probabilistic dependencies among different variables in the system, and calculates the posterior distribution of variables to obtain a measurement of higher accuracy [16], [18]. The Bayesian network and its BP algorithms are fit for smaller power distribution systems and need more information about variables than a linear state estimation [20]. In our simulating experiment, a Bayesian and BP algorithm presented by [17] obtains good accuracy.

For a slightly large power distribution system, we must handle synchronously discrete and continuous measurement values and there is not a suitable state estimation algorithm. Apriori information, such as few candidate states for a sensor, can improve estimation. However, because the current in the main line is the total of the branch currents, even though the assumption of discrete variables is reasonable for every branch current, values of some variables about main lines have lots of candidates, and continuous measurement values become necessary assumptions for them. Most algorithms can treat continuous variables well, and Bayesian network and BP algorithms are fit for discrete variables. With different assumptions, BP algorithms [18], [27], [28] can analyze a system with Gaussian variables. However, to apply those algorithms, every variable in the power system must have continuous measurement values. [29] studied a mixed system with discrete variables and continuous variables and used a deeping learning algorithm to learn the nonlinear relation between discrete and continuous parts of the system, but the algorithm is extremely time-consuming.

When focusing on the industrial application of NILM and SILM, we have few algorithms of state estimation, because low-cost are important for such cases and hence most candidate algorithms have been affected deeply [21], [22]. Such a system installs low frequency sensors for few kinds of electric parameters. For example, [24] proposes a system which removes voltage sensors from the hardware design. Hence, we must handle great noise in a small dataset from such a system. Moreover, low compute-demanding and low memory-intensive are also important for such a system [23], [25], which is a challenge when selecting or designing algorithms.

To that end, a state estimator, which can handle discrete and continuous variables synchronously and can run with limited computing resources, has great potential. This study proposes such a belief propagation (BP) algorithm to calculate the power consumption of electronic appliances in a semi-intrusive load monitoring (SILM) system.

Some symbols have a subscript indicating the involved sensor or SM. For example, a symbol $h_{i,j}$ in the paper represents the current at the j -th sensor connected up to the i -th SM, and

$a_{0,k}$ is for the k -th SM. When considering all SMs and sensors in the system as a whole, we use a matrix $\mathbf{H} = (h_{i,j})$, which represents true values of current in the system. Moreover, a random variable is printed in italic type or bold italic type, and a symbol with roman fonts or bold roman fonts denotes a possible value of the random variable. The nomenclature of important symbols is as follows.

III. PRACTICAL ENVIRONMENT, THE MODEL AND ITS FACTOR GRAPH

A. EXPERIMENTAL SILM SYSTEM AND ITS ENVIRONMENT

We study parts of a simple radial distribution system without any feeder interconnections. The radial system has independent feeders, and each end customer is connected to one feeder. The shadow region in Fig. 1 illustrates a part of the radial system, which is a school building consisting of discussion rooms and classrooms. Inside rooms, monitoring sensors attach the wires near switches and sockets, and outside a room, SM with high accuracy attach feeders. Fig. 2 presents monitoring devices in the experimental SILM system. For several neighbouring rooms, there is a wireless concentrator which collects data of SM and sensors in those rooms, according to the standard IEC61334-4-32/IPv4. The server in the SILM system deals with concentrators, and the server does not receive any data directly from sensors.

Fixed electronic appliances in the building includes lots of lamps with different kinds, air-conditioners, desk computers, projectors. There are several electric kettles and heaters in some rooms. Students, teachers and staff may take personal laptop computers into the building for studying and teaching. As a lecture lasts 50 minutes, the pattern of a laptop computer does not change during this period. The number of laptop computers is not big because most students do not take them. Chargers are not permitted but chargers for mobile phones may be taken into classrooms. Features of this experimental environment includes: the number of sensors is far larger than the number of electronic appliances except lamps; the electronic appliances and their positions do not change during a lecture of 50 minutes; the power of appliances except a few laptop computers and chargers are not continuously varying. Moreover, the key variables are current values of every electronic appliance.

Because there are a few laptop computers and chargers in the system, we use effectively a successive sample of current to identify possible laptop computers [31]. Hence we can assume that the current value of every sensor or electronic appliance has several discrete candidates and we have good apriori information about its distribution. In a test of the SILM system in a few rooms, every sensor can be replaced by an SM as alternative, and another state estimation algorithm of searching a lookup table makes efficient and correct identification for electronic appliances with the accurate value of current from those SM. However, replacing all sensors with SM is an uneconomical plan and we need a practical and efficient state estimation technique based on the dataset from all sensors and SM.

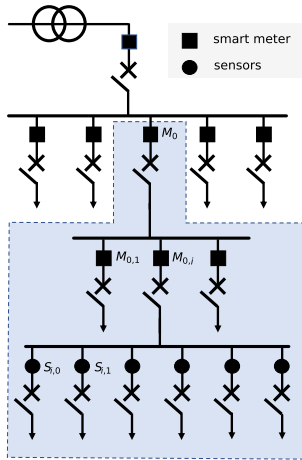


FIGURE 1. The investigated part of radial distribution system is in the shadow region. Except the root, each SM has many sensors, but only sensors from $M_{0,i}$ are illustrated in the figure.

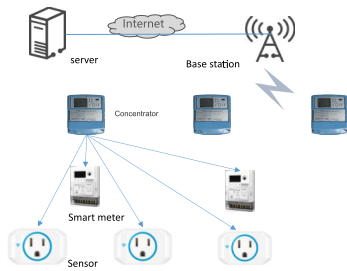


FIGURE 2. Devices in the SILM system. Only parts of SMs and sensors of the wireless concentrator on the left are illustrated in the figure.

B. MODEL OF THE SILM SYSTEM

According to the standard measurement model [30], we have that for the system in Fig. 1,

$$\mathbf{Z} = \mathbf{H} + \mathbf{U}, \tag{1}$$

where $\mathbf{H}(\mathbf{X})$ is the matrix of true values, \mathbf{Z} is the matrix of measurement values, and \mathbf{U} is the matrix of measurement errors. We write the matrix of state variables values $\mathbf{X} = (\mathbf{x}_0^T, \mathbf{x}_1^T, \dots, \mathbf{x}_n^T)^T \in R^{(n+1) \times (m+1)}$, and $\mathbf{x}_0 = (x_{0,0}, x_{0,1}, \dots, x_{0,m})$. For $j = 1, 2, \dots, n$, $x_{0,j}$ denotes the state variable for the j th SM, $M_{0,j}$ in Fig. 1, and $x_{0,0}$ is the state variable for a root SM, M_0 in Fig. 1. For $i \geq 1$, we write $\mathbf{x}_i = (x_{i,0}, x_{i,1}, \dots, x_{i,m})$ and $x_{i,j}$ denotes the state variable for a sensor $S_{i,j}$. The numbers of sensors from different SM may vary, and we replace some $x_{i,j}$ by blanks for this case. Components of other symbols $\mathbf{Z}, \mathbf{H}, \mathbf{U}$ have the similar meaning. There are n SM (except the root one) in the system and an SM has at most $m + 1$ sensors.

For our scenario, the raw data consists of the current value \mathbf{Z} at every SM and sensor. At a time point, the symbols $\mathbf{Z} = (z_{i,j}), \mathbf{H} = (h_{i,j}), \mathbf{U} = (u_{i,j}) \in R^{(n+1) \times (m+1)}$ are matrices. $z_{0,0}, h_{0,0}, u_{0,0}$ are the measurement value, measurement function, and measurement error, respectively, of the current at the root SM, M_0 in Fig. 1. For other SMs, $z_{0,i}, h_{0,i}, u_{0,i}$ are of the i th SM, $M_{0,i}$, and $z_{i,j}, h_{i,j}, u_{i,j}$ are of a sensor $S_{i,j}$. We assume

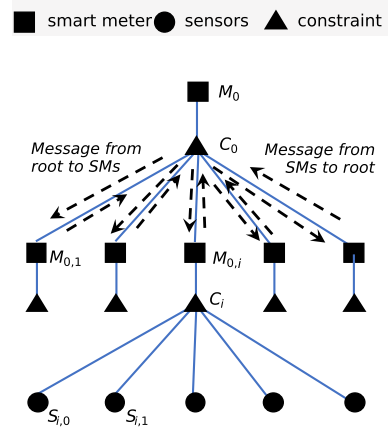


FIGURE 3. The investigated radial distribution system and the flooding schedule of messages in simple diagrammatic form.

that elements in \mathbf{U} are independent random variables with known variance and the mean 0, and the noise vector \mathbf{u}_0 of all SM has a Gaussian distribution.

When the state $x_{i,j}$ of an SM or sensor is known, the corresponding theoretical value of current value $h_{i,j}$ exists. Hence, there is a measurement function connects $h_{i,j}$ with $x_{i,j}$ for each pair of i, j . Sometimes and the measurement function $h_{i,j} = h_{i,j}(x_{i,j})$ is a one-to-one function of the state variable $x_{i,j}$.

The state table or the state space for an SM or sensor consists of all possible states of the corresponding random variable $x_{i,j}$, and we often set the state table larger than necessary. When applying our method to the experimental SILM system, we examine possible kinds of electronic appliances for each room and apply the disaggregation algorithm in Section III-A for a successive sample to identify electronic appliances when necessary. According to those method, we can ensure good apriori information about the state table for each sensor.

