An information theoretic approach for knowledge representation using Petri nets

Manuel Chiachío, Juan Chiachío, Darren Prescott and John Andrews
Resilience Engineering Research Group
Faculty of Engineering
University of Nottingham
Nottingham, NG7 2RD, UK
Email:{manuel.chiachio, juan.chiachioruano, darren.prescott, john.andrews}@nottingham.ac.uk

Abstract—A new hybrid approach for Petri nets (PNs) is proposed in this paper by combining the PNs principles with the foundations of information theory for knowledge representation. The resulting PNs have been named *Plausible Petri nets* (PPNs) mainly because they can handle the evolution of a discrete event system together with uncertain (plausible) information about the system using *states of information*. This paper overviews the main concepts of classical PNs and presents a method to allow uncertain information exchange about a state variable within the system dynamics. The resulting methodology is exemplified using an idealized expert system, which illustrates some of the challenges faced in real-world applications of PPNs.

Keywords—*Petri nets; Expert Systems; Knowledge representation*

I. INTRODUCTION

A. Classical Petri Nets

Petri Nets (PNs) are bipartite directed graphs (digraph) used for modeling the dynamics of systems, which are known after the celebrated thesis dissertation *Kommunikation mit Automaten* by Carl Petri in 1962 [1]. Two types of nodes are represented in a PN: *transitions* and *places*, which are temporarily visited by *tokens*, the abstract moving units of a PN that can adopt different meanings depending on the application (also different meanings within a specific PN). The distribution of tokens over the PN at a specific time of execution is referred to as *marking*.

From a mathematical perspective, a PN can be defined as an ordered 6-tuple $\mathfrak{P}$ as follows [2]:

$$\mathfrak{P} \triangleq \langle \mathbf{P}, \mathbf{T}, \mathbf{E}, \mathbf{M}_0, \mathbf{D}, \mathbf{W} \rangle \tag{1}$$

where $\mathbf{P} \in \mathbb{N}^{n_p}$ and $\mathbf{T} \in \mathbb{N}^{n_t}$ are vectors to denote the set of n_p places and n_t transitions of the PN respectively, $\mathbf{M}_0 \in \mathbb{N}^{n_p}$ is the initial marking, and $\mathbf{D} \in \mathbb{R}^{n_t}$ is the non-negative vector of switching delays of the transitions (0 by default). The connections between transitions and places are expressed through the set of edges $\mathbf{E} \subseteq (\mathbf{P} \times \mathbf{T}) \cup (\mathbf{T} \times \mathbf{P})$, so that $\mathbf{E} \subset \mathbb{N}^{n_p} \times \mathbb{N}^{n_t}$ contains ordered pairs of nodes. Each edge (also referred to as arc) has assigned a weight, a non-negative integer value (1 by default) defined in the set of weights $\mathbf{W}$.

There are three main laws that govern a PN, that can be enunciated as follows:

1) Transitions always consume from all the input arcs at the same time;

2) Transitions always produce from all out-coming arcs the same amount;

3) Transitions cannot consume tokens from an empty place.

The dynamics of classical PNs can be described through a state equation defined as follows:

$$\mathbf{M}_{k+1} = \mathbf{M}_k + \mathbf{A}^T \mathbf{u}_k \tag{2}$$

where $\mathbf{u}_k$ is the *firing vector* at state k, an n_t-dimensional vector of Boolean values whose elements are obtained according to the *firing rule*, as will be explained further below. $\mathbf{A}$ is an $n_t \times n_p$ matrix typically referred to as the *incidence matrix*, whose elements are the result of subtracting the *forward* $(\mathbf{A}^+)$ and *backward* $(\mathbf{A}^-)$ *incidence matrices* respectively, as follows:

$$\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^- \tag{3}$$

where

$$\mathbf{A}^+ = \begin{pmatrix} a_{11}^+ & a_{12}^+ & \cdots & a_{1n_p}^+ \\ a_{12}^+ & a_{22}^+ & \cdots & a_{2n_p}^+ \\ \vdots & & \ddots & \\ a_{n_t1}^+ & a_{n_t2}^+ & \cdots & a_{n_tn_p}^+ \end{pmatrix} \quad \mathbf{A}^- = \begin{pmatrix} a_{11}^- & a_{12}^- & \cdots & a_{1n_p}^- \\ a_{12}^- & a_{22}^- & \cdots & a_{2n_p}^- \\ \vdots & & \ddots & \\ a_{n_t1}^- & a_{n_t2}^- & \cdots & a_{n_tn_p}^- \end{pmatrix} \tag{4}$$

In the last equation, the element a_{ij}^+ is the weight of the arc from transition t_i to output place p_j, whereas a_{ij}^- is the weight of the arc to transition t_i from input place p_j, where $i = 1, \ldots, n_t$, $j = 1, \ldots, n_p$. If transition t_i is activated at state k, then $u_{i,k} \in \mathbf{u}_k$ is modified according to the firing rule, which can be expressed as follows:

$$u_{i,k} = \begin{cases} 1, & \text{if} \quad M(j) \geqslant a_{ij}^- \ \forall p_j \in {}^{\bullet}\mathbf{P}_{t_i} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $M(j) \in \mathbb{N}$ is the marking for place p_j, and ${}^{\bullet}\mathbf{P}_{t_i}$ denotes the set of places that belong to the pre-set of transition t_i, $i = 1, \ldots, n_t$, as will be next described.

B. Tokens, probability, and information

Despite the efficiency that PNs have demonstrated for modelling expert system dynamics for a widespread range of applications, their pragmatism is not commensurate with the

reality of certain systems [3], [4]. From the last years, a significant increase in the research activity has been observed in the literature about the definition of new kind of tokens as a way to make PNs better aligned with the reality of the systems to be idealized. Traditionally, tokens in classical PNs are interpreted as moving objects (typically integer units) in a network of interconnected places, as mentioned in Section I-A. Further approaches encountered in the literature consider different kind of tokens, like for example numerical tokens [5], [6], fuzzy tokens [4], [7], particle tokens [8], [9], to cite just but any, each one providing some changes on the net dynamics, yet defining a variant of the classical PNs.

