INL/JOU-23-73256-Revision-0

Decision Making Based on Markov Decision Process in Integrated Artificial Reasoning Framework - Part I: Theory

August 2025

Junyung Kim

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/JOU-23-73256-Revision-0

Decision Making Based on Markov Decision Process in Integrated Artificial Reasoning Framework - Part I: Theory

Junyung Kim

August 2025

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

RESEARCH ARTICLE

Decision-making based on Markov decision process in integrated artificial reasoning framework—Part I: Theory [version 1; peer review: 2 approved with reservations]

Junyung Kim[1,2], Xinyan Wang[3], Kyle Warns[2], Xingang Zhao[4], Birdy Phathanapirom[4], Michael W. Golay[3], Hyun Gook Kang[2]

[1]Idaho National Laboratory, Idaho Falls, Idaho, USA
[2]Rensselaer Polytechnic Institute, Troy, New York, USA
[3]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[4]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

v1 **First published:** 21 Aug 2024, **2**:64
https://doi.org/10.12688/nuclscitechnolopenres.17491.1
Latest published: 21 Aug 2024, **2**:64
https://doi.org/10.12688/nuclscitechnolopenres.17491.1

Abstract

This paper presents a decision-making framework based on an integrated artificial reasoning framework and Markov decision process (MDP). The integrated artificial reasoning framework provides a physics-based approach that converts system information into state transition models, and the analysis result will be represented by the transition probabilities that can be used with an MDP to find a traceable and explainable optimal pathway. A dynamic Bayesian network (DBN) is well suited for representing the structure of an MDP. The causality information among process variables (or among subsystems) is mathematically represented in a DBN by the conditional probabilities of the node's states provided different probabilities of the parent node's states. To define node states in a physically understandable manner, we used multilevel flow modeling (MFM). An MFM follows the fundamental energy and mass conservation laws and supports the selection of process variables that represent the system of interest so that causal relations among process variables are properly captured. An MFM-based DBN supports developing state transition models in an MDP to capture the effect of process variables of system having physical relations. The operators of the target system can capture stochastic system dynamics as multiple subsystem state transitions based on their physical relations and uncertainties coming from component degradation or random failures. We analyzed a simplified exemplary system to illustrate an optimal operational policy using the suggested approach.

Open Peer Review

Approval Status ? ?

	1	2
version 1 21 Aug 2024	? view	? view

1. **Muhammad Zubair**, University of Sharjah, Sharjah, United Arab Emirates

2. **Elvan Sahin**, Virginia Polytechnic Institute and State University, Blacksburg, USA

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
reinforcement learning, Markov decision process, decision-making, dynamic Bayesian network, multilevel flow modeling

Corresponding author: Hyun Gook Kang (kangh6@rpi.edu)

Author roles: Kim J: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Wang X**: Software, Validation, Writing – Review & Editing; **Warns K**: Validation, Writing – Review & Editing; **Zhao X**: Supervision, Validation, Writing – Review & Editing; **Phathanapirom B**: Supervision, Validation, Writing – Review & Editing; **Golay MW**: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing; **Kang HG**: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work is supported by and performed in conjunction with the U.S. Department of Energy Federal Grant number DE-NE0008873. This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes. The author Kyle Warns is supported by the Nuclear Regulatory Commission Fellowship Program under the grants 31310018M0003 and 31310023M0032.

How to cite this article: Kim J, Wang X, Warns K *et al.* **Decision-making based on Markov decision process in integrated artificial reasoning framework—Part I: Theory [version 1; peer review: 2 approved with reservations]** Nuclear Science and Technology Open Research 2024, **2**:64 https://doi.org/10.12688/nuclscitechnolopenres.17491.1

First published: 21 Aug 2024, **2**:64 https://doi.org/10.12688/nuclscitechnolopenres.17491.1

Abbreviations

AI: artificial intelligence
BN: Bayesian network
BoP: balance of plant
DBN: dynamic Bayesian network
EFS: energy flow structure
FW: feedwater
HX: heat exchanger
IARF: integrated artificial reasoning framework
IFS: information flow structure
MDP: Markov decision process
MFM: multilevel flow modeling
MFS: mass flow structure
ML: machine learning
RL: reinforcement learning

1. Introduction

Making operational decisions for large systems like nuclear power plants is a challenging job because of the complexity and interdependent nature of their systems. There are many factors to consider, including system information and uncertainties from diverse factors. Moreover, the decisions made in such systems often have significant consequences, both positive and negative. For example, a decision to shut down a nuclear reactor may be necessary to prevent an accident, but it could also have significant economic consequences, as well as impacting the energy supply to the surrounding area. Therefore, decision-makers need to use advanced techniques, such as modeling and simulation, to understand the potential outcomes of different decisions and to identify the best course of action. To make informed decisions, it is necessary to understand the system's nature, as well as the timing and sequencing of events. This timing is essential because it can lead to different system trajectories depending on the plant's detailed states, resulting in many possible scenarios that must be considered. To address these challenges, decision-making support tools are needed that can handle all available information about the system and uncertainties that come with complex phenomena. Many traditional approaches for solving decision-making problems involve using hand-crafted heuristics that sequentially construct a solution.[1] Such heuristics are designed by domain experts and are often suboptimal due to the complex nature of the problems.

Machine learning (ML) algorithms are widely used in artificial intelligence to manage complex systems with numerous process variables and connections and to provide operational decision-making support.[2] However, the black box nature of most ML algorithms makes it challenging to understand the artificial reasoning behind their decisions. This lack of explainability has limited the broader application of ML in many large control problems, including supervisory or fully autonomous control systems.[3,4] In the heavily regulated industry such as nuclear power, explainability is particularly crucial for supervisory systems, as operators need to understand why an algorithm recommends a specific decision before taking any action.

Reinforcement learning (RL) is a subfield of ML which can propose a viable alternative to automate the search of the hand-crafted heuristics by training an agent in a supervised or self-supervised manner. Specifically, model-based RL employs the state transition and reward models of any given state-action pair to mathematically formulate the optimization problem as a Markov decision process (MDP).[5] As a result of its training and decision-making based on transition and reward models, optimal solutions can be verified by tracing back state transitions and associated rewards.

One way to construct state transition models is by using the cell-to-cell mapping approach.[6] The objective of cell-to-cell mapping is to divide a system into multiple state cells, with each cell representing a state entity. Probabilistic mapping of the discretized system space in discrete time allows for the quantification of the probabilistic system evolution over time, as well as tracking fault propagation throughout the system.[7] The number of state cells can be managed by space discretization methods, such as equal width discretization[8] and data-driven discretization.[9]

The cell-to-cell mapping technique can be extended to dynamic Bayesian network (DBN) if causality information among process variables is considered.[9] A DBN, which relates variables to each other over adjacent time steps, is a probabilistic graphical model for representing domain knowledge where each node corresponds to process variables and each edge represents causality among nodes.[10] The causality information is represented mathematically by the conditional probability of the node's state provided different probabilities of the parent node's state. Therefore, it is crucial to establish a clear and comprehensible state definition with relevant system process variables so that the state

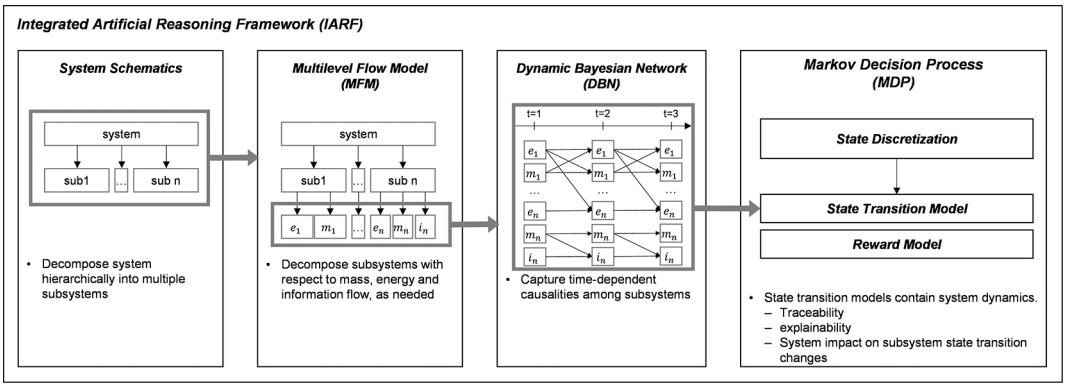

Figure 1. Flowchart of decision-making process in IARF, where e, m, and i in the DBN denote energy, mass, and information nodes, respectively.

transition can be explained physically and state-space explosion due to increased number of process variables can be avoided.