C. FACTOR GRAPH

The features of the system imply constraints of state variables and measurement functions. As the currents satisfy the Kirchhoff Current Law (KCL), we have the following constraint between the root SM and other SMs in the scenario given in Section III-B.

$$h_{0,0} = h_{0,1} + \dots + h_{0,n}. \tag{2}$$

And for each SM and sensors from it, we have that for $i = 1, 2, \dots, n$,

$$h_{0,i} = h_{i,0} + \dots + h_{i,m}. \tag{3}$$

The markovian relations between SM, sensors, and constraints (2), (3) are illustrated by Fig. 3 with a factor graph, which is introduced by [16] for power grids. The node C_0 in Fig. 3 corresponds to the constraint (2). A node $C_i, i = 1, 2, \dots, n$, corresponds to one of the constraint (3). When two variable nodes, such as $M_{0,i}$ and $S_{i,j}$, are connected by a constraint node, C_i , then there exists an electrical correlation

between them. Two variable nodes which are not connected by a constraint node, such as sensors from different SM, are independent. For example, the factor graph in Fig. 3 implies that $\mathbf{x}_1, \dots, \mathbf{x}_n$ are independent random vectors.

Given measurement values \mathbf{Z} , we will calculate the posterior distribution $\mathcal{P}(X = \mathbf{X} | Z = \mathbf{Z})$ of state variables. Note that a symbol with roman fonts or bold roman fonts denotes a possible value of the random variable. Because the factor graph in Fig. 3 is a tree, it implies a straightforward method to compute the posterior distribution.

IV. MAXIMUM A POSTERIORI PROBABILITY ESTIMATOR

For a given observation value of measurement, we will study the posterior distribution of state variables regarding the observation value and will present a BP algorithm to calculate the maximum a posteriori probability (MAP) estimator of state variables.

A. POSTERIOR DISTRIBUTION AND MAP ESTIMATOR

Let a state table $V_{i,j}$ be the set of all possible values of a state variable $x_{i,j}$. Because the state table $V_{i,j}$ is finite, the set

$$V = \{(x_{i,j}) : x_{i,j} \in V_{i,j}\} \quad (4)$$

is a finite set.

Note that each constraint C_i in Fig. 3 influences the real-world probability $\mathcal{P}(X = \mathbf{X})$, where $\mathbf{X} = (x_{i,j}) \in V$ and we write $\mathbf{x}_i = (x_{i,0}, x_{i,1}, \dots, x_{i,m})$. For the scenario in Section III-B, the state variables of an SM and its sensors follow a constraint (3) from the KCL. Hence, for $i = 1, 2, \dots, n$,

$$p_i(x_{0,i}, \mathbf{x}_i) = v_i(x_{0,i}, \mathbf{x}_i) \mathcal{P}(\mathbf{x}_i = \mathbf{x}_i), \quad (5)$$

where v_i describes the constraint C_i given in Fig. 3.

$$v_i(x_{0,i}, \mathbf{x}_i) = \begin{cases} 1, & h_{0,i}(x_{0,i}) = \sum_{j=0}^m h_{i,j}(x_{i,j}), \\ 0, & h_{0,i}(x_{0,i}) \neq \sum_{j=0}^m h_{i,j}(x_{i,j}). \end{cases} \quad (6)$$

Similarly, the constraint (2) for SM corresponds to a function

$$v_0(\mathbf{x}_0) = \begin{cases} 1, & h_{0,0}(x_{0,0}) = \sum_{j=1}^n h_{0,j}(x_{0,j}) \\ 0, & h_{0,0}(x_{0,0}) \neq \sum_{j=1}^n h_{0,j}(x_{0,j}). \end{cases} \quad (7)$$

The same result can be drawn from the factor graph too. $p_i(x_{0,i}, \mathbf{x}_i)$, $i = 1, 2, \dots, n$, describes a local probability about an SM $M_{0,i}$ and sensors $S_{i,j}$, $j = 0, 1, \dots, m$ from the SM in Fig. 1. Note that the local probability $p_i(x_{0,i}, \mathbf{x}_i) = 0$ when the SM's state $x_{0,i}$ and its sensors' state \mathbf{x}_i do not satisfy the constraint (3). Similarly, the local constraint function $v_0(\mathbf{x}_0)$ is about the states of all SM.

Therefore, it follows from the local Markov property of the investigated system, we have the real-world probability

$$\mathcal{P}(X = \mathbf{X}) = v_0(\mathbf{x}_0) \prod_{i=1}^n p_i(x_{0,i}, \mathbf{x}_i), \quad \mathbf{X} \in V. \quad (8)$$

From assumptions of the measurement errors, elements of $U = (u_{i,j})$ are independent, then the pdf of U is

$$q(\mathbf{U}) = \prod_{i,j} q_{i,j}(u_{i,j}), \quad \mathbf{U} = (u_{i,j}) \in R^{(n+1) \times (m+1)}, \quad (9)$$

where $q_{i,j}(u)$ is the pdf of $u_{i,j}$. From assumptions, $q_{0,i}$ is a Gaussian pdf with known variance and its mean is 0. According to the measurement model (1), the joint pdf of state variables \mathbf{X} and measurement values \mathbf{Z} are

$$\mathcal{P}(X = \mathbf{X}, Z = \mathbf{Z}) = q(\mathbf{Z} - \mathbf{H}(\mathbf{X})) \mathcal{P}(X = \mathbf{X}). \quad (10)$$

For a given observation value $\mathbf{Z} = (z_{i,j})$ of measurement values, it follows from (8-10) that the posterior distribution $\mathcal{P}(X = \mathbf{X} | Z = \mathbf{Z})$ is proportional to

$$f(\mathbf{X}) = v_0(\mathbf{x}_0) \prod_{i=1}^n p_i(x_{0,i}, \mathbf{x}_i) \prod_{i,j} q_{i,j}(z_{i,j} - h_{i,j}x_{i,j}), \quad (11)$$

where $h_{i,j} = h_{i,j}(x_{i,j})$ is the value of measurement function for this state. As the traditional statistical method, the observation value \mathbf{Z} does not appear as an argument variable of $f(\mathbf{X})$. We define a function g_0 of variables connected with the constraint node C_0 in Fig. 3,

$$g_0(\mathbf{x}_0) = v_0(\mathbf{x}_0) \prod_{j=0}^n q_{0,j}(z_{0,j} - h_{0,j}x_{0,j}), \quad (12)$$

and function g_i , $i = 1, 2, \dots, n$, for variables connected with the node C_i ,

$$g_i(x_{0,i}, \mathbf{x}_i) = p_i(x_{0,i}, \mathbf{x}_i) \prod_{j=0}^m q_{i,j}(z_{i,j} - h_{i,j}x_{i,j}). \quad (13)$$

Then $f(\mathbf{X})$ has a concise form as follows.

$$f(\mathbf{X}) = g_0(\mathbf{x}_0) \prod_{i=1}^n g_i(x_{0,i}, \mathbf{x}_i). \quad (14)$$

The sequence MAP estimator \mathbf{X}^* of the state variables vector for the given observation value \mathbf{Z} satisfies

$$f(\mathbf{X}^*) = \max_{\mathbf{X} \in V} f(\mathbf{X}). \quad (15)$$

Although V is finite, to solve (15) with a brute force method is an intractable problem because $|V|$ is very large and there are constraints (2), (3) and factors $g_i(x_{0,i}, \mathbf{x}_i)$ of f in the sequel. To overcome the difficulty, the BP algorithm studies a posteriori probability of a state variable $x_{i,j}$ of an SM or a sensor with the index i, j , with respect to the given observation value \mathbf{Z} of measurement values [35]. The BP algorithm presents the following bit-wise MAP estimator $\hat{x}_{i,j}$ for the state variable. For convenience, multiplying the posteriori probability with a constant factor $|V|^{-1}$, where $|V|$ is the number of elements in V , we have that the bit-wise MAP estimator $\hat{x}_{i,j}$ satisfies

$$\bar{f}(\hat{x}_{i,j}) = \max_{\alpha \in V_{i,j}} \bar{f}(\alpha, i, j), \quad (16)$$

$$\bar{f}(\alpha, i, j) = \frac{1}{|V|} \sum_{\mathbf{X} \in V(\alpha, i, j)} f(\mathbf{X}), \quad \alpha \in V_{i,j}, \quad (17)$$

where the set $V_{i,j}$ is defined at the beginning of this subsection, and $V(\alpha, i, j) \subset V$ consists of all $\mathbf{X} = (x_{i,j}) \in V$ whose i, j element $x_{i,j} = \alpha$.

B. BP ALGORITHMS AND THE FLOODING SCHEDULE

Not all elements of the state variable matrix \mathbf{X} are independent under the real-world probability \mathcal{P} . To overcome this and to apply the BP algorithm easily to the tree given in Fig. 3, we will consider a reference probability \mathcal{P}^\dagger . We follow the terminology of a reference probability developed by Elliot [32], and its recent applications include [33], [34]. B shows details of the technique.

We assume that the random matrix \mathbf{X} is drawn uniformly from V under the reference probability \mathcal{P}^\dagger , i.e., $\mathcal{P}^\dagger(\mathbf{X} = \mathbf{X}) = 1/|V|$ for every $\mathbf{X} \in V$. Hence all elements of \mathbf{X} are independent under \mathcal{P}^\dagger . For every random variable $\xi = \xi(\mathbf{X})$, the expectation of ξ under \mathcal{P}^\dagger is

$$E^\dagger(\xi) = \sum_{\mathbf{X} \in V} \xi(\mathbf{X})\mathcal{P}^\dagger(\mathbf{X} = \mathbf{X}) = \frac{1}{|V|} \sum_{\mathbf{X} \in V} \xi(\mathbf{X}). \quad (18)$$

For $\alpha \in V_{i,j}$ and $\mathbf{X} = (x_{i,j})$, the index function $I_{\alpha,i,j}$ is

$$I_{\alpha,i,j}(\mathbf{X}) = I_{\alpha,i,j}(x_{i,j}) = \begin{cases} 1, & x_{i,j} = \alpha, \\ 0, & x_{i,j} \neq \alpha. \end{cases} \quad (19)$$

Then the bit-wise posterior distribution $\bar{f}(\alpha, i, j)$ in (17) can be expressed by the reference expectation E^\dagger .