In this work, a hybrid approach is proposed by combining traditional tokens with a new class of tokens that represent a degree of belief about a state variable at a certain mode which is specified by a place. The resulting framework has been named *Plausible Petri Nets* (PPN) since certain tokens enclose a set of numerical values about a state variable $\mathbf{x}$ from a finite-dimensional set $\mathcal{X} \subset \mathbb{R}^d$, along with a mapping over $\mathcal{X}$ denoted by $f : \mathcal{X} \subset \to \mathbb{R}^+$ that assigns each element $\mathbf{x}$ with a non-negative value interpreted as its *relative plausibility*, as will be discussed in the next section. In PPNs, the moving units are objects (in the sense of integer units as in classical PNs) and also *states of information* about $\mathbf{x}$. Moreover, the consumption and production of tokens in PPNs are not only associated with the concept of adding and subtracting integer values, but also with the operation of adding and subtracting information about the state variable $\mathbf{x}$.

The next section provides further insight into the concepts of plausibility, uncertainty, and states of information, since they are fully adopted throughout the text. Section III is devoted to providing the basis and main rules of the proposed PPNs. In Section IV, our framework is illustrated using an example of application. Finally, Section V gives concluding remarks.

II. BASIC OPERATIONS IN PLAUSIBLE PETRI NETS

Let $\mathcal{A} \subseteq \mathcal{X}$ be a subset representing a certain event or proposition over $\mathcal{X}$ in such a way that there exist a probability measure $P(\mathcal{A})$ through a density function $f(\mathbf{x})$ that can be normalized such that $\int_{\mathcal{X}} f(\mathbf{x})d\mathbf{x} = 1$. In this sense, $P(\mathcal{A})$ is interpreted as the plausibility of a set of possible values $\mathbf{x}$ given a *state of information* about them provided by $f(\mathbf{x})$ [10]–[12]. Let us now evoke the first principles of Boolean logic by recalling the concepts of the logic operators AND ($\wedge$) and OR ($\vee$) for the conjunction and disjunction of propositions, respectively. Then, the following logical relationships are compatible according to the De Morgan's law [13], [14] for two probability measures $P_a(\mathcal{A})$ and $P_b(\mathcal{A})$, each one ascribed to the state of information $f_a(\mathbf{x})$ and $f_b(\mathbf{x})$, respectively:

$$P_a(\mathcal{A}) = 0 \quad \text{OR} \quad P_b(\mathcal{A}) = 0 \quad \Rightarrow \quad (P_a \wedge P_b)(\mathcal{A}) = 0$$
$$P_a(\mathcal{A}) \neq 0 \quad \text{OR} \quad P_b(\mathcal{A}) \neq 0 \quad \Rightarrow \quad (P_a \vee P_b)(\mathcal{A}) \neq 0$$
$$(6)$$

where $(P_a \wedge P_b)(\mathcal{A})$ and $(P_a \vee P_b)(\mathcal{A})$ can be expressed as [15]:

$$(P_a \wedge P_b)(\mathcal{A}) = \int_{\mathcal{A}} (f_a \wedge f_b)(\mathbf{x})d\mathbf{x} \quad (7a)$$

$$(P_a \vee P_b)(\mathcal{A}) = \int_{\mathcal{A}} (f_a \vee f_b)(\mathbf{x})d\mathbf{x} \quad (7b)$$

In the last equation, the densities $(f_a \wedge f_b)(\mathbf{x})$ and $(f_a \vee f_b)(\mathbf{x})$ represent the *conjunction* and *disjunction of states of information* given by $f_a(\mathbf{x})$, $f_b(\mathbf{x})$, respectively [12], [16]. Besides, $(P_a \wedge P_b)(\mathcal{A})$ and $(P_a \vee P_b)(\mathcal{A})$ stand for the plausibility of the set of values $\mathbf{x} \in \mathcal{A}$ in compliance with the information given by $(f_a \wedge f_b)(\mathbf{x})$ and $(f_a \vee f_b)(\mathbf{x})$, respectively.

Next, because commutativity is allowed in logic of propositions, i.e. $(P_a \vee P_b)(\mathcal{A}) = (P_b \vee P_a)(\mathcal{A})$ and $(P_a \wedge P_b)(\mathcal{A}) = (P_b \wedge P_a)(\mathcal{A})$, then a simple solution for the unnormalized densities $(f_a \wedge f_b)(\mathbf{x})$ and $(f_a \vee f_b)(\mathbf{x})$ are such that [17]:

$$(f_a \vee f_b)(\mathbf{x}) = f_a(\mathbf{x}) + f_b(\mathbf{x}) \quad (8a)$$

$$(f_a \wedge f_b)(\mathbf{x}) = \frac{f_a(\mathbf{x})f_b(\mathbf{x})}{\mu(\mathbf{x})} \quad (8b)$$

where $\mu(\mathbf{x})$ is the *homogeneous density function* [12], [18], a reference state of information that can be understood as a probability model for $\mathbf{x} \in \mathcal{X}$ in absence of any other information, which actually represents the *state of total ignorance* about $\mathcal{X}$ [18], [19]. In case that $\mathcal{X}$ is a linear space, it is demonstrated [17] that $\mu(\mathbf{x}) = const$, i.e., $\mu(\mathbf{x})$ is a uniform over $\mathcal{X}$. See Fig. 1 for a conceptual illustration of the conjunction and disjunction of states of information over some arbitrary densities $f_a(\mathbf{x})$ and $f_b(\mathbf{x})$. Moreover, it is worth noting that the conjunction and disjunction operations over states of information can be extended to the case of multiple states of information (e.g. $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})$), as follows [18]:

$$(f_1 \vee \cdots \vee f_n)(\mathbf{x}) = \sum_{i=1}^{n} f_i(\mathbf{x}) \quad (9a)$$

$$(f_1 \wedge \cdots \wedge f_n)(\mathbf{x}) = \prod_{i=1}^{n} \frac{f_i(\mathbf{x})}{\mu(\mathbf{x})} \quad (9b)$$