In previous studies, authors proposed an integrated artificial reasoning framework (IARF),[11–13] which is a physics-based approach of developing the DBN structure. This study presents an extended application of IARF in decision-making. Figure 1 shows the system modeling and decision-making process flowchart. First, the system is decomposed into multiple subsystems from descriptive system information (or system schematics). Each subsystem is further decomposed with respect to mass, energy, and information flow as needed. Next, a multilevel flow modeling (MFM) based DBN is formulated to perform a mathematical and graphical system modeling. The state in each DBN node is defined by using a state discretization method, and state transition models are developed based on DBN and state discretization results. Lastly, model-based RL is utilized to find optimal solutions maximizing (or minimizing) an objective function. Since reward modeling is dependent on the problem domain, this research focuses on state transition modeling and decision-making in IARF.

In this paper, Section 2 introduces IARF and MDP fundamentals, Section 3 demonstrates the proposed approach through an optimal decision-making problem with a simplified balance of plant (BoP) considering component degradation, and Section 4 concludes the study.

2. Integrated artificial reasoning for decision-making

2.1 Framework overview

There are three steps for system modeling and one step for decision process in the IARF. A hierarchical decomposition of the system in a top-down fashion is the first step. The system is divided into subsystems as needed. MFM modeling[14] is the next step and is a qualitative modeling and reasoning technique to represent system characteristics and phenomena. An MFM follows the fundamental energy and mass conservation laws and supports the selection of process variables that represent the system of interest so that causal relations among process variables are properly captured. Applying the MFM technique, one can decompose a system into several mass, energy, and information flow structures, and each flow structure of the MFM model can be modeled as a node in a DBN.[12] After converting the DBN to an MDP and solving it, the optimal solutions found can be explained by the relationships among nodes in the DBN, which were connected based on physical laws and reasoning. An example of MFM-based DBN structure design can be found in.[12]

Figure 2 shows how a heat exchanger (HX) system in the BoP is modeled in IARF. From the system information in Figure 2(a), one can develop an MFM model, as shown in Figure 2(b), where the tube and shell sides of the HX are further decomposed with respect to energy flow structure (EFS) and mass flow structure (MFS) following the MFM modeling syntax. In each flow structure, the basic flow functions[15] used for modeling are *storage* (○), *transport* (⌖), source (⊙), and sink (⊗). Each flow function represents a process variable of the system or physical phenomenon. Table 1 shows process variables and their representative MFM flow functions used for the HX modeling. For instance, tra1.m represents mass transport inside the HX tube and st2.m represents the HX water level (shell side). In the model development, the following thermohydraulic phenomena have been captured in the relationships between the HX EFS and MFS:

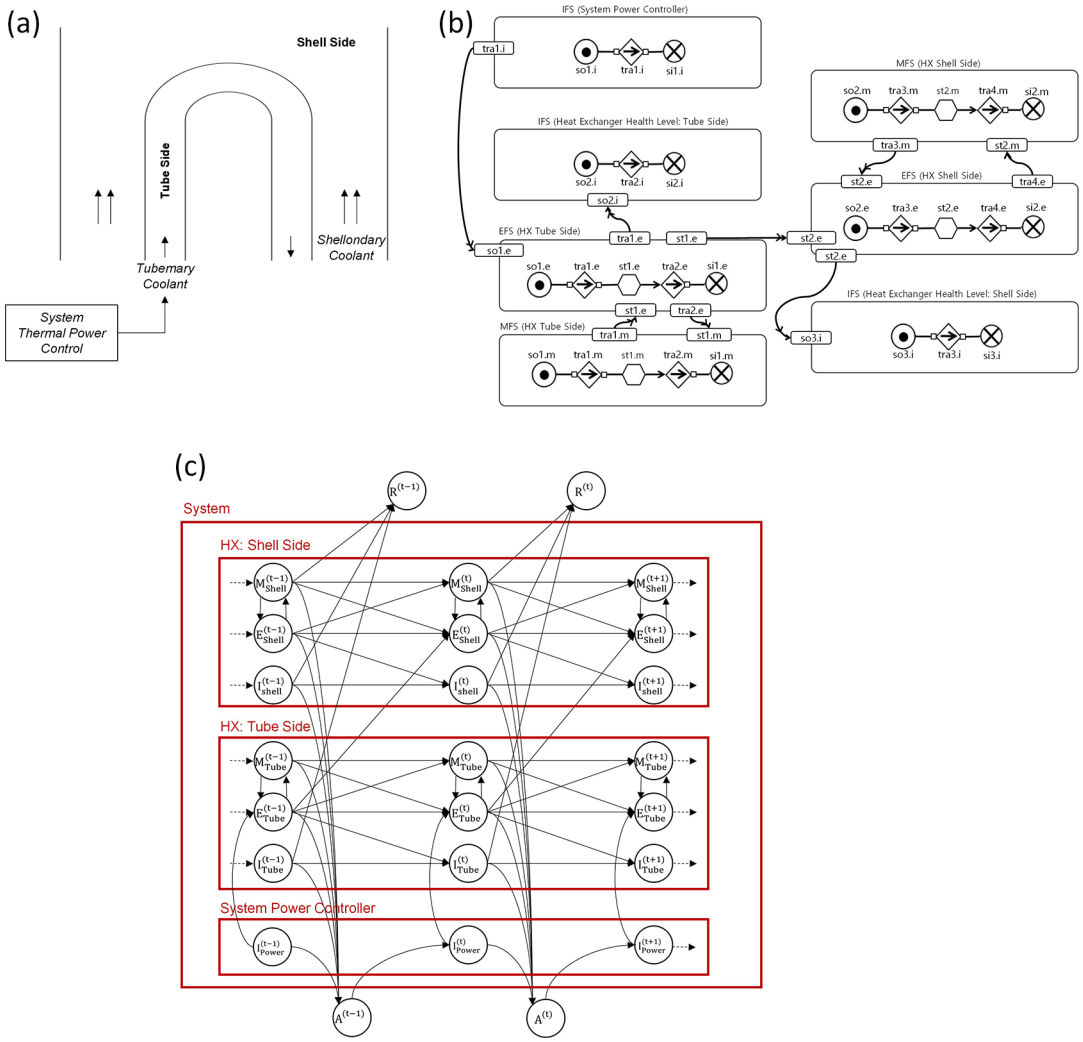

Figure 2. From system schematic to static MFM and MDP model structure. In (a) the physical diagram of the system is given. In (b) the MFM qualitative causal reasoning decomposition of the system is depicted. In (c) the MDP structure of the system is given. R, E, M, I, and A denote reward, energy, mass, information, and action, respectively.

Table 1. MFM flow structures and functions used for an HX model.