Concretely, for $i = 1, 2, \dots, n$, we can calculate the bit-wise posterior distribution $\bar{f}(\alpha, i, j)$ that the state of a sensor $S_{i,j}$ is α as follows.

$$\begin{aligned} \bar{f}(\alpha, i, j) &= E^\dagger \left(I_{\alpha,i,j}(x_{i,j})g_0(\mathbf{x}_0) \prod_{k=1}^n g_k(x_{0,k}, \mathbf{x}_k) \right) \\ &= E^\dagger \left(g_i I_{\alpha,i,j} E^\dagger \left(\prod_{k=0, \dots, n, k \neq i} g_k \middle| x_{0,i}, \mathbf{x}_i \right) \right) \\ &= E^\dagger \left(g_i I_{\alpha,i,j} E^\dagger \left(\prod_{k=0, \dots, n, k \neq i} g_k \middle| x_{0,i} \right) \right). \end{aligned} \quad (20)$$

Here we omit the arguments of $I_{\alpha,i,j}$ and g_i in the last two expressions. The second last equality follows from the tower property of conditional expectations [36] and the last equality follows from the facts that all elements of \mathbf{X} are independent under \mathcal{P}^\dagger and x_i is not among arguments of every g_k as $k \neq i$. Moreover, according to the terminology of the BP algorithm [16], we define the message from the node C_0 to an SM $M_{0,i}$ as

$$\mathcal{M}_{C_0 \rightarrow M_{0,i}} = E^\dagger \left(\prod_{k=0, \dots, n, k \neq i} g_k \middle| x_{0,i} \right). \quad (21)$$

It follows from the relations between g_k in Fig. 3 that

$$\mathcal{M}_{C_0 \rightarrow M_{0,i}} = E^\dagger \left(g_0 \prod_{k=1, \dots, n, k \neq i} \mathcal{M}_{M_{0,k} \rightarrow C_0} \middle| x_{0,i} \right), \quad (22)$$

where the message from the SM node $M_{0,k}$ to the constraint node C_0 is

$$\mathcal{M}_{M_{0,k} \rightarrow C_0} = E^\dagger (g_k(x_{0,k}, \mathbf{x}_k) | x_{0,k}) \quad (23)$$

A general form of (22) is proved in A.

The bit-wise posterior distribution $\bar{f}(\alpha, 0, i)$ that the state of an SM $M_{0,i}$ is α can be obtained similarly.

$$\bar{f}(\alpha, 0, i) = E^\dagger \left(I_{\alpha,0,i}(x_{0,i})g_0(\mathbf{x}_0) \prod_{k=1}^n \mathcal{M}_{M_{0,k} \rightarrow C_0}(x_{0,k}) \right). \quad (24)$$

Messages $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ and $\mathcal{M}_{M_{0,i} \rightarrow C_0}$ are generalizations of an issue in the communication theory [26] where every state variable has two possible states. In our problem, because a state variable often has many possible states, a message is a random variable. Although the form of involved messages is different, the flooding schedule of the BP algorithm [37] can still be employed for our problem. According to the flooding schedule, all SM pass extrinsic information $\mathcal{M}_{M_{0,i} \rightarrow C_0}$ up to the constraint node C_0 ; the constraint node C_0 then processes its inputs and passes new information $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ down to its neighboring SM. The tree in Fig. 3 is suited for the flooding schedule.

C. PSEUDO-CODE OF THE BP ALGORITHM

We can use (23), (22) in turn to calculate conditional expectations, use the last equality in (20) and (24) to obtain the bit-wise posterior distribution, and apply (16) to obtain the bit-wise MAP estimator $\hat{x}_{i,j}$ for each state variable. The conditional expectation for a random variable $\xi = \xi(\mathbf{X})$

$$E^\dagger(\xi | x_{0,i} = \alpha) = \frac{E^\dagger(\xi I_{\alpha,0,i})}{E^\dagger(I_{\alpha,0,i})} \quad (25)$$

is a ratio of two expectations. Because the random variable $x_{0,j}$ is drawn from V uniformly under the probability \mathcal{P}^\dagger ,

$$E^\dagger(I_{\alpha,0,i}) = \mathcal{P}^\dagger(x_{0,i} = \alpha) = |V_{0,i}|^{-1} \quad (26)$$

is a constant for every $\alpha \in V_{0,i}$. Multiplying a constant to a posteriori probability does not change an MAP estimator. Therefore, we replace hereafter conditional expectations (22), (23) of messages by their corresponding expectations. And we do not change the notation because introducing two many new symbols eventually becomes confusing.

The pseudo-code of the BP algorithm to calculate the bit-wise MAP estimators is as follows.

- 1: Each SM $M_{0,i}$, $i = 1, 2, \dots, n$ calculates its message function $\mathcal{M}_{M_{0,i} \rightarrow C_0}$ for every state α in its state table $V_{0,i}$ as follows.

$$\mathcal{M}_{M_{0,i} \rightarrow C_0}(\alpha) = E^\dagger(I_{\alpha,0,i}(x_{0,i})g_i(x_{0,i}, \mathbf{x}_i)). \quad (27)$$

And then each SM $M_{0,i}$, $i = 1, 2, \dots, n$ sends the message $\mathcal{M}_{M_{0,i} \rightarrow C_0}$ up to the constraint node C_0 .

- 2: The constraint node C_0 computes the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ for each SM $M_{0,i}$, $i = 1, 2, \dots, n$, and sends

TABLE 1. State tables and measurement functions of 5 sensors.

State	1	2	3	4
$h_{1,0}$	0	0.1	0.2	-
$h_{1,1}$	0	0.1	0.2	-
$h_{1,2}$	0	0.2	0.4	0.6
$h_{1,3}$	0	0.4	-	-
$h_{1,4}$	0	0.3	-	-

it down to $M_{0,i}$. For every state $\alpha \in V_{0,i}$, $\mathcal{M}_{C_0 \rightarrow M_{0,i}}(\alpha)$ is

$$E^\dagger \left(I_{\alpha,0,i}(x_{0,i}) g_0(\mathbf{x}_0) \prod_{k=1,\dots,n,k \neq i} \mathcal{M}_{M_{0,k} \rightarrow C_0}(x_{0,k}) \right). \quad (28)$$

- 3: Each SM $M_{0,i}$, $i = 1, 2, \dots, n$ sends the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ to its all sensors $S_{i,j}$. Each sensor computes

$$\bar{f}(\alpha, i, j) = E^\dagger \left(I_{\alpha,i,j}(x_{i,j}) g_i(\mathbf{x}_i) \mathcal{M}_{C_0 \rightarrow M_{0,i}}(x_{0,i}) \right) \quad (29)$$

for every state α in its state table $V_{i,j}$, and then computes its bit-wise MAP estimator $\hat{x}_{i,j}$ according to (16).

- 4: The root SM M_0 computes $\bar{f}(\alpha, 0, j)$, $j = 0, \dots, n$ for itself and other SM according to (24), and then computes their bit-wise MAP estimators from (16).

A constraint node is not a physical object, and the root SM takes its duty in our implementation, and Fig. 3 illustrates the flooding schedule of messages too. Every SM except the root executes Step 1 in parallel, and after the root sends its message, all SM executes Steps 3 or 4 in parallel again. In our implementation, an SM takes the duty of computation for all its sensors and a sensor sends only its measurement values $z_{i,j}$ up to its SM.

D. DISCRETE AND CONTINUOUS STATE TABLES

We assume hereafter that state variables $x_{i,j}$ of all sensors are independent under the real-world probability \mathcal{P} . We will illustrate by a toy system the relationship between a continuous state table and a discrete one for a summation variable. The system has an SM $M_{0,1}$ and its 5 sensors whose states and the corresponding values of measurement functions are in Table 1. Then from (3) the state table $V_{0,1}$ of the SM $M_{0,1}$ has 18 possible states from 0 to 1.7. As a simple expectation, we have shown in (26) that $E^\dagger(I_{\alpha,0,1}) = 1/18$ for every $\alpha \in V_{0,1}$. Because numbers of all possible values are not large, we can explore every state and obtain exact expectations.

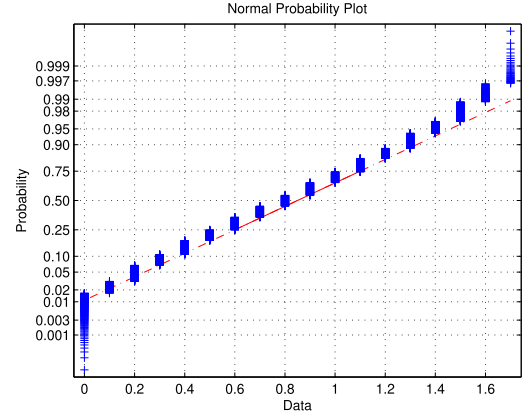
For example, we calculate, for a state $\alpha \in V_{0,1}$, a simple and typical expectation value $E^\dagger(I_{\alpha,0,1} p_1(x_{0,1}, \mathbf{x}_1))$, where $p_1(x_{0,1}, \mathbf{x}_1)$ is defined by (5) with the real-world probability \mathcal{P} and it is a factor in g_1 , see (13). From (18) we have that

$$|V| E^\dagger(I_{\alpha,0,1} p_1(x_{0,1}, \mathbf{x}_1)) = \sum_{\mathbf{x} \in V} p_1(x_{0,1}, \mathbf{x}_1) = \mathcal{P}(A), \quad (30)$$

where the random event A is

$$A = \{\mathbf{x}_1 : h_{1,0}(x_{1,0}) + \dots + h_{1,m}(x_{1,m}) = \alpha\}. \quad (31)$$

Other complex expectations in the BP algorithm can turn to similar issues with the real-world probability \mathcal{P} .