Finally, observe that the normalization of the density function in (8b) may require the evaluation of an intractable integral such that $\int_{\mathcal{X}} \frac{1}{\alpha} \frac{f_a(\mathbf{x})f_b(\mathbf{x})}{\mu(\mathbf{x})} d\mathbf{x} = 1$, where α is a normalizing constant. Moreover, note that there might be situations where the conjunction is conducted with density functions which are not completely known, perhaps because they are defined trough samples. Hence, sampling-based algorithms (e.g. particle methods) [20], [21] can be used in those cases to circumvent the evaluation of the normalizing constant. In particle methods, a set of N samples $\{\mathbf{x}^{(n)}\}_{n=1}^{N}$ with associated weights $\{w^{(n)}\}_{n=1}^{N}$ are used to obtain an approximation for the required density function [e.g. $(f_a \wedge f_b)(\mathbf{x})$] with a feasible computational cost, as follows:

$$(f_a \wedge f_b)(\mathbf{x}) \approx \sum_{n=1}^{N} w^{(n)} \delta(\mathbf{x} - \mathbf{x}^{(n)}) \quad (10)$$

where $\mathbf{x}^{(n)} \sim (f_a \wedge f_b)(\mathbf{x})$, and δ is the Dirac delta. The particle weight $w^{(n)}$ represents a likelihood value of the n-th particle, which can be obtained as follows:

$$w^{(n)} = \frac{f_a(\mathbf{x}^{(n)})f_b(\mathbf{x}^{(n)})}{\sum_{n=1}^{N} f_a(\mathbf{x}^{(n)})f_b(\mathbf{x}^{(n)})} \quad (11)$$

where $\mathcal{X}$ is assumed to be a linear space.

(a) Conjunction of $f_a(\mathbf{x})$ and $f_b(\mathbf{x})$

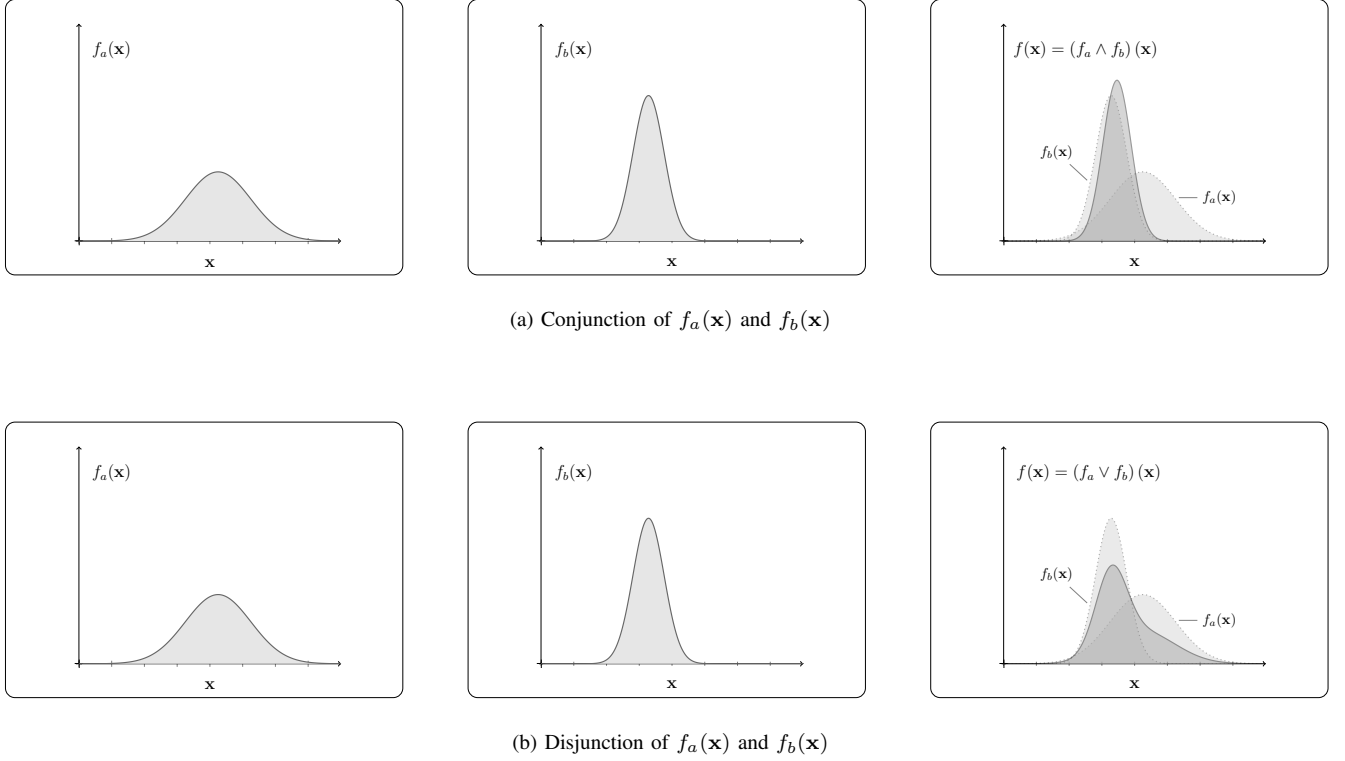(b) Disjunction of $f_a(\mathbf{x})$ and $f_b(\mathbf{x})$

Fig. 1. Illustrative example of the conjunction and disjunction operation over two states of information, namely $f_a(\mathbf{x})$ (left panels) and $f_b(\mathbf{x})$ (center panels). The right panels represent the resulting density $f(\mathbf{x})$ from the conjunction (upper right) and disjunction (lower right) of f_a and f_b. The resulting PDF is represented superimposed over f_a and f_b for better understanding.