MFS – HX (Tube Side)						
	MFM flow function	**so1.m**	**tra1.m**	**st1.m**	**tra2.m**	**si1.m**
	Representing variables	HX inlet (tube side) mass flow	Mass transport	Amount of coolant inside the tube	Mass transport	HX outlet (tube side) mass flow
EFS – HX (Tube Side)						
	MFM flow function	**so1.e**	**tra1.e**	**st1.e**	**tra2.e**	**si1.e**
	Representing variables	HX inlet (tube side) temperature, pressure	Energy transport	Enthalpy of coolant inside HX tube	Energy transport	HX outlet (tube side) temperature, pressure

Table 1. *Continued*

MFS – HX (Shell Side)						
	MFM flow function	**so2.m**	**tra3.m**	**st2.m**	**tra4.m**	**si2.m**
	Representing variables	HX inlet (shell side) mass flow	Mass transport	HX water level	Mass transport	HX outlet (shell side) mass flow
EFS – HX (Shell Side)						
	MFM flow function	**so2.e**	**tra3.e**	**st2.e**	**tra4.e**	**si2.e**
	Representing variables	HX inlet (shell side) temperature, pressure	Energy transport	Enthalpy of coolant inside HX shell	Energy transport	HX outlet (shell side) temperature, pressure
IFS – System Power Control						
	MFM flow function	**so1.i**	**tra1.i**	**si1.i**		
	Representing Variables	Initial power controller status	Power changes	Terminal power controller status		
IFS – Health Status of HX (Tube Side)						
	MFM flow function	**so2.i**	**tra2.i**	**si2.i**		
	Representing Variables	Initial structural deformation level	Structure deformation	Terminal structural deformation level		
IFS – Health Status of HX (Shell Side)						
	MFM flow function	**so3.i**	**tra3.i**	**si3.i**		
	Representing Variables	Initial structural deformation level	Structure deformation	Terminal structural deformation level		

- Energy gives an effect to a mass (e.g., high pressure increases mass flow rate).
- Mass gives an effect to an energy (e.g., amount of steam influences pressure).
- Energy gives an effect to the HX health level (e.g., thermal deformation).
- The HX health level affects neither energy nor mass.

In addition to energy and mass flow inside the system, we can consider system control and component health information, which is modeled as information flow structures (IFS) in the MFM. For instance, when system operators increase the system power level (tra1.i ↑), it affects the inlet temperature and pressure of the HX tube side (so1.e ↑) and consecutively the outlet temperature and pressure of the HX tube side (si1.e ↑). Once the temperature and pressure of the HX (tube side) increases (so1.e ↑), the HX health level will get a negative impact (tra1.i ↓). In principle, the health information of every component in the system can be modeled.

The DBN's structure, as depicted in Figure 2(c), was modeled in accordance with the causal information among flow structures in the MFM. System power controller information nodes connect to the energy nodes of the HX tube side. Additionally, nodes for mass flow in the HX shell side and nodes for HX health information are connected to reward nodes considering rewards from electricity generation and costs of a plant shutdown caused by severe HX structure deformation.

2.2 Markov Decision Process with Dynamic Bayesian Network

DBN is well suited for representing the structure of an MDP.[16] Generally, there are three types of DBN nodes at each time step in an MDP: reward ($R^{(t)}$), system ($S^{(t)}$), and action ($A^{(t)}$), where *t* represents time step. In this research, we propose splitting $S^{(t)}$ into nodes representing system process variables (i.e., $P^{(t)}$) and nodes for components' statuses (or operational mode) (i.e., $C^{(t)}$), where status can affect process variables. Component statuses were given individual nodes so that component failure under certain actions could be explicitly considered in the modeling. Figure 3 shows the structure of an MDP model in the form of a DBN. The action ($A^{(t)}$) taken at the current time step will change the state of the component status node in the next time step, whereas the component status nodes affect the process variable nodes in their own time step (i.e., there is no temporal delay between control input and process variable change).

The MDP is based on the set of Bellman equations,[17] describing the following process [1]:

1) State transition is modeled as a conditional probability.

2) The conditional probability distribution of a future state, given both past and present states, is solely dependent on the present state (i.e., Markov property).

3) An action ($a^{(t)}$) is taken based on monitoring the state of process variables ($p^{(t)}$) and component statuses ($c^{(t)}$).

4) An action ($a^{(t)}$) affects the state of the component status at the next time step ($c^{(t+1)}$).

5) The reward ($r^{(t)}$) of each state is determined by the state of process variables ($p^{(t)}$) and component statuses ($c^{(t)}$).

6) Gain ($g^{(t)}$) is the summation of rewards with a discount factor, γ:

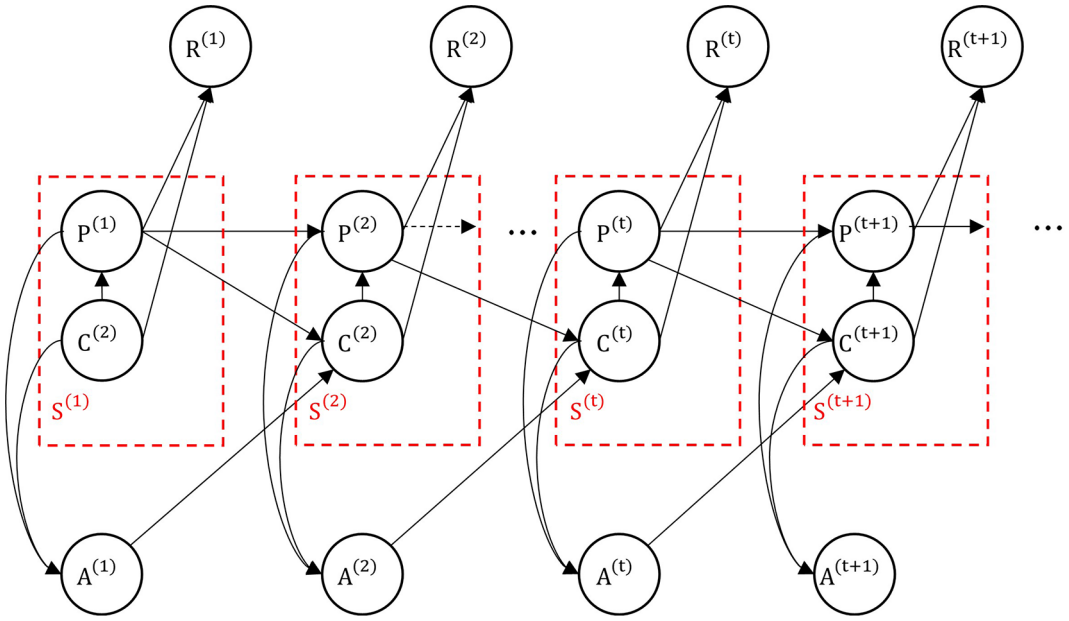

Figure 3. Structure of an MDP model as a DBN, where *R*, *P*, *C*, and *A* denote state of reward, state of process variable, state of components' status (or operational mode), and state of actions, respectively; the super-scripts are for time points; and the dotted red line represents system state.

[1]An uppercase letter represents a random variable or function. A lowercase letter is for a realization of a random variable or function value. This convention also applies to state variables in other equations.

$$G^{(t)} = R^{(t)} + \gamma \cdot R^{(t+1)} + \gamma^2 \cdot R^{(t+2)} \ldots = R^{(t)} + \gamma \cdot G^{(t+1)} \tag{1}$$