**FIGURE 4. The normal probability plot of a sum $h_{1,0} + \dots + h_{1,4}$.**

In general, when $x_{1,0}, \dots, x_{1,m}$ are independent random variables under the real-world probability, the random variables $h_{1,0}, \dots, h_{1,m}$ are independent too. Then $\mathcal{P}(A)$ is the probability of the sum $h_{1,0} + \dots + h_{1,m}$ is α . The sum tends to a Gaussian random variable for a large m . Fig. 4 shows the normal probability plot of the sum $h_{1,0} + \dots + h_{1,4}$ for the data in Table 1 with such a real-world probability, and the distribution of sum approximates a Gaussian distribution well in this case.

Based on the above idea, we can change some discrete random variable with its continuous alternative and do not influence the mechanism of the BP algorithm. In fact, to apply the BP algorithm in Section IV-C, we just need calculate expectations with a general form

$$E^\dagger \left(I_{\alpha,i,j} g_i \prod_k \mathcal{M}_k \right), \quad (32)$$

where involved random variables are of variable nodes connected by one constraint node in Fig. 3. The details to calculate such an expectation are the topics of B.

V. SIMULATION RESULTS AND DISCUSSION

In this Section, we will test the performance of the BP algorithm by simulations and discuss its performance in some extreme cases.

The proposed BP algorithm is concluded as the following pseudo-code.

- 1: Each SM $M_{0,i}$ calculates parameters $\hat{m}_{0,i}$, $\hat{\sigma}_{0,i}^2$ from (45),(50)-(52) for its message $\mathcal{M}_{M_{0,i} \rightarrow C_0}$ and sends the parameters up to the constraint node C_0 .
- 2: From (55), the node C_0 computes parameters $\hat{m}_{0,i}$, $\hat{\sigma}_{0,i}^2$ of the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ for each SM $M_{0,i}$, $i = 1, 2, \dots, n$, and sends it down to $M_{0,i}$.
- 3: Each SM $M_{0,i}$, $i \neq 0$ sends the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ to its all sensors $S_{i,j}$. Each sensor computes $\bar{f}(\alpha, i, j)$, $\alpha \in V_{i,j}$ and calculates its bit-wise MAP estimator $\hat{x}_{i,j}$ from (15).
- 4: The root SM M_0 computes bit-wise MAP estimators $\hat{x}_{0,i} = \hat{\mu}_i$, $i = 0, 1, \dots, n$, for itself and other SMs.

TABLE 2. Real-world probability of different state for 5 sensors.

State	1	2	3	4
$x_{1,0}$	0.5	0.25	0.25	-
$x_{1,1}$	0.5	0.25	0.25	-
$x_{1,2}$	0.25	0.25	0.25	0.25
$x_{1,3}$	0.5	0.5	-	-
$x_{1,4}$	0.5	0.5	-	-

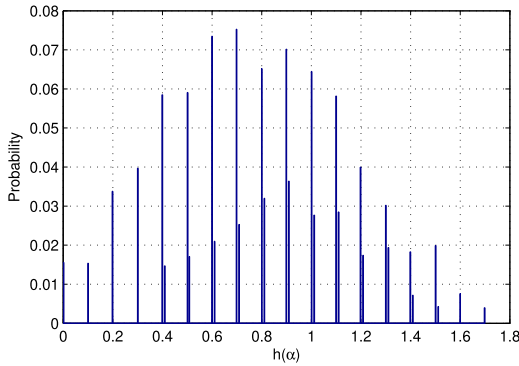


FIGURE 5. The approximate formula of messages may perform poorly.

A. ACCURACY OF APPROXIMATE MESSAGES

The number of an SM’s sensors affects the accuracy of the SM’s message up to C_0 , and the simulation shown by Fig. 4 indicates that the performance of an SM with 5 sensors is good enough.

Messages up to C_0 , messages down to SM, and all kinds of posterior distribution are probability values for various events. When we replace such a value with a pdf at a value $h_{i,j}(\alpha)$, we miss a factor that describes the neighborhood of $h_{i,j}(\alpha)$. For instance, if $h_{0,1}(\alpha)$ does not arrange evenly for every $\alpha \in V_{0,1}$, the accuracy of the approximate formula for $\mathcal{M}_{C_0 \rightarrow M_{0,1}}(\alpha)$ is questionable. Based on the method given in Section IV-D and the data in Table 1, replacing $h_{1,2}(3)$ by 0.41, and with the real-world probability in Table 2, we calculate the exact histogram of $\mathcal{M}_{C_0 \rightarrow M_{0,1}}(\alpha)$ and present it in Fig. 5. The random variable is far away from a Gaussian variable. This anomalous histogram is due to the increment 0.01 of $h_{1,2}(3)$, which is a clear signature of $x_{1,2} = 3$ in the current value at the SM $M_{0,1}$. Tests show the good performance of the approximate method when $h_{0,1}(\alpha)$ arranges evenly.

B. INDEPENDENCE OF SENSORS

The theoretical formula of B is based on the assumption of sensors’ independence, i.e., state variables $x_{i,j}$ of all sensors are independent under the real-world probability \mathcal{P} . To compare performances of methods without the assumptions, we study an environment has only one SM, i.e., there is only one level in the radial distribution system given in Fig. 3, and the SM’s sensors are dependent.

A modification of the proposed method is based on the following discussion on the algorithm’s mechanism. In Step 1, each SM $M_{0,j}$ collects currents at its all sensors and creates a

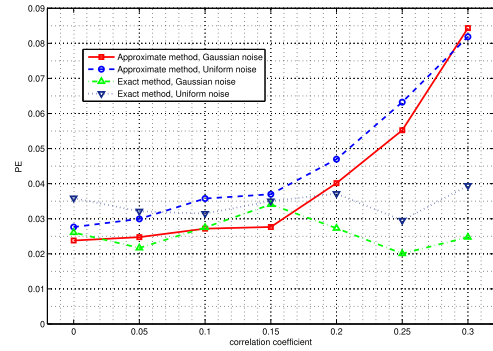


FIGURE 6. The performance of the exact and approximate methods without independence under different kinds of noise.

new virtual measurement $\mathcal{M}_{M_{0,j} \rightarrow C_0}$ of its current $h_{0,j}(x_{0,j})$. In Step 2, the node C_0 computes the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ for an SM $M_{0,i}$, from all SMs’ measurements z_0 and virtual measurements of SMs except $M_{0,i}$ itself. Then in Step 3, the SM uses the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ and data from its sensor to calculate bit-wise posterior distributions and MAP estimators. When the radial distribution system has only one level, we have that the message down to $C_{0,i}$ is its measurement and the variance of its noise, i.e., parameters $\hat{m}_{0,i} = z_{0,i}$, $\hat{\sigma}_{0,i}^2 = \sigma_U^2$. Then for the data in Table 1, where the SM $M_{0,1}$ has 18 possible states and every sensor has 3 or 4 states, both exact and approximate methods can apply.

Because the measurement function $h_{i,j}(x_{i,j})$ is a one-to-one function of the state variable $x_{i,j}$, in the test, for sensors states in Table 1, we replace $h_{1,0}$ with

$$\tilde{h}_{1,0} = [(h_{1,0} + kh_{1,1})/(1 + k)], \tag{33}$$

and then the independence between two sensors $S_{1,0}$ and $S_{1,1}$ breaks down. Here $[h]$ is the nearest integer less than or equal to h , and we choose a suitable k to ensure a preset correlation coefficient ρ between $\tilde{h}_{1,0}$ and $h_{1,1}$.

We apply exact and approximate methods to the issue with two kind of sensors’ noise u_1 . The first is Gaussian noise $N(0, 2.5 \times 10^{-3})$, the second is a uniform distribution on $[-0.087, 0.087]$. Two kinds of noise has the same variance. The noise $u_{0,1}$ of the SM has a Gaussian distribution $N(0, 10^{-4})$, corresponding to be of 0.5 accuracy class. We run two methods 400 times at every given parameters. Fig. 6 shows the percentage of error (PE) $\hat{x}_{1,0} \neq x_{1,0}$ for different methods and noise. The approximate method performs still well when the correlation coefficient $\rho \leq 0.2$, and the influence of different kinds of sensor’s noise is similar. Therefore, we recommend the exact method only when it can execute because it does not need the assumption of independence and its performance is robust.

C. WRONG CONFIGURATION OF A PRIOR PROBABILITY

From a perspective of Bayesian analysis, the real-world probability is a prior probability. The probability is preset in our issue, and it can be obtained from the historical data, or learned by a stochastic model [20], [39]. In this simulation,

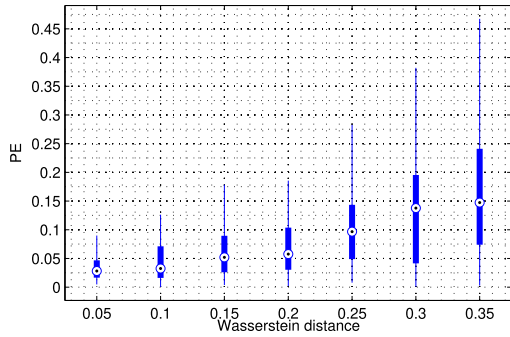


FIGURE 7. The performance of the approximate method with a drifting prior probability.