III. INFORMATION FLOW DYNAMICS IN PPNs

A. Modelling assumptions

Let $\{\mathbf{x}_k, k \in \mathbb{N}\}$ be a stochastic process taking values in $\mathcal{X}$, which is manifested through different modes corresponding to each place of the Petri net. Next, let us consider that $\mathbf{P} \in \mathfrak{P}$ can be partitioned into two disjoint sets: 1) numerical places $\mathbf{P}^{(\mathcal{N})} \in \mathbb{N}^{n_p}$, and 2) symbolic places $\mathbf{P}^{(\mathcal{S})} \in \mathbb{N}^{n'_p}$, such that $\mathbf{P}^{(\mathcal{N})} \cup \mathbf{P}^{(\mathcal{S})} = \mathbf{P} \in \mathbb{N}^{n_p+n'_p}$, and $\mathbf{P}^{(\mathcal{N})} \cap \mathbf{P}^{(\mathcal{S})} = \emptyset$. Analogously, the set of transitions $\mathbf{T}$ is partitioned into numerical transitions $\mathbf{T}^{(\mathcal{N})} \in \mathbb{N}^{n_t}$, and symbolic transitions $\mathbf{T}^{(\mathcal{S})} \in \mathbb{N}^{n'_t}$, where $\mathbf{T}^{(\mathcal{N})} \cup \mathbf{T}^{(\mathcal{S})} = \mathbf{T} \in \mathbb{N}^{n_t+n'_t}$, and $\mathbf{T}^{(\mathcal{N})} \cap \mathbf{T}^{(\mathcal{S})} \neq \emptyset$. Those transitions that belong to $\mathbf{T}^{(\mathcal{N})} \cap \mathbf{T}^{(\mathcal{S})}$ are referred to as *mixed transitions*. Finally, let us call by ${}^{\bullet}\mathbf{P}_{t_i}$, the subset of places from the pre-set of transition t_i, $i = 1, \ldots, n_t$. Analogously $\mathbf{P}_{t_i}^{\bullet}$ is to denote the subset of places that belongs to the post-set of t_i. From this standpoint, the dynamics of PPNs is formulated under the adoption of the following items:

1) At a certain time k, the place $p_j^{(\mathcal{N})}$ encloses a state of information about $\mathbf{x}_k$ given by $f_k^{p_j}$, where $f_k^{p_j} : \mathcal{A}_j \to \mathbb{R}^+$ and $\mathcal{A}_j$ is a subset of $\mathcal{X}$;

2) Equivalently, any transition $t_i \in \mathbf{T}^{(\mathcal{N})}$ carries a state of information about $\mathbf{x}_k$ defined over $\mathcal{A}_i \subseteq \mathcal{X}$, namely $f_k^{t_i}(\mathbf{x}_k)$;

3) As in classical PNs, there exist arc weights for the symbolic places, denoted by $a_{ij}^{'+}, a_{ij}^{'-} \in \mathbf{W}^{(\mathcal{S})} \subset \mathbb{N}$, whereby the incidence matrix $\mathbf{A}^{(\mathcal{S})}$ can be obtained as: $\mathbf{A}^{(\mathcal{S})} = a_{ij}^{'+} - a_{ij}^{'-}$, $i = 1, \ldots, n'_t$, $j = 1, \ldots, n'_p$, where n'_p, n'_t are the number of symbolic places and transitions of the PPNs, respectively. The arc weights for the numerical places are denoted by $a_{ij}^+, a_{ij}^- \in \mathbf{W}^{(\mathcal{N})} \subset \mathbb{R}^+$, where $\mathbf{A}^{(\mathcal{N})} = a_{ij}^+ - a_{ij}^-$, and $i = 1, \ldots, n_t$, $j = 1, \ldots, n_p$, being in this case n_t, n_p the amount of numerical transitions and numerical places of the PPN, respectively. These weights provide us with a measure of the importance of the information that flows from/to the corresponding transition;

4) Any transition t_i always consumes from all its input arcs at the same time. Moreover, an input arc from place $p_j^{(\mathcal{N})}$ to transition $t_i \in \mathbf{T}^{(\mathcal{N})}$ conveys a state of information given by the conjunction of states of information of $f_k^{p_j}(\mathbf{x}_k)$ and $f_k^{t_i}(\mathbf{x}_k)$, namely $(f_k^{p_j} \wedge f_k^{t_i})(\mathbf{x}_k)$;

5) Any transition t_i produces to its output arcs the same amount of information given by the conjunction of states of information between $f_k^{t_i}(\mathbf{x}_k)$ and $f^{{}^{\bullet}\mathbf{P}_{t_i}}(\mathbf{x}_k)$, i.e., $(f_k^{t_i} \wedge f^{{}^{\bullet}\mathbf{P}_{t_i}})(\mathbf{x}_k)$. In the last expression, $f^{{}^{\bullet}\mathbf{P}_{t_i}}(\mathbf{x}_k)$ results from the disjunction of the states of information of the pre-set of transition t_i, i.e., $f^{{}^{\bullet}\mathbf{P}_{t_i}}(\mathbf{x}_k) = (f^{p_{t_i,1}} \vee$

$f^{p_{t_i,2}} \vee \cdots \vee f^{p_{t_i,m}})(\mathbf{x}_k)$ (recall [9a]), where places $p_{t_i,1}, p_{t_i,2}, \ldots, p_{t_i,m} \in {}^{\bullet}\mathbf{P}_{t_i} \subset \mathbf{P}^{(\mathcal{N})}$;

6) If numerical place $p_j^{(\mathcal{N})}$ belongs to the pre-set of t_i, and assuming that[1] $p_j^{(\mathcal{N})} \notin \mathbf{P}_{t_i}^{\bullet} \ \forall \tilde{i} = 1, \ldots, n_t$, then the state of information that remains in $p_j^{(\mathcal{N})}$ at $k+1$ after firing t_i, is the conjunction of $f_k^{p_j}(\mathbf{x}_k)$ and $f_k^{t_i}(\mathbf{x}_k)$ weighted according to a_{ij}^-, i.e. $f_{k+1}^{p_j}(\mathbf{x}_{k+1}) = a_{ij}^-(f_k^{p_j} \wedge f_k^{t_i})(\mathbf{x}_k)$;

7) Correspondingly, the state of information resulting in place $p_{\tilde{j}}^{(\mathcal{N})}$ after firing t_i given that $p_j^{(\mathcal{N})} \in \mathbf{P}_{t_i}^{\bullet}$, is the disjunction of the states of information $f_k^{p_j}(\mathbf{x}_k)$ and $(f_k^{t_i} \wedge f^{\bullet \mathbf{P}_{t_i}})(\mathbf{x}_k)$, where the latter is multiplied by its output weight a_{ij}^+. Mathematically $f_{k+1}^{p_j}(\mathbf{x}_{k+1}) = \left(f_k^{p_j} \vee a_{ij}^+ \left(f_k^{t_i} \wedge f^{\bullet \mathbf{P}_{t_i}} \right) \right)(\mathbf{x}_k)$.