7) The expected gain, called the state value ($v\left(S^{(t)}\right)$), is:

$$\begin{aligned} v\left(S^{(t)}\right) &= E\left[G^{(t)}|S^{(t)} = \left\{P^{(t)}, C^{(t)}\right\}\right] = E\left[R^{(t)} + \gamma \cdot G^{(t+1)}|S^{(t)} = \left\{P^{(t)}, C^{(t)}\right\}\right] \\ &= E\left[R^{(t)}|S^{(t)} = \left\{P^{(t)}, C^{(t)}\right\}\right] + \gamma \cdot E\left[G^{(t+1)}|S^{(t)} = \left\{P^{(t)}, C^{(t)}\right\}\right] \end{aligned} \tag{2}$$

which means that the state value $v\left(S^{(t)}\right)$ equals the expected reward, $E\left[R^{(t)}|P^{(t)}, C^{(t)}\right]$, plus the expected gain of reaching the successor states, $E\left[G^{(t+1)}|S^{(t)} = \{P^{(t)}, C^{(t)}\}\right]$, with a discount factor, γ. The first term of Equation (2) can be further expanded after applying the sum rule (Eq. (3)), product rule (Eq. (4)), and Markovian property (Eq. (5)):

$$E\left[R^{(t)}|S^{(t)} = \left\{P^{(t)}, C^{(t)}\right\}\right] = \sum_{r^{(t)}} r^{(t)} \cdot \Pr\left(r^{(t)}|p^{(t)}, c^{(t)}\right) = \sum_{p^{(t+1)}} \sum_{r^{(t)}} \sum_{a^{(t)}} \sum_{c^{(t+1)}} r^{(t)} \cdot \Pr\left(p^{(t+1)}, r^{(t)}, a^{(t)}, c^{(t+1)}|p^{(t)}, c^{(t)}\right) \tag{3}$$

$$= \sum_{p^{(t+1)}} \sum_{r^{(t)}} \sum_{a^{(t)}} \sum_{c^{(t+1)}} r^{(t)} \cdot \Pr\left(a^{(t)}|p^{(t)}, c^{(t)}\right) \cdot \Pr\left(c^{(t+1)}|a^{(t)}, p^{(t)}, c^{(t)}\right) \cdot \Pr\left(p^{(t+1)}, r^{(t)}|c^{(t+1)}, p^{(t)}, c^{(t)}\right) \tag{4}$$

$$= \sum_{p^{(t+1)}} \sum_{r^{(t)}} \sum_{a^{(t)}} \sum_{c^{(t+1)}} r^{(t)} \cdot \Pr\left(a^{(t)}|p^{(t)}, c^{(t)}\right) \cdot \Pr\left(c^{(t+1)}|a^{(t)}, p^{(t)}, c^{(t)}\right) \cdot \Pr\left(p^{(t+1)}, r^{(t)}|c^{(t+1)}, p^{(t)}\right) \tag{5}$$

Similarly, the second term of Equation (2) can be expanded after applying the definition of expectation (Eq. (6)), sum rule (Eq. (7)), product rule (Eq. (8) and Eq. (9)), and the definition of state value (Eq. (10) and Eq. (11)):

$$E\left[G^{(t+1)}|S^{(t)} = P^{(t)}, C^{(t)}\right] = \sum_{g^{(t+1)}} g^{(t+1)} \cdot \Pr\left(g^{(t+1)}|p^{(t)}, c^{(t)}\right) \tag{6}$$

$$= \sum_{p^{(t+1)}} \sum_{r^{(t)}} \sum_{a^{(t)}} \sum_{c^{(t+1)}} \sum_{g^{(t+1)}} g^{(t+1)} \cdot \Pr\left(p^{(t+1)}, r^{(t)}, a^{(t)}, c^{(t+1)}, g^{(t+1)}|p^{(t)}, c^{(t)}\right) \tag{7}$$

$$= \sum_{p^{(t+1)}} \sum_{r^{(t)}} \sum_{a^{(t)}} \sum_{c^{(t+1)}} \sum_{g^{(t+1)}} g^{(t+1)} \cdot \Pr\left(p^{(t+1)}, r^{(t)}, a^{(t)}, c^{(t+1)}|p^{(t)}, c^{(t)}\right) \cdot \Pr\left(g^{(t+1)}|p^{(t)}, c^{(t)}, p^{(t+1)}, r^{(t)}, a^{(t)}, c^{(t+1)}\right) \tag{8}$$

$$\begin{aligned} &= \sum_{p^{(t+1)}} \sum_{r^{(t)}} \sum_{a^{(t)}} \sum_{c^{(t+1)}} \sum_{g^{(t+1)}} g^{(t+1)} \cdot \Pr\left(a^{(t)}|p^{(t)}, c^{(t)}\right) \cdot \Pr\left(c^{(t+1)}|a^{(t)}, p^{(t)}, c^{(t)}\right) \cdot \Pr\left(p^{(t+1)}, r^{(t)}|p^{(t)}, c^{(t+1)}\right) \\ &\quad \cdot \Pr\left(g^{(t+1)}|p^{(t+1)}, c^{(t+1)}\right) \end{aligned} \tag{9}$$

$$\begin{aligned} &= \sum_{a^{(t)}} \Pr\left(a^{(t)}|p^{(t)}, c^{(t)}\right) \sum_{r^{(t)}} \sum_{p^{(t+1)}} \sum_{c^{(t+1)}} \sum_{g^{(t+1)}} g^{(t+1)} \cdot \Pr\left(g^{(t+1)}|p^{(t+1)}, c^{(t+1)}\right) \cdot \Pr\left(c^{(t+1)}|a^{(t)}, p^{(t)}, c^{(t)}\right) \\ &\quad \cdot \Pr\left(p^{(t+1)}, r^{(t)}|p^{(t)}, c^{(t+1)}\right) \end{aligned} \tag{10}$$

$$= \sum_{a^{(t)}} \Pr\left(a^{(t)}|p^{(t)}, c^{(t)}\right) \sum_{r^{(t)}} \sum_{p^{(t+1)}} \sum_{c^{(t+1)}} \Pr\left(c^{(t+1)}|a^{(t)}, p^{(t)}, c^{(t)}\right) \cdot \Pr\left(p^{(t+1)}, r^{(t)}|p^{(t)}, c^{(t+1)}\right) \cdot v\left(S^{(t+1)}\right) \tag{11}$$

which is the probability-weighted sum of the state value of possible future states, $v\left(S^{(t+1)}\right)$. The state value is the immediate reward plus an expected value of risk in the future, where the likelihood of the event (i.e., $p^{(t+1)}$ and $c^{(t+1)}$ happen simultaneously, given $a^{(t)}$, $p^{(t)}$, and $c^{(t)}$) is $\Pr\left(c^{(t+1)}|a^{(t)}, p^{(t)}, c^{(t)}\right) \cdot \Pr\left(p^{(t+1)}, r^{(t)}|p^{(t)}, a^{(t)}, c^{(t+1)}\right)$ and the consequence of event is $v\left(S^{(t+1)}\right)$ (i.e., risk = likelihood × consequence). The summation chain is long because at state $s^{(t)}$ there are several actions ($a^{(t)}$) that could be taken and at each possible $a^{(t)}$ are several possible future states ($s^{(t+1)}$). Therefore:

$$
\begin{aligned}
\nu\left(S^{(t)}\right) &= \sum_{p^{(t+1)}}\sum_{r^{(t)}}\sum_{a^{(t)}}\sum_{c^{(t+1)}} r^{(t)}\cdot \Pr\left(a^{(t)}|p^{(t)},c^{(t)}\right)\cdot \Pr\left(c^{(t+1)}|a^{(t)},p^{(t)},c^{(t)}\right)\cdot \Pr\left(p^{(t+1)},r^{(t)}|c^{(t+1)},p^{(t)}\right)\\
&\quad+\gamma\cdot\sum_{a^{(t)}}\Pr\left(a^{(t)}|p^{(t)},c^{(t)}\right)\sum_{r^{(t)}}\sum_{p^{(t+1)}}\sum_{c^{(t+1)}}\Pr\left(c^{(t+1)}|a^{(t)},p^{(t)},c^{(t)}\right)\cdot \Pr\left(p^{(t+1)},r^{(t)}|p^{(t)},c^{(t+1)}\right)\cdot\nu\left(S^{(t+1)}\right)\\
&= \sum_{a^{(t)}}\Pr\left(a^{(t)}|p^{(t)},c^{(t)}\right)\cdot\sum_{c^{(t+1)}}\sum_{r^{(t)}}\sum_{p^{(t+1)}}\Pr\left(c^{(t+1)}|a^{(t)},p^{(t)},c^{(t)}\right)\cdot \Pr\left(p^{(t+1)},r^{(t)}|c^{(t+1)},p^{(t)}\right)\cdot\left(r^{(t)}+\gamma\cdot\nu\left(S^{(t+1)}\right)\right)\\
&= r^{(t)}+\gamma\cdot\sum_{a^{(t)}}\Pr\left(a^{(t)}|p^{(t)},c^{(t)}\right)\cdot\sum_{c^{(t+1)}}\sum_{p^{(t+1)}}\Pr\left(c^{(t+1)}|a^{(t)},p^{(t)},c^{(t)}\right)\cdot \Pr\left(p^{(t+1)}|c^{(t+1)},p^{(t)}\right)\cdot\nu\left(S^{(t+1)}\right)
\end{aligned}
\tag{12}
$$