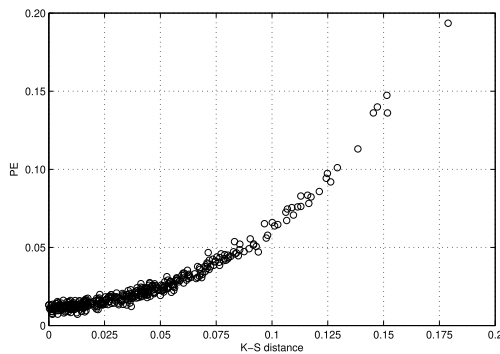


FIGURE 8. PE vs Kolmogorov-Smirnov distance for an SM.

4 SM are connected to the root SM and each SM has 5 sensors whose features are in Tables 1 and 2. Based on the approximate algorithm, We calculate the mean PE of all sensors of 400 simulations with different prior probabilities of sensors's state table. Fig. 7 shows the boxplot for the PE of state variables vs the Wasserstein distance [40] between the prior probabilities and the real distributions. The relation between the PE and the Wasserstein distance is similar to the relation between the PE and another distance of distributions, the Kullback-Leibler divergence. The result can be accepted when the Wasserstein distance between the prior probability and the true one is less than 0.1.

Note that the size of an SM's state table is not less than 17, and the size of the root SM is 65. As the true value for SMs is almost a continuous random variable, we can use the Kolmogorov-Smirnov distance to describe the difference between the prior distribution and the real one. Fig. 8 is the scatter plot of the PE of current vs the Kolmogorov-Smirnov distance between the distributions and the real distributions of the first SM. Here a point in Fig. 8 represents a simulation, and in a simulation, we run 100 tests for a pair of prior probability and real distributions of all sensors. As the current at the SM is the sum of its sensors and we cannot control it completely in the simulation, the difference between the prior probability and real distributions is not arranged uniformly as Fig. 7. Nevertheless, the relation between the PE and the Kolmogorov-Smirnov distance is clear in Fig. 8.

VI. RESULT OF THE EXPERIMENTAL SILM SYSTEM

The proposed state estimation method undertakes the main supervision and measurement in the experimental SILM system. However, state estimation is an intermediate step in a full solution of the SILM system which has routines to set an initial state table and to recover failures. For the experimental SILM system, every a hour, there is a successive sample for two minutes of every sensor and SM. From the sample, we can set a suitable state table for most of the sensors by a disaggregation algorithm [31]. Moreover, once our algorithm detects an extreme state for a variable, that is, a very high or a very low current value for a sensor, the algorithm calls a routine to rearrange the state table for the variable with a new successive sample. The routine can identify some high power electrical appliances such as electric kettles and heaters, and in most cases, it can present a suitable state table for lamps, computers, air-conditioners and projectors.

Besides the proposed method, we tested three other methods for the experimental SILM system. The first one is based on a linear state estimation with constraints (2), (3) which gives means of current for all sensors and SM, then the benchmark searches a lookup table to make identification to electronic appliances. A parameter of the benchmark is the lock time during which the system collects data every 15 seconds for computing those means. The second method is an HMM based on [13], [14] which has better performance and has less time-consuming than other learning-based techniques [29]. The third method, a full-discrete BP algorithm, sets a very large state table for SM in the buildings and deals with discrete random variables for SM. Because of the time and space complexities, in applying the third method, all data are uploaded to the server and the server runs the computation. The third method and our proposed method do not need the parameter of lock time, which can present results dynamically. We just presented their results recorded at $T = 3$ minutes for comparison.

The test is based on data from sensors and SM of the experimental SILM system and corresponding survey results of manual patrol inspection in a set of high-density school buildings lasting one week (five working days). According to manual inspection, we can determine whether a method detects a working device or not. Besides the above data, there is another set of data from three rooms, where every sensor and an alternative SM are connected in series for the test. From this set of data, we can obtain current of a device correctly and hence we can calculate R^2 of its current, where

$$R^2 = 1 - \frac{\sum_i |\hat{I}_i - I_i|^2}{\sum_i |I_i|^2}. \quad (34)$$

Here I_i is measured by an SM, and \hat{I}_i is reported by a method. When a method gives the estimator $\hat{x}_{k,j}$ of the state of a sensor or SM indicating by k, j , the corresponding current \hat{I} is determined by the measurement function, i.e., $\hat{I} = h_{k,j}(\hat{x}_{k,j})$. A shortcoming of R^2 is that two states with similar true current values are hard to distinguish by it.

TABLE 3. The percentage of error (PE) for methods' detecting a device. T is the lock time. A number less than 0.01 is marked as 0 in the table.

Method	lamp	D.C.	projector
Linear Est., $T = 3min$	0.32	0.02	0.08
Linear Est., $T = 10min$	0.26	0.	0.02
Linear Est., $T = 50min$	0.18	0.	0.02
HMM, $T = 10min$	0.12	0.01	0.05
HMM, $T = 50min$	0.03	0.	0.01
Full-discrete BP	0.	0.	0.
Proposed BP	0.	0.	0.

Method	k&h	laptop	charger
Linear Est., $T = 3min$	0.07	0.88	0.98
Linear Est., $T = 10min$	0.15	0.82	0.98
Linear Est., $T = 50min$	0.23	0.85	0.99
HMM, $T = 10min$	0.09	0.08	0.21
HMM, $T = 50min$	0.04	0.04	0.10
Full-discrete BP	0.04	0.04	0.11
Proposed BP	0.04	0.04	0.08

TABLE 4. R^2 of different kinds of device's current. Symbols and abbreviations are the same as Table 3.

Method	lamp	D.C.	laptop	charger
Linear Est., $T = 3$	0.96	0.90	0.87	0.67
Linear Est., $T = 10$	0.97	0.93	0.89	0.71
Linear Est., $T = 50$	0.98	0.96	0.92	0.79
HMM, $T = 10$	0.98	0.94	0.89	0.74
HMM, $T = 50$	0.98	0.93	0.87	0.74
Full-discrete BP	0.98	0.93	0.85	0.70
Proposed BP	0.98	0.93	0.88	0.75

Table 3 presents the PE of detecting a device for different methods with respect to manual patrol inspections. As PE of air-conditioner is 0.00 for all methods, its data are not presented in Table 3. It seems that the proposed method can obtain similar results as the full-discrete BP algorithm, and both of them have better performances than other methods. The performance of HMM is good just for a long lock time. Low PE of linear state estimation for lamps may be due to different kinds of lamps among which some have similar powers. Low PE of electric kettles and heaters may be due to faults in inspections, because there are many reports of working heaters by algorithms and no working heaters in inspections. There are some breaks in patrols, whose details are not reported.

Table 4 presents R^2 of different devices for the second set of data. There is no electric kettle or heater in the testing rooms and hence there is no corresponding column in the table. Table 4 reveals an interesting result such that the linear state estimation has a better R^2 for lamps, laptop computers and chargers than other methods. The cause of the phenomenon is that states of a device have similar true current values and hence they are hard to distinguish by R^2 . Hence we suggest the linear state estimation to calculate the mean of power for a device. As an index to distinguish devices with varying power, PE is better than R^2 . Table 3 shows that the proposed method and full-discrete method can report devices' working state more accurately.

To investigate the time and space complexities of the four methods, we use a computer with an Intel(R) Core(TM) i7

TABLE 5. Time and space complexities of four methods.

Method	Time (sec.)	Memory (Mbyte)
Linear Est.	0.732	6
HMM, $T = 10$	8.35	9
HMM, $T = 50$	22.7	12
Full-discrete BP	235	250
Proposed BP	0.914	16

CPU to store and compute data from all SM and sensors. The execution time in seconds of four methods and maximal memory in megabytes used by four methods are presented in Table 5. For different lock times, the execution time of linear state estimation is similar, and only data of $T = 3$ minutes are presented in Table 5. The proposed method and the linear state estimation have the least execution time and the full-discrete BP needs a very long execution time and a large memory.

Besides facts presented by Tables 3 to 5, it is worthwhile to note that a large part of the work of the proposed method can run on concentrators. With the proposed method, data from a sensor are collected and processed by a concentrator located in the same room as the sensor. For other methods, the process can not be implemented by concentrators because the linear state estimation and HMM must analyze all data together, and the complexities of a full-discrete BP can not be treated by an SoC on a concentrator. Because of this feature, the proposed method is a technique of SILM, and the other three methods are ILM techniques as sensors' data leak outside a room by other methods.

VII. CONCLUSION

We proposed an approximate BP algorithm for a state estimator of SM and sensors of an experimental SILM system in a radial distribution system. Although every sensor in the SILM system has a few states, the number of states in the state table of some SM in the system is enormous, and hence the complexities of a full-discrete BP algorithm are unacceptable. In this paper, we establish a new BP algorithm for such a system. The algorithm collects discrete states of sensors of an SM and creates a virtual continuous measurement for the SM. And a discrete MAP estimators for a sensor is derived from its measurement value and received continuous messages.

Simulation results indicate that the proposed BP algorithm is acceptable and can apply to some environments where some theoretical assumptions are slightly wrong. Those theoretical assumptions include sensors' independence and a correct prior probability. Tests of an experimental SILM system in a set of high-density school buildings within a secondary distribution network show that the proposed method detects and identifies a working device well. The results show that the proposed method achieves a percentage of error (8%), which outperforms the percentage achieved by the other state-of-the-art methods, i.e., a linear state estimation of 99%, a hidden Markov model of 21%, and a full-discrete BP algorithm of 11%. In addition, the time and space complexities of the

proposed method and the linear state estimation are the least among all methods, and the proposed algorithm is the only one which can run by SoC on concentrators.