Observe that the rules given above, specifically those from items 3) and 5), are in agreement with the rules of classical PNs given in Section I-A except for the consideration that here, part of the flow of tokens is based on states of information.

B. Marking evolution

In PPNs, the marking at state k consists of both types of information given by $\mathbf{M}_k^{(\mathcal{N})}$ for the numerical places, and $\mathbf{M}_k^{(\mathcal{S})}$ for the case of symbolic places, so that $\mathbf{M}_k = \left(\mathbf{M}_k^{(\mathcal{N})}, \mathbf{M}_k^{(\mathcal{S})} \right)$. Mathematically, $\mathbf{M}_k^{(\mathcal{N})}$ is expressed through a column vector specified by $\mathbf{M}_k^{(\mathcal{N})} = \left(f_k^{p_1}(\mathbf{x}_k), f_k^{p_2}(\mathbf{x}_k), \ldots, f_k^{p_{n_p}}(\mathbf{x}_k) \right)^T$, such that each $f_k^{p_j}(\mathbf{x}_k)$ is a normalized density function that provides us with a measure of the relative plausibility of $\mathbf{x}_k$ at place $p_j^{(\mathcal{N})}$, $j = 1, \ldots, n_p$. Similarly, $\mathbf{M}_k^{(\mathcal{S})}$ is expressed through a column vector of integer values so that its jth component represents the number of tokens present in place $p_j^{(\mathcal{S})}$ at state k. The marking evolution of $\mathbf{M}_k^{(\mathcal{S})}$ for the symbolic subnet has been given in (2), which corresponds to the state equation of an ordinary PN [2]. However, the marking evolution corresponding to the numerical places gives response to the information flow dynamics described in Section III-A, and in particular, results from applying the rules 4) to 7). Both markings, namely $\mathbf{M}_k^{(\mathcal{N})}$ and $\mathbf{M}_k^{(\mathcal{S})}$, evolve in a synchronized manner driven by the firing rule for PPNs, as will be explained below.

C. Firing rule for PPNs

In PPNs, any transition $t_i \in \mathbf{T}$ is fired at state k if the delay time has passed and:

1) $\forall p_j^{(\mathcal{S})} \in {}^{\bullet}\mathbf{P}_{t_i}$, $M_k^{(\mathcal{S})}(j) \geqslant a_{ij}^{'-}$, for $t_i \in \mathbf{T}^{(\mathcal{S})}$, $t_i \notin (\mathbf{T}^{(\mathcal{S})} \cap \mathbf{T}^{(\mathcal{N})})$;

2) $\forall p_j^{(\mathcal{N})} \in {}^{\bullet}\mathbf{P}_{t_i}$, $(f_k^{p_j} \wedge f_k^{t_i})(\mathbf{x}_k) \neq \emptyset$, which, in this case, applies for numerical transitions, i.e., $t_i \in \mathbf{T}^{(\mathcal{N})}$, $t_i \notin (\mathbf{T}^{(\mathcal{S})} \cap \mathbf{T}^{(\mathcal{N})})$;

3) Condition 1) and 2) are both satisfied for numerical places and symbolic places from the pre-set of t_i, when t_i is a mixed transition, i.e. $t_i \in (\mathbf{T}^{(\mathcal{S})} \cap \mathbf{T}^{(\mathcal{N})})$.

Condition 1) means that each symbolic place from the pre-set of t_i has enough amount of tokens according to their input arc weight, as in classical PNs. On the other hand, condition 2) allows us to ensure that every conjunction of states of information between f^{t_i} and each of the density functions from the pre-set of t_i, are possible. Note that a conjunction (e.g. $(f_k^{p_j} \wedge f_k^{t_i})(\mathbf{x}_k)$) is possible if $(f_k^{p_j} \wedge f_k^{t_i})(\mathbf{x}_k) \neq \emptyset$ for any subset $\mathcal{B} \in \mathcal{X}$ [12].

IV. APPLICATION EXAMPLE

In this section, a numerical example is provided to illustrate our proposed methodology. To this end, let us consider that there exist a measurement about a bi-dimensional state variable $\mathbf{x}_k = (x_{1,k}, x_{2,k})$, performed using two sensors denoted by s_1 and s_2. The information from s_1 and s_2 is initially represented in places $p_1^{(\mathcal{N})}$ and $p_2^{(\mathcal{N})}$ using probability densities as respective states of information about $\mathbf{x}_k$. Suppose now that the sensors are imperfect so that they introduce uncertainty that can modeled as a zero mean Gaussian density function with covariance matrix given by $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively. Next, consider that the component-wise measurements are stochastically independent so that $\boldsymbol{\Sigma}_1 = diag(\sigma_{v_1}^2, \sigma_{v_2}^2)$, $\boldsymbol{\Sigma}_2 = diag(\sigma_{w_1}^2, \sigma_{w_2}^2)$, where $\sigma_{v_1}, \sigma_{v_2}, \sigma_{w_1}, \sigma_{w_2}$ are the standard deviations of measurement errors $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$, for sensors s_1 and s_2, respectively. The system being modeled is based on an expert system that controls the activation of a discrete-event subsystem (which can represent an automata, machine, or similar, in a real-world application). Once a new data point arrives, then the activation/deactivation of the discrete subsystem occurs conditioned upon the quality of the information coming from the sensors. The differential entropy, as a measure of the data uncertainty, is used in this example as a quality indicator of the information given by the sensors. Fig. 2 provides a geometric description of the PPN being modeled using four numerical places ($p_1^{(\mathcal{N})}$ to $p_4^{(\mathcal{N})}$), five symbolics places ($p_1^{(\mathcal{S})}$ to $p_5^{(\mathcal{S})}$), and three mixed transitions labeled as t_1, t_2, and t_3. A cold transition (ϵ) is used to represent the data arrival to the system. In our PPN model, the information from sensors s_1 and s_2, initially represented in places $p_1^{(\mathcal{N})}$ and $p_2^{(\mathcal{N})}$, is further gathered in place $p_3^{(\mathcal{N})}$ provided that transition t_1 is fired. The joint information in place $p_3^{(\mathcal{N})}$ is then used to activate a sequence of discrete-event system (e.g. a machine) if the differential entropy of $\mathbf{x}_k$ is lower than a certain reference value $\xi \in \mathbb{R}$, which is taken as $\xi = 7$ in this example. Next, the state of information about $\mathbf{x}_k$ is finally collected in place $p_4^{(\mathcal{N})}$. Otherwise, transition t_2 is activated whereby the information in place $p_2^{(\mathcal{N})}$ is improved using the joint information about $\mathbf{x}_k$ given in place $p_3^{(\mathcal{N})}$. Note that $p_2^{(\mathcal{S})}$ helps in synchronizing transitions t_1 and t_3, so as to avoid them to be activated simultaneously.