The goal of solving the Bellman equation is to find an optimal operational policy, denoted by π, which comprises a sequence of actions from the set of actions (A), that provides the maximum state values at each time step:

$$
\nu_\pi\left(S^{(t)}\right) = r^{(t)}+\gamma\cdot\max_a\cdot\sum_{c^{(t+1)}}\sum_{p^{(t+1)}}\Pr\left(c^{(t+1)}|a^{(t)},p^{(t)},c^{(t)}\right)\cdot\Pr\left(p^{(t+1)}|p^{(t)},c^{(t+1)}\right)\cdot\nu_\pi\left(S^{(t+1)}\right) \tag{13}
$$

Since $\nu_\pi\left(S^{(t)}\right)$ depends on $\nu_\pi\left(S^{(t+1)}\right)$, the MDP agent will iterate over all the states and actions to solve the equation.

In this research, we derived the Bellman equation in a form that includes the concept of functional decomposition of the system into multiple subsystems. The conditional probability distribution of the system state in Eq. (13) (i.e., $\Pr\left(p^{(t+1)}|p^{(t)},c^{(t)}\right)$) extends to the multiplication of multiple conditional probability distributions of subsystems. Causality information among subsystems captured in a DBN is represented when one calculates the state value using the Bellman equation. Eq. (17) shows how the DBN in Figure 2(c) is represented as a Bellman equation, and a derivation of Eq. (17) promptly follows.

$$
\begin{aligned}
&\nu_\pi\left(P^{(t)}=\{M^{(t)}_{Tube},E^{(t)}_{Tube},I^{(t)}_{Tube},M^{(t)}_{Shell},E^{(t)}_{Shell},I^{(t)}_{Shell}\},C^{(t)}=\{I^{(t)}_{Power}\}\right)\\
&=r^{(t)}+\gamma\cdot\max_a\sum_{i^{(t+1)}_{power}}\Pr\left(i^{(t+1)}_{Power}|a^{(t)},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},i^{(t)}_{Power}\right)\\
&\quad\cdot\sum_{m^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Shell}}\sum_{i^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Tube}}\sum_{m^{(t+1)}_{Tube}}\sum_{i^{(t+1)}_{Tube}}\Pr\left(m^{(t+1)}_{Shell},e^{(t+1)}_{Shell},i^{(t+1)}_{Shell},m^{(t+1)}_{Tube},e^{(t+1)}_{Tube},i^{(t+1)}_{Tube}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power}\right)\\
&\quad\cdot\nu_\pi\left(S^{(t+1)}\right)
\end{aligned}
\tag{14}
$$

$$
\begin{aligned}
&=r^{(t)}+\gamma\cdot\max_a\sum_{i^{(t+1)}_{Power}}\Pr\left(i^{(t+1)}_{Power}|a^{(t)},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},i^{(t)}_{Power}\right)\\
&\quad\cdot\sum_{m^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Shell}}\sum_{i^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Tube}}\sum_{m^{(t+1)}_{Tube}}\sum_{i^{(t+1)}_{Tube}}\Pr\left(m^{(t+1)}_{Tube},e^{(t+1)}_{Tube},i^{(t+1)}_{Tube}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power}\right)\\
&\quad\cdot\Pr\left(m^{(t+1)}_{Shell},e^{(t+1)}_{Shell},i^{(t+1)}_{Shell}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power},m^{(t+1)}_{Tube},e^{(t+1)}_{Tube},i^{(t+1)}_{Tube}\right)\cdot\nu_\pi\left(S^{(t+1)}\right)
\end{aligned}
\tag{15}
$$

$$
\begin{aligned}
&=r^{(t)}+\gamma\cdot\max_a\sum_{i^{(t+1)}_{Power}}\Pr\left(i^{(t+1)}_{Power}|a^{(t)},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},i^{(t)}_{Power}\right)\\
&\quad\cdot\sum_{m^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Shell}}\sum_{i^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Tube}}\sum_{m^{(t+1)}_{Tube}}\sum_{i^{(t+1)}_{Tube}}\Pr\left(m^{(t+1)}_{Tube},e^{(t+1)}_{Tube}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power}\right)\\
&\quad\cdot\Pr\left(i^{(t+1)}_{Tube}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power},m^{(t+1)}_{Tube},e^{(t+1)}_{Tube}\right)\\
&\quad\cdot\Pr\left(m^{(t+1)}_{Shell},e^{(t+1)}_{Shell}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power},m^{(t+1)}_{Tube},e^{(t+1)}_{Tube},i^{(t+1)}_{Tube}\right)\\
&\quad\cdot\Pr\left(i^{(t+1)}_{Shell}|m^{(t)}_{Shell},e^{(t)}_{Shell},i^{(t)}_{Shell},m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t)}_{Tube},i^{(t)}_{Power},i^{(t+1)}_{Power},m^{(t+1)}_{Tube},e^{(t+1)}_{Tube},i^{(t+1)}_{Tube},m^{(t+1)}_{Shell},e^{(t+1)}_{Shell}\right)\cdot\nu_\pi\left(S^{(t+1)}\right)
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
&=r\left(i^{(t)}_{Tube},i^{(t)}_{Shell},m^{(t)}_{shell}\right)+\gamma\cdot\max_a\sum_{i^{(t+1)}_{Power}}\Pr\left(i^{(t+1)}_{Power}|a^{(t)},i^{(t)}_{Power}\right)\\
&\quad\cdot\sum_{m^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Shell}}\sum_{i^{(t+1)}_{Shell}}\sum_{e^{(t+1)}_{Tube}}\sum_{m^{(t+1)}_{Tube}}\sum_{i^{(t+1)}_{Tube}}\Pr\left(m^{(t+1)}_{Tube},e^{(t+1)}_{Tube}|m^{(t)}_{Tube},e^{(t)}_{Tube},i^{(t+1)}_{Power}\right)\cdot\Pr\left(i^{(t+1)}_{Tube}|i^{(t)}_{Tube},e^{(t)}_{Tube}\right)\\
&\quad\cdot\Pr\left(m^{(t+1)}_{Shell},e^{(t+1)}_{Shell}|m^{(t)}_{Shell},e^{(t)}_{Shell},e^{(t)}_{Tube}\right)\cdot\Pr\left(i^{(t+1)}_{Shell}|e^{(t)}_{Shell},i^{(t)}_{Shell}\right)\cdot\nu_\pi\left(S^{(t+1)}\right)
\end{aligned}
\tag{17}
$$

Conditional probability terms in Eq. (16) were simplified into conditional probability terms in Eq. (17) after applying causal relations between nodes in DBN. Each conditional probability term in Eq. (17) encapsulates the stochastic process of a subsystem. $\Pr\left(m_{Tube}^{(t+1)}, e_{Tube}^{(t+1)} | m_{Tube}^{(t)}, e_{Tube}^{(t)}, i_{Power}^{(t+1)}\right)$ and $\Pr\left(m_{Shell}^{(t+1)}, e_{Shell}^{(t+1)} | m_{Shell}^{(t)}, e_{Shell}^{(t)}, e_{Tube}^{(t)}\right)$ describe the probabilistic transient behaviors of process variables (e.g., pressure, temperature, mass flow rate) in the primary loop and secondary loop, respectively. $\Pr\left(i_{Tube}^{(t+1)} | i_{Tube}^{(t)}, e_{Tube}^{(t)}\right)$ and $\Pr\left(i_{Shell}^{(t+1)} | e_{Shell}^{(t)}, i_{Shell}^{(t)}\right)$ describe state transitions of HX tube and shell side health status given the energy state of tube and shell sides.