**APPENDIX A
A GENERAL FORM OF (22)**

Lemma 1: For a subset W of $\{1, 2, \dots, n\}$ and $i = 0, 1, 2, \dots, n$, we have

$$E^\dagger \left(g_0 \prod_{k \in W} g_k \middle| x_{0,i} \right) = E^\dagger \left(g_0 \prod_{k \in W} \mathcal{M}_{M_{0,k} \rightarrow C_0} \middle| x_{0,i} \right). \quad (35)$$

Proof. It follows from the tower property of conditional expectations that

$$E^\dagger \left(g_0 \prod_{k \in W} g_k \middle| x_{0,i} \right) = E^\dagger \left(g_0 E^\dagger \left(\prod_{k \in W} g_k \middle| \mathbf{x}_0 \right) \middle| x_{0,i} \right). \quad (36)$$

We only need to prove that

$$E^\dagger \left(\prod_{k \in W} g_k \middle| \mathbf{x}_0 \right) = \prod_{k \in W} \mathcal{M}_{M_{0,k} \rightarrow C_0}. \quad (37)$$

The proof is by induction on the size of W . When $W = \{k\}$ has only one element k , because $\mathbf{x}_0 = (x_{0,0}, \dots, x_{0,n})$ and the argument $\mathbf{x}_k, x_{0,k}$ of g_k and random variables $x_{0,i}, i \neq k$ are independent under \mathcal{P}^\dagger , it is clear that

$$E^\dagger (g_k | \mathbf{x}_0) = E^\dagger (g_k | x_{0,k}) = \mathcal{M}_{M_{0,k} \rightarrow C_0}. \quad (38)$$

Suppose then that we have (37) for any subset W' with size less than l . For a subset W with size l , let $W' = W \setminus \{s\}$ where $s \in W$, and we have that

$$\begin{aligned} E^\dagger \left(\prod_{k \in W} g_k \middle| \mathbf{x}_0 \right) &= E^\dagger \left(E^\dagger \left(\prod_{k \in W} g_k \middle| \mathbf{x}_0, \mathbf{x}_k, k \in W' \right) \middle| \mathbf{x}_0 \right) \\ &= E^\dagger \left(E^\dagger (g_r | \mathbf{x}_0, \mathbf{x}_k, k \in W') \prod_{k \in W'} g_k \middle| \mathbf{x}_0 \right). \end{aligned} \quad (39)$$

The last equality follows from the fact that the set $\{\mathbf{x}_0, \mathbf{x}_r, r \in W'\}$ includes every argument of $g_k, k \in W'$. Moreover, because all elements of \mathbf{X} are independent under \mathcal{P}^\dagger and \mathbf{x}_i is not among arguments of every g_r as $i \neq r$.

$$E^\dagger (g_r | \mathbf{x}_0, \mathbf{x}_k, k \in W') = E^\dagger (g_r | x_{0,r}) = \mathcal{M}_{M_{0,r} \rightarrow C_0}. \quad (40)$$

Because $\mathcal{M}_{M_{0,r} \rightarrow C_0}$ is a function of the random variable $x_{0,r}$, we have that

$$\begin{aligned} E^\dagger \left(\prod_{k \in W} g_k \middle| \mathbf{x}_0 \right) &= \mathcal{M}_{M_{0,r} \rightarrow C_0} E^\dagger \left(\prod_{k \in W'} g_k \middle| \mathbf{x}_0 \right) \\ &= \prod_{k \in W} \mathcal{M}_{M_{0,k} \rightarrow C_0}, \end{aligned} \quad (41)$$

where the last equality follows from the induction hypothesis for the subset W' .

APPENDIX B

RANDON-NIKODYM DENSITY AND MESSAGE $\mathcal{M}_{M_{0,i} \rightarrow C_0}$

To establish an approximate method to calculate expectations, we use the tool of Randon-Nikodym density. For a probability \mathcal{P}' on V , consider the following random variable on V ,

$$\frac{d\mathcal{P}'}{d\mathcal{P}^\dagger}(\mathbf{X}) = |V| \mathcal{P}'(\mathbf{X} = \mathbf{X}), \mathbf{X} \in V. \quad (42)$$

It can be verified that for every $\mathbf{X} \in V$,

$$\mathcal{P}'(\mathbf{X} = \mathbf{X}) = \frac{d\mathcal{P}'}{d\mathcal{P}^\dagger}(\mathbf{X}) \mathcal{P}^\dagger(\mathbf{X} = \mathbf{X}), \quad (43)$$

and hence for every random variable $\xi = \xi(\mathbf{X})$ on V ,

$$E'(\xi) = \sum_{\mathbf{X} \in V} \xi(\mathbf{X}) \mathcal{P}'(\mathbf{X} = \mathbf{X}) = E^\dagger \left(\xi \frac{d\mathcal{P}'}{d\mathcal{P}^\dagger} \right), \quad (44)$$

where E' is the expectation under \mathcal{P}' . We refer to [32], [33], [34] for more details of the reference probability and its Randon-Nikodym densities.

Given $i \geq 1$, to calculate a message $\mathcal{M}_{M_{0,i} \rightarrow C_0}(\alpha)$ given in (27), let a probability \mathcal{P}' satisfies that random variables $x_{i,0}, \dots, x_{i,m}$ are independent under \mathcal{P}' , and that

$$\mathcal{P}'(x_{i,j} = x_{i,j}) = \frac{\mathcal{P}(x_{i,j} = x_{i,j}) q_{i,j}(z_{i,j} - h_{i,j} x_{i,j})}{K_{i,j}}, \quad (45)$$

where $x_{i,j}$ is any state in $V_{i,j}$, and the constant factor

$$K_{i,j} = \sum_{\alpha \in V_{i,j}} \mathcal{P}(x_{i,j} = \alpha) q_{i,j}(z_{i,j} - h_{i,j}(\alpha)). \quad (46)$$

See (9-12) for the meaning of $q_{i,j}, z_{i,j}$, and $h_{i,j}$. Then we can be verified

$$\frac{d\mathcal{P}'}{d\mathcal{P}^\dagger}(\mathbf{x}_i) = \mathcal{P}(\alpha_i = \mathbf{x}_i) \prod_j \left(q_{0,j}(z_{0,j} - h_{0,j}(x_{0,j})) \frac{|V_{j,k}|}{K_{j,k}} \right). \quad (47)$$

Therefore, for the given $i \geq 1$, from (5), (6), and (13) we have that for a random variable $\xi = \xi(x_{0,i}, \mathbf{x}_i)$,

$$E'(\xi v_i(x_{0,i}, \mathbf{x}_i)) = E^\dagger (\xi g_i(x_{0,i}, \mathbf{x}_i)), \quad (48)$$

where we omit a constant factor. For example, by letting ξ in (44) be $I_{\alpha,0,i}$, the message $\mathcal{M}_{M_{0,i} \rightarrow C_0}(\alpha)$ in (27) is

$$\begin{aligned} E^\dagger (I_{\alpha,0,i} g_i(x_{0,i}, \mathbf{x}_i)) &= E'(I_{\alpha,0,i} v_i(x_{0,i}, \mathbf{x}_i)) \\ &= \mathcal{P}'(h_{i,0}(x_{i,0}) + \dots + h_{i,m}(x_{i,m}) = \alpha). \end{aligned} \quad (49)$$

The sum $h_i = h_{i,0} + \dots + h_{i,m}$ tends to a Gaussian random variable under \mathcal{P}' , it has a mean $\acute{m}_{0,i}$ and variance $\acute{\sigma}_{0,i}^2$, where

$$\acute{m}_{0,i} = \sum_{j=1}^m \acute{m}_{i,j}, \acute{\sigma}_{0,i}^2 = \sum_{j=1}^m \acute{\sigma}_{i,j}^2, \quad (50)$$

and $\acute{m}_{i,j}, \acute{\sigma}_{i,j}^2$ are the mean and variance of $h_{i,j}$ under \mathcal{P}' .

$$\acute{m}_{i,j} = \sum_{x_{i,j} \in V_{i,j}} h_{i,j}(x_{i,j}) \mathcal{P}'(x_{i,j} = x_{i,j}), \quad (51)$$

$$\hat{\sigma}_{i,j}^2 = \sum_{x_{i,j} \in V_{i,j}} (h_{i,j}(x_{i,j}) - \hat{m}_{i,j})^2 \mathcal{P}'(x_{i,j} = x_{i,j}). \quad (52)$$

Then the probability that the sum h_i is α in (49) is proportional to the Gaussian pdf.

$$\mathcal{M}_{M_{0,i} \rightarrow C_0}(\alpha) \propto \exp\left(-\frac{(h_{0,i}(\alpha) - \hat{m}_{0,i})^2}{2\hat{\sigma}_{0,i}^2}\right). \quad (53)$$

The above (53) is the proposed approximate formula for the message sent up to the constraint node C_0 . When the number of random variables in the sum is small, the arrangement of the state table of the sum is irregular, or not all sensors are independent under the real-world probability \mathcal{P} , then the accuracy of the approximate formula may decrease.

**APPENDIX C
MESSAGE $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$**

From Section B, the message $\mathcal{M}_{M_{0,i} \rightarrow C_0}$ can be regarded as a Gaussian random variable with mean $\hat{m}_{0,i}$ and variance $\hat{\sigma}_{0,i}^2$. Because all SMs in the system are often of the same kind, we assume that every SM's measurement error $u_{0,i}$ is a Gaussian random variable with mean 0 and variance σ_U^2 . Then from the discussion in Appendix E we can obtain the following approximate formula of the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}(\alpha)$.