---

[1]The assumption implies that place $p_j^{(\mathcal{N})}$ does not belong to the post-set of any transition.

The initial marking for the numerical places is $\mathbf{M}_0^{(\mathcal{N})} = \left(f_0^{p_1}(\mathbf{x}_0), f_0^{p_2}(\mathbf{x}_0), f_0^{p_3}(\mathbf{x}_0), f_0^{p_4}(\mathbf{x}_0) \right)^T$, where $f_0^{p_1} \sim \mathcal{N}(\hat{\mathbf{x}}, \boldsymbol{\Sigma}_1)$, $f_0^{p_2} \sim \mathcal{N}(\hat{\mathbf{x}} + \boldsymbol{\beta}, \boldsymbol{\Sigma}_2)$, $f_0^{p_3} = f_0^{p_4} = \emptyset$. In the last expressions, $\hat{\mathbf{x}}$ is the data point that arrives to the system, and $\boldsymbol{\beta}$ is a bias that applies to the measurements from sensor s_2. In this example, $\hat{\mathbf{x}} = (20, 15)$, and $\boldsymbol{\beta} = (5, 5)$, both expressed in arbitrary units. The covariance matrices for $f_0^{p_1}(\mathbf{x}_0)$ and $f_0^{p_2}(\mathbf{x}_0)$ are given by $\boldsymbol{\Sigma}_1 = diag(4^2, 3^2)$ and $\boldsymbol{\Sigma}_2 = diag(3^2, 3^2)$, respectively. See Fig. 3 for a particle-based representation of the states of information for places $p_1^{(\mathcal{N})}$ and $p_2^{(\mathcal{N})}$ at $k = 0$.

On the other side, the initial marking for symbolic places is given by $\mathbf{M}_0^{(\mathcal{S})} = (0, 0, 0, 0, 1)^T$. The mixed transitions t_i, $i = \{1, 2, 3\}$ are defined using Diract Delta density functions, i.e., $f^{t_i}(\mathbf{x}_k) \sim \mathbb{I}_{\mathcal{B}_i}(\mathbf{x}_k)$, thus their firing is prescribed for the state variable $\mathbf{x}_k$ on fulfilling the condition $\mathbf{x}_k \in \mathcal{B}_i$, where:

$$\mathcal{B}_1 = \left\{ \mathbf{x}_k \in \mathcal{X} : (x_{1,k} - 20)^2 + (x_{2,k} - 20)^2 \leqslant 10 \right\} \quad (12a)$$

$$\mathcal{B}_2 = \left\{ \mathbf{x}_k \in \mathcal{X} : H(\mathbf{x}_k) \geqslant \xi \right\} \quad (12b)$$

$$\mathcal{B}_3 = \left\{ \mathbf{x}_k \in \mathcal{X} : H(\mathbf{x}_k) < \xi \right\} \quad (12c)$$

In (12b) and (12c), $H : \mathcal{X} \to \mathbb{R}$ denotes the differential entropy of $\mathbf{x}_k$, that can be obtained by calculating[2] $^1/_2 \ln \left[(2\pi e)^d \det \left[cov(\mathbf{x}_k) \right] \right]$ as a measure quantifying the uncertainty of $\mathbf{x}_k$. Here, we assume for simplicity that the time spent by the machine in performing transitions t_4 and t_5 is negligible. The rules for the information flow dynamics of PPNs [recall rules 4) to 7) in Section III-A] along with (2), are applied in confluence with the firing rule for the system state evolution described through the marking $\mathbf{M}_k$. The particle approximation for conjunction of states of information described in Section II, has been applied to evaluate the conjunction of states of information resulting from applying

---

[2]This expression for the differential entropy is actually an upper-bound approximation to the actual differential entropy, where the exactness is achieved when the density function is Gaussian.

Fig. 3. Representation of the density functions $f_k^{p_1}(\mathbf{x}_k)$ and $f_k^{p_1}(\mathbf{x}_k)$ for $k = 0$ using samples (circles) in the $\mathbf{x}$-space. The green circle represents a data point arrived to the system.
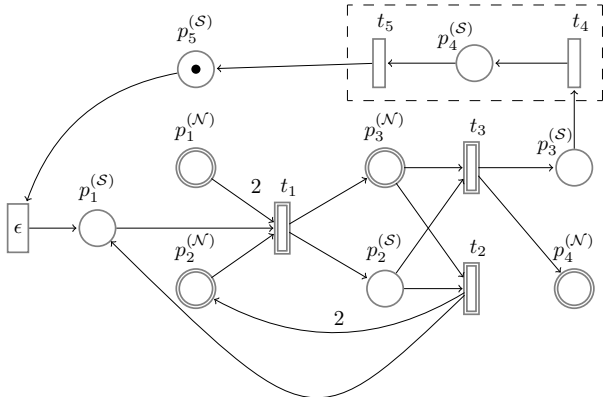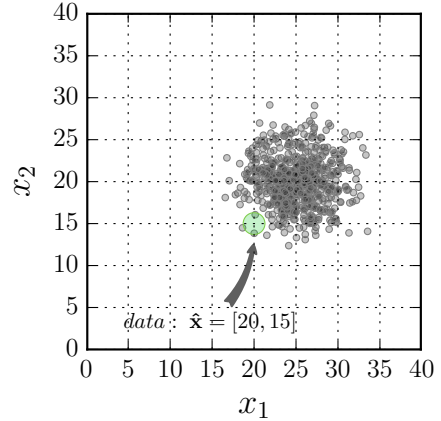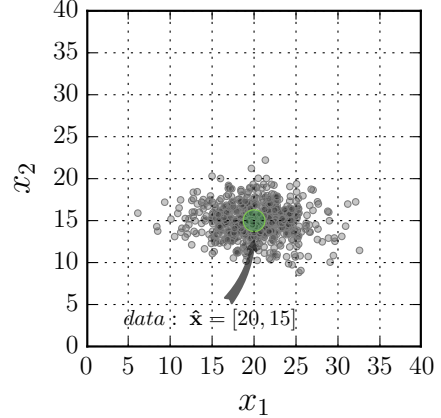

Fig. 2. Plausible Petri net of example in Section IV. Numerical places and transitions are represented using double line to distinguish them from the symbolic ones. Transitions t_4, t_5, and place $p_4^{(\mathcal{S})}$ under the dashed box represent a discrete-event subsystem in a simplified manner.

rules 4) to 7) from the information flow dynamics for the numerical subnet. The disjunction of states of information are straightforwardly evaluated using samples by just joining the samples from the component-wise density functions, and affecting their particle weights using an appropriate normalizing constant so as to obtain a bone fide density.