By expanding basic state transition models (i.e., $\Pr\left(c^{(t+1)} | a^{(t)}, p^{(t)}, c^{(t)}\right)$ and $\Pr\left(p^{(t+1)} | p^{(t)}, c^{(t+1)}\right)$ in Eq. (13)), the effect of process variables having physical relations can be captured in state transition models. For instance, the component health level, which can be defined by time to failure (t_f), is stochastic process, and the t_f of heat exchanger due to creep can be determined by Eq. (18):

$$t_f = 10^{\frac{LMP}{T} - 20} \tag{18}$$

where LMP is the Larson-Miller parameter[18] of HX that depends on applied stress, material properties, and failure criteria, and T is applied temperature on HX. The physical relations shown in Eq. (18) are then represented in the state transition models (i.e., $\Pr\left(i_{Tube}^{(t+1)} | e_{Tube}^{(t)}, i_{Tube}^{(t)}\right)$ and $\Pr\left(i_{Shell}^{(t+1)} | e_{Shell}^{(t)}, i_{Shell}^{(t)}\right)$) in Eq. (17). Another advantage of the expansion is that one can validate component characteristics that affect overall system performance by substituting one or more component state transition models.

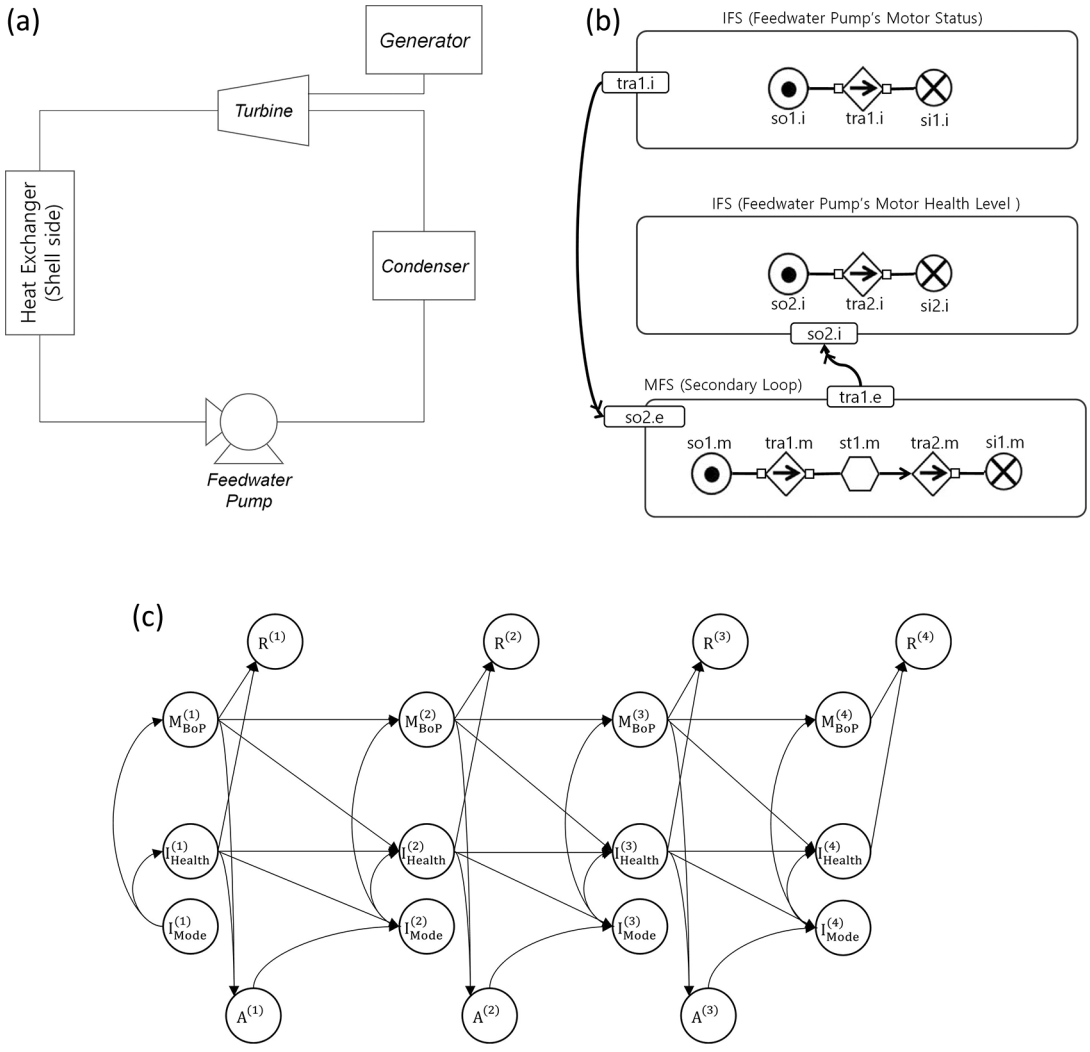

Figure 4. System schematic, static MFM model, and MDP model structure of a simplified BoP. In (c), R, M, I, and A denote reward, mass, information, and action, respectively.

3. Decision-making Example: Power generation optimization with component failure risk

This chapter presents a simplified power plant BoP in which both the health state of components and process variables of the system change over time. Figure 4 is a system schematic of the simplified BoP. The main feedwater (FW) pump provides coolant to the HX (shell side), and operators can maneuver the FW pump to control the mass flow of steam coming out of the HX, which will affect the amount of electricity that the steam turbine generates. The plant is initially operating in a steady-state condition and will continue to operate for four time steps for an illustrative purpose. During operation, degradation occurs in the pump-driving electric motor, meaning the health state of the motor will change over time, leading to different sudden failure likelihoods of the FW pump. At each time step, one decides to operate the system at either a higher power to prioritize electricity generation or a lower power to minimize the likelihood of failure. The goal, as it is for any power plant, is to maximize the amount of profit generated from electricity production while also maintaining safe operations without a high likelihood of motor failure, which will lead to plant shutdown. The MDP model was developed with the following assumptions:

- There are two system operational modes: 100% and 80% rated electric power.
- The health state of the motor in the FW pump can change over time.
- Motor degradation depends on the power level: the higher the power is, the faster the degradation occurs.
- The likelihood of the pump having a locked rotor depends on the health state of the FW pump. A gradual performance degradation of the FW pump is not considered to simplify prognostics.
- The component status of the next time step is dependent on the action and health state of the current time step.
- Three health states of the motor are assumed: healthy (State 1), in danger (State 2), and failed (State 3).

Eq. (19) is the corresponding Bellman equation for Figure 4. The likelihood of the FW pump locked rotor given the current action and motor health state (i.e., $\Pr\left(c_{\text{Mot}}^{(t+1)}|a^{(t)},i_{\text{Mot}}^{(t)}\right)$) and prognostics of the motor health (i.e., $\Pr\left(i_{\text{Mot}}^{(t+1)}|m_{\text{BoP}}^{(t)},i_{\text{Mot}}^{(t)},c_{\text{Mot}}^{(t+1)}\right)$) are encapsulated in Eq. (19). Conditional probabilities used in Eq. (19) are given in Table 2. A key to the state definitions is given in Table 3.