From Appendix E, the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}$ can be regard as a Gaussian pdf with mean $\hat{m}_{0,i}$ and variance $\hat{\sigma}_{0,i}^2$,

$$\mathcal{M}_{C_0 \rightarrow M_{0,i}}(\alpha) \propto \exp\left(-\frac{(h_{0,i}(\alpha) - \hat{m}_{0,i})^2}{2\hat{\sigma}_{0,i}^2}\right). \quad (54)$$

Here $\hat{m}_{0,i}$ and $\hat{\sigma}_{0,i}^2$ are

$$\hat{m}_{0,i} = d_i \left(y_i - \frac{\mu}{s^2 + \sigma_U^2} \right), \quad \hat{\sigma}_{0,i}^2 = \frac{s^2 \sigma_U^2}{s^2 + \sigma_U^2}, \quad (55)$$

where $\mu = \sum_{j=1}^n d_j y_j$, $s^2 = \sum_{j=1}^n d_j^2$, $d_i = \sigma_U^2$, $y_i = \frac{z_{0,i} + z_{0,0}}{\sigma_U^2}$ and for $j \neq i$, $d_j = \frac{\sigma_U^2 \hat{\sigma}_{0,j}^2}{\sigma_U^2 + \hat{\sigma}_{0,j}^2}$, $y_j = \frac{z_{0,j} + z_{0,0}}{\sigma_U^2} + \frac{\hat{m}_{0,j}}{\hat{\sigma}_{0,j}^2}$. Appendix VII discusses more details of (55).

**APPENDIX D
BIT-WISE POSTERIOR DISTRIBUTION AND MAP ESTIMATORS**

For every $j = 1, 2, \dots, n$, we write $\hat{d}_j = \frac{\sigma_U^2 \hat{\sigma}_{0,j}^2}{\sigma_U^2 + \hat{\sigma}_{0,j}^2}$, $\hat{y}_j = \frac{z_{0,j} + z_{0,0}}{\sigma_U^2} + \frac{\hat{m}_{0,j}}{\hat{\sigma}_{0,j}^2}$ and let $\hat{\mu} = \sum_{j=1}^n \hat{d}_j \hat{y}_j$, $\hat{s}^2 = \sum_{j=1}^n \hat{d}_j^2$. Appendix VII shows that for $i \neq 0$, the bit-wise posterior distribution

$$\bar{f}(\alpha, 0, i) \propto \exp\left(-\frac{(h_{0,i}(\alpha) - \hat{\mu}_i)^2}{2\hat{\sigma}_{i,i}^2}\right), \quad (56)$$

where the posterior mean and variance are

$$\hat{\sigma}_{i,i}^2 = \hat{d}_i - \frac{\hat{d}_i^2}{\sigma_U^2 + \hat{s}^2}, \quad \hat{\mu}_i = \hat{d}_i \left(\hat{y}_i - \frac{\hat{\mu}}{\sigma_U^2 + \hat{s}^2} \right). \quad (57)$$

When $i = 0$, the bit-wise approximate posterior distribution $\bar{f}(\alpha, 0, 0)$ follows (56) too, where

$$\hat{\sigma}_{0,0}^2 = \frac{\hat{s}^2 \sigma_U^2}{\sigma_U^2 + \hat{s}^2}, \quad \hat{\mu}_0 = \sum_{i=1}^n \hat{\mu}_i. \quad (58)$$

For $i \neq 1$, we have similarly that the approximate formulae of the bit-wise posterior distribution $\bar{f}(\alpha, i, j)$ is given in the following equation, with omitting a constant factor.

$$\bar{f}(\alpha, i, j) \propto \frac{\mathcal{P}(x_{i,j} = \alpha)}{\sqrt{\hat{\sigma}_{0,i}^2 - \sigma_{i,j}^2 + \hat{\sigma}_{0,i}^2}} \left(\exp(A) + \frac{B}{2} \right),$$

Where

$$A = -\frac{(\hat{m}_{0,i} - \hat{m}_{i,j} + h_{i,j}(\alpha))^2}{\hat{\sigma}_{0,i}^2 - \sigma_{i,j}^2}$$

and

$$B = \frac{(\hat{\sigma}_{0,i}^2 - \sigma_{i,j}^2) \hat{\sigma}_{0,i}^2}{\hat{\sigma}_{0,i}^2 - \sigma_{i,j}^2 + \hat{\sigma}_{0,i}^2} \left(\frac{\hat{m}_{0,i} - \hat{m}_{i,j} + h_{i,j}(\alpha)}{\hat{\sigma}_{0,i}^2 - \sigma_{i,j}^2} + \frac{\hat{m}_{0,i}}{\hat{\sigma}_{0,i}^2} \right)^2.$$

Because the pdf in (56) is Gaussian and the approximate state table is R , an approximate formula of the bit-wise MAP estimator defined in (16) for $x_{0,i}$ is $\hat{x}_{0,i} = \hat{\mu}_i$, $i = 0, 1, \dots, n$. For $i \neq 0$ and $j = 0, 1, \dots, m$, the bit-wise posterior distribution $\bar{f}(\alpha, i, j)$, $\alpha \in V_{i,j}$ has a complex expression. But because $|V_{i,j}|$ is small, we can calculate $\bar{f}(\alpha, i, j)$ for each α in $V_{i,j}$ to search the bit-wise MAP estimator $\hat{x}_{i,j}$.

**APPENDIX E
APPROXIMATE FORMULAE WITH THE NODE C_0**

It follows from (53) of Section B that $h_{0,i}(\alpha)$, instead of α , appears in a Gaussian pdf. As an example, we study first the following auxiliary function with an argument row vector $\mathbf{h} = (h_{0,1}, \dots, h_{0,n})$.

$$w(\mathbf{h}) = \prod_{j=0}^n q_{0,j}(z_{0,j} - h_{0,j} x_{0,j}) \prod_{i=1}^n \mathcal{M}_{M_{0,k} \rightarrow C_0}(x_{0,i}), \quad (59)$$

where $h_{0,0}$ and \mathbf{h} satisfy (3). Then the function $w(\mathbf{h})$, multiplying a constant factor, is a joint Gaussian pdf of \mathbf{h} . Concretely, let a row vector $\mathbf{y} = (y_1, \dots, y_n)$ and consider a symmetric matrices $\mathbf{A} = (a_{i,j}) \in R^{n \times n}$ such that

$$y_i = \frac{z_{0,i} + z_{0,0}}{\sigma_U^2} + \frac{\hat{m}_{0,i}}{\hat{\sigma}_{0,i}^2}, \quad (60)$$

$$a_{i,i} = \frac{2}{\sigma_U^2} + \frac{1}{\hat{\sigma}_{0,i}^2}, \quad (61)$$

$$a_{i,j} = \frac{1}{\sigma_U^2}, \quad i \neq j. \quad (62)$$

Then we have that

$$\begin{aligned} \log(w(\mathbf{h})) &= -\frac{1}{2} (\mathbf{h} \mathbf{A} \mathbf{h}^T - \mathbf{h} \mathbf{y}^T) + c \\ &= -\frac{1}{2} ((\mathbf{h} - \mu) \mathbf{A} (\mathbf{h} - \mu)^T) + c', \end{aligned} \quad (63)$$

where c and c' are two constant terms with the given observation value \mathbf{Z} , and the row vector $\mu = (\mu_1, \dots, \mu_n)$ satisfies $\mu = \mathbf{y}\mathbf{A}^{-1}$. Note that $\mathbf{A} = \mathbf{D} + \beta\beta^T$, where \mathbf{D} is a diagonal matrix, $\mathbf{D} = \text{diag}(\frac{1}{\sigma_U^2} + \frac{1}{\sigma_{0,1}^2}, \dots, \frac{1}{\sigma_U^2} + \frac{1}{\sigma_{0,n}^2})$, and the row vector $\beta = \frac{1}{\sigma_U}(1, 1, \dots, 1)$. It follows from the Sherman-Morrison formula [38] that

$$\mathbf{A}^{-1} = \mathbf{D}^{-1} - \frac{\mathbf{D}^{-1}\beta^T\beta\mathbf{D}^{-1}}{1 + \beta\mathbf{D}^{-1}\beta^T}, \quad (64)$$

and each element of \mathbf{A}^{-1} as well as μ can be obtained.

We define an auxiliary probability \mathcal{P}' such that

$$\frac{d\mathcal{P}'}{d\mathcal{P}^\dagger} \propto w(\mathbf{h}). \quad (65)$$

Then the row random vector \mathbf{h} is a Gaussian random vector with mean $\hat{\mathbf{m}}$ and covariance matrix \mathbf{A}^{-1} .

Using the technique of Randon-Nikodym densities in Section B, we can obtain from (24) that the bit-wise posterior distribution $\bar{f}(\alpha, 0, i)$, $i \neq 0$ is

$$\bar{f}(\alpha, 0, i) = \mathcal{P}'(h_{0,i} = h_{0,i}(\alpha)). \quad (66)$$

Under \mathcal{P}' , the random variable $h_{0,i}$ is Gaussian with mean μ_i , and the (i, i) element of \mathbf{A}^{-1} is its variance. The concrete expressions of the random variable's mean and variance, and the bit-wise posterior distribution $\bar{f}(\alpha, 0, i)$ can be calculated by (64) and they are given in (56) and (57).

It follows from (24) that

$$\bar{f}(\alpha, 0, 0) \propto \mathcal{P}'(h_{0,1} + \dots + h_{0,n} = h_{0,0}(\alpha)). \quad (67)$$

Because the sum $h = h_{0,1} + \dots + h_{0,n} = \mathbf{h}\boldsymbol{\gamma}^T$, where every element in the row vector $\boldsymbol{\gamma}$ is 1, the sum h is a Gaussian random variable under the probability \mathcal{P}' . Its mean is $\mu\boldsymbol{\gamma}^T$ and its variance is $\boldsymbol{\gamma}\mathbf{A}^{-1}\boldsymbol{\gamma}^T$, and the results are given in (58).