The results for the numerical places are presented for states $k = 0$ and $k = 2$ in Fig. 4 (the states of information about $\mathbf{x}_k$ remain unchanged in $\mathbf{P}^{(\mathcal{N})}$ after $k = 2$, hence they are not represented). A summary of the results for the analysis of the PPN is provided in Table I. Note from Table I that the results for symbolic marking and active transitions are column vectors, although they are not explicitly reflected as so in this table for clarity. Also, the fourth column represents the continuum entropy of $f_k^{p_3}(\mathbf{x}_k), k = 1, 2, 3$. The incidence matrices used for calculations of the numerical and symbolic

TABLE I.   SUMMARY OF THE RESULTS FROM THE ANALYSIS OF THE
PPN SHOWN IN FIG. 2 FOR THREE STATES STARTING FROM $k = 0$.

State	$\mathbf{M}_k^{(\mathcal{S})}$	$\mathbf{u}_k$	CE
$k = 0$	$(1\,0\,0\,0\,0)$	$(1\,0\,0\,0\,0)$	–
$k = 1$	$(0\,1\,0\,0\,0)$	$(0\,0\,1\,0\,0)$	6.21
$k = 2$	$(0\,0\,1\,0\,0)$	$(0\,0\,0\,1\,0)$	6.25
$k = 3$	$(0\,0\,0\,1\,0)$	$(0\,0\,0\,0\,1)$	6.25

subnets are given by:

$$\mathbf{A}^{(\mathcal{N})} = \begin{pmatrix} -2 & -1 & 1 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \mathbf{A}^{(\mathcal{S})} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{13}$$

respectively. Observe from the analysis that numerical and symbolic subnets interact theirselves in a synchronized manner. To serve as an example, note that transition t_1 is activated after the data arrival at $k = 0$ because $p_1^{(\mathcal{S})}$ is assigned one token and $(f_0^{p_i} \wedge f_0^{t_1})(\mathbf{x}_k) \neq \emptyset$, for $i = 1, 2$. Note also that t_1 is no longer fired at $k \geqslant 1$ because $p_1^{(\mathcal{S})}$ is empty, hence condition 1) from the firing rule is not fulfilled. The sequence of discrete-event actions (marked using a dashed box in Fig. 2) are activated at $k = 2$ after t_3 is fired, since the entropy condition given through t_3 is fulfilled at such time. Note that this example demonstrates that the flow of information can be altered (activate/deactivate) in our PPNs by combining numerical and symbolic places so as to make their tokens conveniently interact in a synchronized manner. The results for the numerical places $p_1^{(\mathcal{N})}$ and $p_4^{(\mathcal{N})}$ are presented for states $k = 0$ to $k = 2$ in Fig 4.

V. CONCLUSION

A novel hybrid approach for PNs has been proposed in this paper by combining the algebra of classical PNs with the basis of information theory. As in classical PNs, there exist a set of places and transitions (labeled here as "symbolic") to model discrete-event systems using regular tokens as moving units. In addition, a set of numerical places and transitions are considered to model knowledge representation about a system state variable, which interact with the regular tokens in a synchronized manner. The simulated results demonstrate that PPNs are a versatile tool to fuse uncertain information (which can be diverse through the different numerical nodes) about the state variable with sequences of Boolean events, provided that an adequate net architecture is adopted. This fact makes them useful for analyzing hybrid systems with interaction of diverse sources of information, like in expert systems. One future research direction is to explore their ability to model cybersystems, since they can receive, store, exchange, and process information so as to use it for control. Moreover, further research effort is needed to investigate formal aspects of the resulting hybrid system, as well as to explore efficient implementations using a variety of examples of application.