$$\begin{aligned}v_\pi\left(S^{(t)}\right) = {} & r\left(m_{\text{BoP}}^{(t)},i_{\text{Health}}^{(t)}\right)+\gamma\cdot\max_a\sum_{i_{\text{Mode}}^{(t+1)}}\Pr\left(i_{\text{Mode}}^{(t+1)}|a^{(t)},i_{\text{Health}}^{(t)}\right)\cdot\sum_{m_{\text{BoP}}^{(t+1)}}\sum_{i_{\text{Health}}^{(t+1)}}\Pr\left(i_{\text{Health}}^{(t+1)}|m_{\text{BoP}}^{(t)},i_{\text{Health}}^{(t)},i_{\text{Health}}^{(t+1)}\right)\\ & \cdot\Pr\left(m_{\text{BoP}}^{(t+1)}|m_{\text{BoP}}^{(t)},i_{\text{Mode}}^{(t+1)}\right)\cdot v_\pi\left(S^{(t+1)}\right)\end{aligned} \tag{19}$$

where $r\left(m_{\text{BoP}}^{(t)},i_{\text{Health}}^{(t)}\right)$ is the reward function determined by the mass flow of the BoP $\left(m_{\text{BoP}}^{(t)}\right)$ and the motor health state $\left(i_{\text{Health}}^{(t)}\right)$, as shown in Eq. (20)–(22):

$$r\left(m_{\text{BoP}}^{(t)},i_{\text{Health}}^{(t)}\right)=f\left(m_{\text{BoP}}^{(t)}\right)+g\left(i_{\text{Health}}^{(t)}\right) \tag{20}$$

$$f\left(m_{\text{BoP}}^{(t)}\right)=\begin{cases}1 & \text{if } m_{\text{BoP}}^{(t)}=\text{state 1 (BoP massflow rate with 100\%power)}\\ 0.8 & \text{if } m_{\text{BoP}}^{(t)}=\text{state 2(BoP massflow rate with 80\%power)}\\ 0 & \text{if } m_{\text{BoP}}^{(t)}=\text{state 3(BoP massflow rate when motor fails)}\end{cases} \tag{21}$$

$$g\left(i_{\text{Health}}^{(t)}\right)=\begin{cases}0 & \text{if } i_{\text{Health}}^{(t)}=\text{state 1 (Healthy)}\\ 0 & \text{if } i_{\text{Health}}^{(t)}=\text{state 2 (In Danger)}\\ -100 & \text{if } i_{\text{Health}}^{(t)}=\text{state 3 (Failed)}\end{cases} \tag{22}$$

where $f\left(m_{\text{BoP}}^{(t)}\right)$ is the reward from electricity generation and $g\left(i_{\text{Health}}^{(t)}\right)$ is the cost due to motor health status. $g\left(i_{\text{Health}}^{(t)}=3\right)$ is set to –100 because motor failure will lead to a plant shutdown, which is very costly. The optimal

Table 2. State transition models with prognostics of a motor in slow degradation.

$\Pr\left(i_{Mode}^{(t+1)} \mid a^{(t)}, i_{Health}^{(t)}\right)$	$\Pr\left(i_{Health}^{(t+1)} \mid i_{Health}^{(t)}, m_{BoP}^{(t)}, i_{Mode}^{(t+1)}\right)$	$\Pr\left(m_{BoP}^{(t+1)} \mid m_{BoP}^{(t)}, i_{Mode}^{(t+1)}\right)$
$t = 1$		
$a^{(1)} = 100\%, 1_{i\_Health}^{(1)}$ $\quad 1_{i\_Mode}^{(2)}\ 2_{i\_Mode}^{(2)}\ 3_{i\_Mode}^{(2)}$ $\begin{bmatrix} 0.99 & 0.00 & 0.01 \\ 0.00 & 0.99 & 0.01 \end{bmatrix}$ $a^{(1)} = 80\%, 1_{i\_Health}^{(1)}$	$1_{i\_Health}^{(1)}, 1_{BoP}^{(1)}, 1_{i\_Mode}^{(2)}$ $1_{i\_Health}^{(1)}, 1_{BoP}^{(1)}, 2_{i\_Mode}^{(2)}$ $1_{i\_Health}^{(1)}, 1_{BoP}^{(1)}, 3_{i\_Mode}^{(2)}$ $1_{i\_Health}^{(2)}\ 2_{i\_Health}^{(2)}\ 3_{i\_Health}^{(2)}$ $\begin{bmatrix} 0.990 & 0.010 & 0.00 \\ 0.999 & 0.001 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$1_{BoP}^{(1)}, 1_{i\_Mode}^{(2)}$ $1_{BoP}^{(1)}, 2_{i\_Mode}^{(2)}$ $1_{BoP}^{(1)}, 3_{i\_Mode}^{(2)}$ $1_{BoP}^{(2)}\ 2_{BoP}^{(2)}\ 3_{BoP}^{(2)}$ $\begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$
t = 2 and 3		
$1_{i\_Mode}^{(t+1)}\ 2_{i\_Mode}^{(t+1)}\ 3_{i\_Mode}^{(t+1)}$ $a^{(t)} = 100\%, 1_{i\_Health}^{(t)}$ $a^{(t)} = 100\%, 2_{i\_Health}^{(t)}$ $a^{(t)} = 100\%, 3_{i\_Health}^{(t)}$ $a^{(t)} = 80\%, 1_{i\_Health}^{(t)}$ $a^{(t)} = 80\%, 2_{i\_Health}^{(t)}$ $a^{(t)} = 80\%, 3_{i\_Health}^{(t)}$ $\begin{bmatrix} 0.99 & 0.00 & 0.01 \\ 0.95 & 0.00 & 0.05 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.99 & 0.01 \\ 0.00 & 0.95 & 0.05 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$1_{i\_Health}^{(t+1)}\ 2_{i\_Health}^{(t+1)}\ 3_{i\_Health}^{(t+1)}$ $1_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $\begin{bmatrix} 0.990 & 0.010 & 0.00 \\ 0.999 & 0.001 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.95 & 0.05 & 0.00 \\ 0.98 & 0.02 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.90 & 0.10 \\ 0.00 & 0.95 & 0.05 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.95 & 0.01 \\ 0.00 & 0.99 & 0.01 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$1_{BoP}^{(t+1)}\ 2_{BoP}^{(t+1)}\ 3_{BoP}^{(t+1)}$ $1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $\begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$

Table 3. State definitions for Eq. (19).

$m_{BoP}^{(t)}$	$i_{Health}^{(t)}$	$i_{Mode}^{(t)}$
state 1:= BoP massflow rate with 100% power	state 1:=Healthy	state 1 := Motor configuration for 100% power
state2:=BoP massflow rate with 80% power	state 2:=In Danger	state2 := Motor configuration with 80% power
state3:=BoP massflow rate with 0% power	state 3:=Failed	state3 := locked rotor

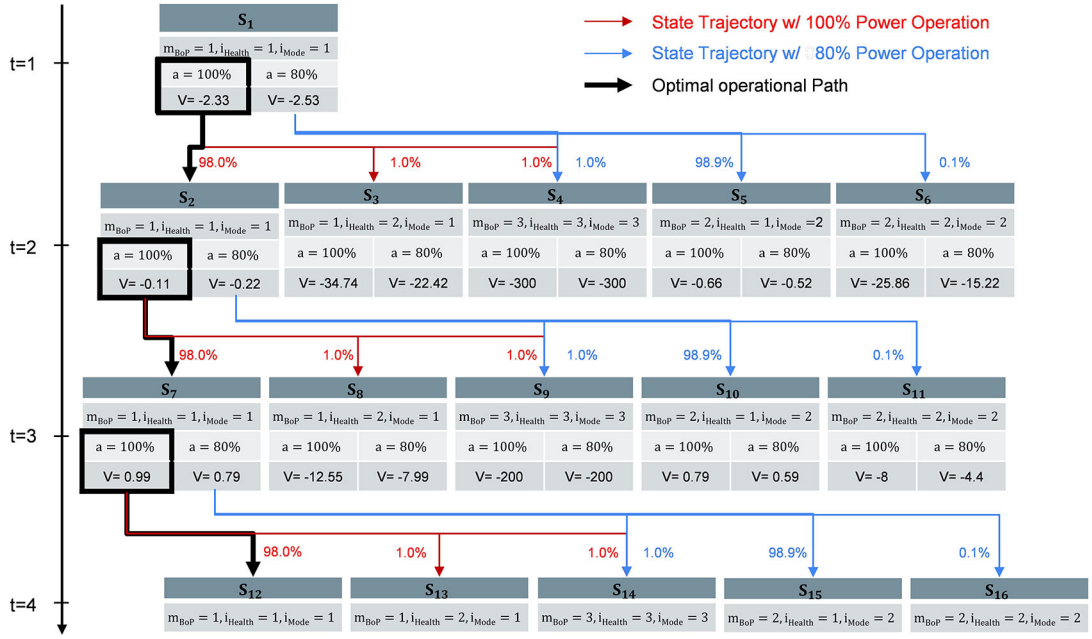

Figure 5. System state flowchart (slow FW pump's motor degradation case), where V is state-action value.