Now we turn to the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}(\alpha)$ from the constraint node C_0 down to an SM $M_{0,i}$. According to (28), we can define a new probability \mathcal{P}'' such that

$$\frac{d\mathcal{P}''}{d\mathcal{P}^\dagger} \propto \frac{w(\mathbf{h})}{\mathcal{M}_{M_{0,k} \rightarrow C_0}(x_{0,i})}. \quad (68)$$

Then we have that the message

$$\mathcal{M}_{C_0 \rightarrow M_{0,i}}(\alpha) \propto \mathcal{P}'(h_{0,i} = h_{0,i}(\alpha)), \quad (69)$$

and under the probability \mathcal{P}'' , the row random vector \mathbf{h} has a Gaussian pdf determined by \mathbf{A}_i and \mathbf{y}_i . Here

$$\mathbf{A}_i = \mathbf{A} - \frac{\dot{m}_{0,i}}{\sigma_{0,i}^2} \mathbf{e}_i \mathbf{e}_i^T, \quad \mathbf{y}_i = \mathbf{y} - \frac{\dot{m}_{0,i}}{\sigma_{0,i}^2} \mathbf{e}_i, \quad (70)$$

and \mathbf{e}_i is the i -th column of the identity matrix. The concrete expression of the message $\mathcal{M}_{C_0 \rightarrow M_{0,i}}(\alpha)$ is presented by (54).

REFERENCES

- [1] M. El-Hawary, "The smart grid—State-of-the-art and future trends," *Electr. Power Compon. Syst.*, vol. 42, nos. 3–4, pp. 239–250, 2014.
- [2] S. Chakraborty, "Smart meters for enhancing protection and monitoring functions in emerging distribution systems," *Int. J. Electr. Power Energy Syst.*, vol. 127, pp. 121–135, May 2021.
- [3] A. Langevin, G. Gagnon, and M. Cheriet, "Crosstalk suppression in semi-intrusive load monitoring systems using Hall effect sensors," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5019–5027, Nov. 2020.
- [4] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.
- [5] E. J. Aladesanmi and K. A. Folly, "Overview of non-intrusive load monitoring and identification techniques," *IFAC-PapersOnLine*, vol. 48, no. 30, pp. 415–420, 2015.
- [6] T. D. Huang, W.-S. Wang, and K.-L. Lian, "A new power signature for nonintrusive appliance load monitoring," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1994–1995, Jul. 2015.
- [7] N. Henao, K. Agbossou, S. Kelouwani, Y. Dubé, and M. Fournier, "Approach in nonintrusive type I load monitoring using subtractive clustering," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 812–821, Mar. 2017.
- [8] D. A. Haughton and G. T. Heydt, "A linear state estimation formulation for smart distribution systems," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1187–1195, May 2013.
- [9] C. Ma and L. Yin, "Deep flexible transmitter networks for non-intrusive load monitoring of power distribution networks," *IEEE Access*, vol. 9, pp. 107424–107436, 2021.
- [10] Z. Ling, Q. Tao, J. Zheng, P. Xiong, M. Liu, Z. Xiao, and W. Gang, "A nonintrusive load monitoring method for office buildings based on random forest," *Buildings*, vol. 11, no. 10, p. 449, Sep. 2021.
- [11] Z. D. Tekler, R. Low, Y. Zhou, C. Yuen, L. Blessing, and C. Spanos, "Near-real-time plug load identification using low-frequency power data in office spaces: Experiments and applications," *Appl. Energy*, vol. 275, Oct. 2020, Art. no. 115391.
- [12] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Effective non-intrusive load monitoring of buildings based on a novel multi-descriptor fusion with dimensionality reduction," *Appl. Energy*, vol. 279, Dec. 2020, Art. no. 115872.
- [13] Y. Sun, "Non-intrusive home-load identification based on improved hidden Markov model," *Electr. Power Construct.*, vol. 41, no. 4, pp. 73–80, 2020.
- [14] Q. Duan, J. Liu, and D. Zhao, "A fast algorithm for short term electric load forecasting by a hidden semi-Markov process," *J. Stat. Comput. Simul.*, vol. 89, no. 5, pp. 831–843, Mar. 2019.
- [15] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Mach. Learn.*, vol. 29, pp. 245–273, Nov. 1997.
- [16] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Comput. Intell. Mag.*, vol. 6, no. 3, pp. 36–46, Aug. 2011.
- [17] X. Wu, D. Jiao, and L. You, "Nonintrusive on-site load-monitoring method with self-adaption," *Int. J. Electr. Power Energy Syst.*, vol. 119, Jul. 2020, Art. no. 105934.
- [18] M. Cosovic and D. Vukobratovic, "Distributed Gauss–Newton method for state estimation using belief propagation," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 648–658, Jan. 2019.
- [19] F. Ni, P. H. Nguyen, J. F. G. Cobben, H. E. Van den Brom, and D. Zhao, "Three-phase state estimation in the medium-voltage network with aggregated smart meter data," *Int. J. Electr. Power Energy Syst.*, vol. 98, pp. 463–473, Jun. 2018.
- [20] W. Jiang, J. Chen, H. Tang, S. Cheng, Q. Hu, M. Cai, and S. Rahman, "A physical probabilistic network model for distribution network topology recognition using smart meter data," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6965–6973, Nov. 2019.
- [21] B. Jing and M. Chen, "Advanced nonintrusive load monitor for equipment diagnostics on fan coil unit," in *Proc. 41st Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2015, pp. 339–344.
- [22] M. V. Bavaresco, S. D'Oca, E. Ghisi, and R. Lamberts, "Technological innovations to assess and include the human dimension in the building-performance loop: A review," *Energy Buildings*, vol. 202, Nov. 2019, Art. no. 109365.
- [23] L. Pereira and N. J. Nunes, "Low cost framework for non-intrusive home energy monitoring and research," in *Proc. 1st Int. Conf. Smart Grids Green IT Syst. (SMARTGREENS)*, 2012, pp. 191–196.

- [24] F. Rossier, P. Lang, and J. Hennebert, "Near real-time appliance recognition using low frequency monitoring and active learning methods," in *Proc. 7th Int. Conf. Future Buildings Districts, Energy Efficiency Nano Urban Scale (CISBAT)*, 2017, pp. 691–696.
- [25] E. Tabanelli, D. Brunelli, A. Acquaviva, and L. Benini, "Trimming feature extraction and inference for MCU-based edge NILM: A systematic approach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 943–952, Feb. 2022.
- [26] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*, Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [27] T. Ritmeester and H. Meyer-Ortmanns, "State estimation of power flows for smart grids via belief propagation," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 102, no. 1, Jul. 2020, Art. no. 012311.
- [28] Y. Weng, R. Negi, and M. D. Ilic, "Probabilistic joint state estimation for operational planning," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 601–612, Jan. 2019.
- [29] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4910–4920, Nov. 2019.
- [30] A. Abur and A. Expósito, *Power System State Estimation: Theory and Implementation* (Power Engineering Series). New York, NY, USA: Taylor & Francis, 2004.
- [31] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, Jr., "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Trans. Power Del.*, vol. 10, no. 3, pp. 1200–1210, Jul. 1995.
- [32] R. J. Elliott and W. P. Malcolm, "General smoothing formulas for Markov-modulated Poisson observations," *IEEE Trans. Autom. Control*, vol. 50, no. 8, pp. 1123–1134, Aug. 2005.
- [33] R. J. Elliott, Z. Chen, and Q. Duan, "Insurance claims modulated by a hidden Brownian marked point process," *Insurance, Math. Econ.*, vol. 45, no. 2, pp. 163–172, Oct. 2009.
- [34] Q. Duan and J. Liu, "Filtering and smoothing formulas of AR(p)-modulated Poisson processes," *Commun. Statist., Simul. Comput.*, vol. 49, no. 6, pp. 1575–1591, Jun. 2020.
- [35] Y. Hu, "Micro-grid state estimation using belief propagation on factor graphs," in *Proc. Annu. Summit Conf.*, Singapore, Dec. 2010, pp. 1–10.
- [36] R. Durrett, *Probability: Theory Examples*, Cambridge, U.K.: Cambridge Univ. Press, 2019.
- [37] T. Richardson, "Error floors of LDPC codes," in *Proc. 41st Allerton Conf. Commun., Control, Comput.* Monticello, IL, USA: Allerton House, Oct. 2003, pp. 1019–1087.
- [38] W. W. Hager, "Updating the inverse of a matrix," *SIAM Rev.*, vol. 31, no. 2, pp. 221–239, Jun. 1989.
- [39] Q. Duan, J. Liu, and D. Zhao, "Short term electric load forecasting using an automated system of model choice," *Int. J. Electr. Power Energy Syst.*, vol. 91, pp. 92–100, Oct. 2017.
- [40] E. del Barrio, E. Giné, and C. Matrán, "Central limit theorems for the Wasserstein distance between the empirical and the true distributions," *Ann. Probab.*, vol. 27, no. 2, pp. 1009–1071, Apr. 1999.



QIHONG DUAN received the B.S. and M.S. degrees in computational mathematics, and the Ph.D. degree in probability and statistics from Xi'an Jiaotong University, in 1993, 1996, and 2001, respectively.

He is currently working as an Associate Professor at Xi'an Jiaotong University. He has (co)authored more than 30 scientific journal articles. His research interests include monitoring of power systems and probability model.



FENG LI received the B.S. degree in theoretical and applied mechanics and the M.S. degree in computer science and engineering from Jilin University, in 2007 and 2011, respectively. He is currently pursuing the Ph.D. degree in mathematics with Xi'an Jiaotong University.

His research interest includes the development and applications of wireless communication using artificial intelligence, big data, and mathematical techniques.



JUNRONG LIU received the B.S. and M.S. degrees in mathematics from Shanxi Normal University, in 2000 and 2004, respectively, and the Ph.D. degree in mathematics from Northwest University, in 2016.

She is currently working as a Lecturer at Northwestern University. She has (co)authored more than ten scientific journal articles. Her research interests include mathematical modeling and partial differential equations.

• • •