(a) States of information about $\mathbf{x}_0$

(b) States of information about $\mathbf{x}_1$

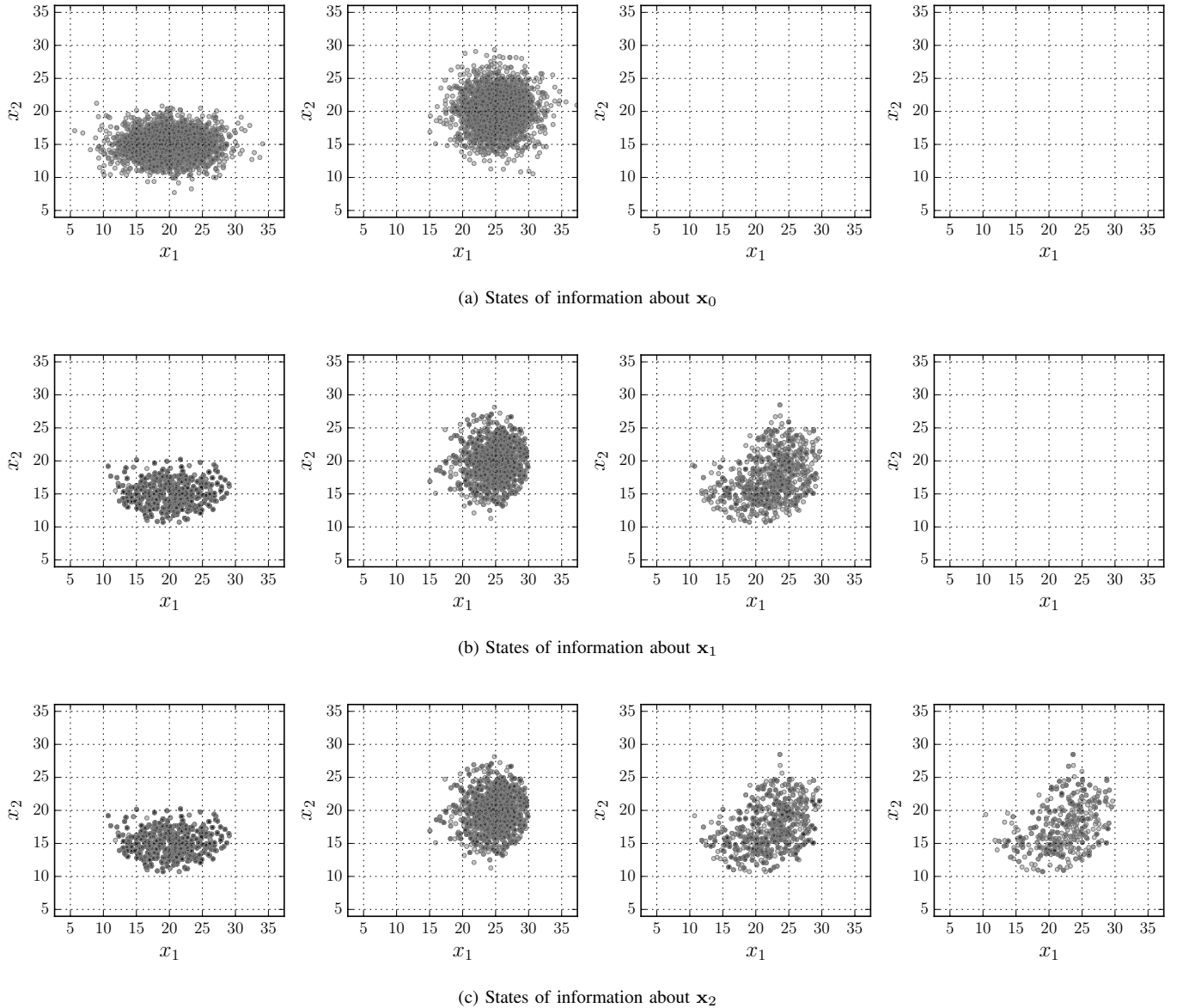(c) States of information about $\mathbf{x}_2$

Fig. 4. PPN output for numerical places $p_1^{(\mathcal{N})}$ (most-left panel) to $p_4^{(\mathcal{N})}$ (most-right panel). Each subplot represents the state of information about $\mathbf{x}_k$, using samples (circles) in the state space $\mathcal{X}$.

REFERENCES

[1] C. A. Petri, "Kommunikation mit automaten," Ph.D. dissertation, Institut für Instrumentelle Mathematik an der Universität Bonn, 1962.

[2] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.

[3] J. Cardoso, R. Valette, and D. Dubois, "Possibilistic Petri nets," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 5, pp. 573–582, 1999.

[4] Z. Ding, Y. Zhou, and M. Zhou, "Modeling self-adaptive software systems with learning Petri nets," in *Companion Proceedings of the 36th International Conference on Software Engineering*. ACM, 2014, pp. 464–467.

[5] G. Horton, V. G. Kulkarni, D. M. Nicol, and K. S. Trivedi, "Fluid stochastic Petri nets: Theory, applications, and solution techniques," *European Journal of Operational Research*, vol. 105, no. 1, pp. 184–201, 1998.

[6] M. Silva, J. Júlvez, C. Mahulea, and C. R. Vázquez, "On fluidization of discrete event models: observation and control of continuous Petri nets," *Discrete Event Dynamic Systems*, vol. 21, no. 4, pp. 427–497, 2011.

[7] S.-M. Chen, J.-S. Ke, and J.-F. Chang, "Knowledge representation using fuzzy Petri nets," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 2, no. 3, pp. 311–319, 1990.

[8] C. Lesire and C. Tessier, "Particle Petri nets for aircraft procedure monitoring under uncertainty," in *Applications and Theory of Petri Nets 2005: 26th International Conference, ICATPN 2005, Miami, USA, June 20-25, 2005. Proceedings*, G. Ciardo, Ed. Springer, 2005, pp. 329–348.

[9] L. Zouaghi, A. Alexopoulos, A. Wagner, and E. Badreddin, "Modified particle Petri nets for hybrid dynamical systems monitoring under environmental uncertainties," in *System Integration (SII), 2011 IEEE/SICE International Symposium on*. IEEE, 2011, pp. 497–502.

[10] R. T. Cox, "Probability, frequency and reasonable expectation," *American journal of physics*, vol. 14, p. 1, 1946.

[11] E. Jaynes, *Papers on probability, statistics and statistical physics*. (Ed. R.D Rosenkrantz), Kluwer Academic Publishers, 1983.

[12] A. Tarantola and B. Valette, "Inverse problems = quest for information,"

Journal of Geophysics, vol. 50, no. 3, pp. 159–170, 1982.

[13] I. M. Copi, *Introduction to logic*. New York, Macmillan, 1953.

[14] R. L. Goodstein, *Boolean algebra*. Dover Publications, Inc., 2007.

[15] A. Kolmogorov, *Foundations of the Theory of Probability (Translation of 1933 original in German)*. Chelsea Publishing: New York, 1950.

[16] G. Rus, J. Chiachío, and M. Chiachío, "Logical inference for inverse problems," *Inverse Problems in Science and Engineering*, vol. 24, no. 3, pp. 448–464, 2016.

[17] A. Tarantola, *Inverse problem theory and methods for model parameters estimation*. SIAM, 2005.

[18] K. Mosegaard and A. Tarantola, "Probabilistic approach to inverse problems," in *International Handbook of Earthquake and Engineering Seismology, Part A*, ser. International Geophysics. Academics Press Ltd, 2002, vol. 81, pp. 237–265.

[19] E. T. Jaynes, "Prior probabilities," *Systems Science and Cybernetics, IEEE Transactions on*, vol. 4, no. 3, pp. 227–241, 1968.

[20] M. Arumlampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[21] A. Doucet, N. De Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo methods in practice*, A. Doucet, N. De Freitas, and N. Gordon, Eds. Springer, 2001, pp. 3–14.