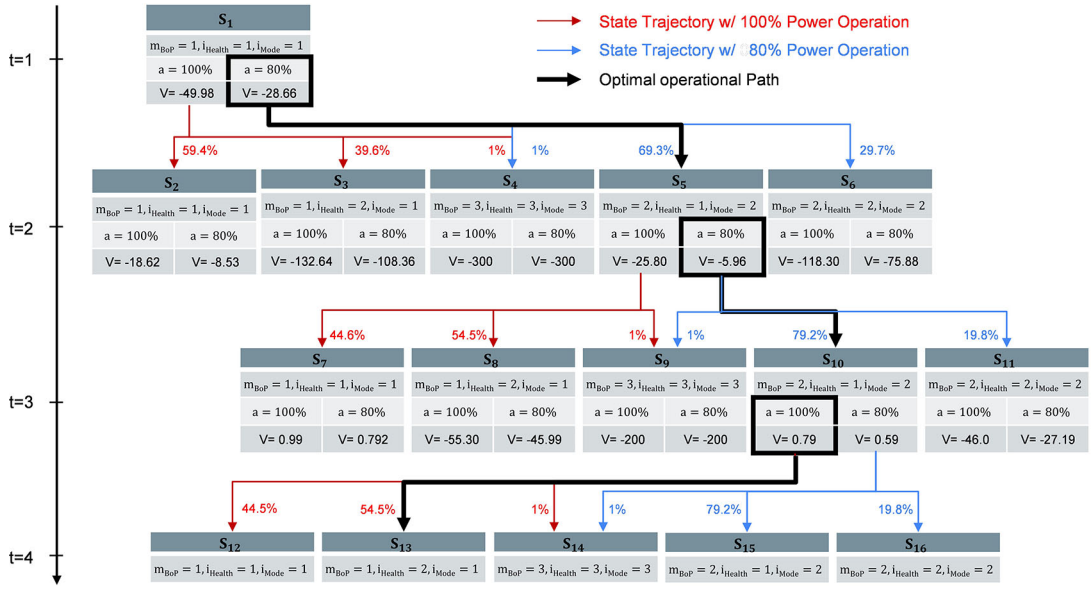

Figure 6. System state flowchart (fast FW pump's motor degradation case). System state flowchart (fast FW pump's motor degradation case). Decisions made are different from Figure 5 due to different component degradation.

Table 4. State transition models with prognostics for a motor in fast degradation.

$\Pr\left(i_{Mode}^{(t+1)} \mid a^{(t)}, i_{Health}^{(t)}\right)$	$\Pr\left(i_{Health}^{(t+1)} \mid i_{Health}^{(t)}, m_{BoP}^{(t)}, i_{Mode}^{(t+1)}\right)$	$\Pr\left(m_{BoP}^{(t+1)} \mid m_{BoP}^{(t)}, i_{Mode}^{(t+1)}\right)$
$t = 1$		
$a^{(1)} = 100\%, 1_{i\_Health}^{(1)}$ $\quad 1_{i\_Mode}^{(2)}\ 2_{i\_Mode}^{(2)}\ 3_{i\_Mode}^{(2)}$ $\begin{bmatrix} 0.99 & 0.00 & 0.01 \\ 0.00 & 0.99 & 0.01 \end{bmatrix}$ $a^{(1)} = 80\%, 1_{i\_Health}^{(1)}$	$1_{i\_Health}^{(2)}\ 2_{i\_Health}^{(2)}\ 3_{i\_Health}^{(2)}$ $1_{i\_Health}^{(1)}, 1_{BoP}^{(1)}, 1_{i\_Mode}^{(2)}$ $1_{i\_Health}^{(1)}, 1_{BoP}^{(1)}, 2_{i\_Mode}^{(2)}$ $1_{i\_Health}^{(1)}, 1_{BoP}^{(1)}, 3_{i\_Mode}^{(2)}$ $\begin{bmatrix} 0.60 & 0.40 & 0.00 \\ 0.70 & 0.30 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$1_{BoP}^{(2)}\ 2_{BoP}^{(2)}\ 3_{BoP}^{(2)}$ $1_{BoP}^{(1)}, 1_{i\_Mode}^{(2)}$ $1_{BoP}^{(1)}, 2_{i\_Mode}^{(2)}$ $1_{BoP}^{(1)}, 3_{i\_Mode}^{(2)}$ $\begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$
t = 2 and 3		
$1_{i\_Mode}^{(t+1)}\ 2_{i\_Mode}^{(t+1)}\ 3_{i\_Mode}^{(t+1)}$ $a^{(t)} = 100\%, 1_{i\_Health}^{(t)}$ $a^{(t)} = 100\%, 2_{i\_Health}^{(t)}$ $a^{(t)} = 100\%, 3_{i\_Health}^{(t)}$ $a^{(t)} = 80\%, 1_{i\_Health}^{(t)}$ $a^{(t)} = 80\%, 2_{i\_Health}^{(t)}$ $a^{(t)} = 80\%, 3_{i\_Health}^{(t)}$ $\begin{bmatrix} 0.99 & 0.00 & 0.01 \\ 0.95 & 0.00 & 0.05 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.99 & 0.01 \\ 0.00 & 0.95 & 0.05 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$1_{i\_Health}^{(t+1)}\ 2_{i\_Health}^{(t+1)}\ 3_{i\_Health}^{(t+1)}$ $1_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{i\_Health}^{(t)}, 3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $\begin{bmatrix} 0.60 & 0.40 & 0.00 \\ 0.70 & 0.30 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.45 & 0.55 & 0.00 \\ 0.80 & 0.20 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.45 & 0.55 \\ 0.00 & 0.55 & 0.45 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.55 & 0.45 \\ 0.00 & 0.75 & 0.25 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$1_{BoP}^{(t+1)}\ 2_{BoP}^{(t+1)}\ 3_{BoP}^{(t+1)}$ $1_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $1_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $1_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $2_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $2_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $2_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $3_{BoP}^{(t)}, 1_{i\_Mode}^{(t+1)}$ $3_{BoP}^{(t)}, 2_{i\_Mode}^{(t+1)}$ $3_{BoP}^{(t)}, 3_{i\_Mode}^{(t+1)}$ $\begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$

action should yield the highest possible sum of the reward value and probability-weighted average of state values of next states.

This formulation constitutes the framing of the problem. Iteratively solving Eq. (19) utilizing the state transition models and reward functions, we developed a state transition flowchart, with part of the flowchart shown in Figure 5. The blocks are different system states that contain the value of each state variable, the possible actions that can be taken from the state, and the state value corresponding to taking each action. Arrows represent actions (100% and 80%) and point to destination states with the corresponding probabilities of reaching these states. The plots do not include all the states but follow only the most probable states selected by optimal actions. The MDP agent chooses actions based on the value of each state, which has been solved by the MDP solver iteratively. For instance, at $t = 0$, the state value could be either –2.33 or –2.53, depending on which action is taken. Because –2.33 > –2.53, the MDP agent will choose a 100% power level. Because of the uncertainty in motor degradation, the state transition from $t = 0$ to $t = 1$ is a stochastic process, whereas in this case, S_2 will be the most likely state after taking a 100% power level operation at $t = 0$. Following same process, the optimal policy becomes {t = 0: 100% Power Operation → t = 1: 100% Power Operation → t = 2: 100% Power Operation → t = 3: 100% Power Operation}.

One advantage of capturing the probabilistic nature of system dynamics with state transition models is that one can easily validate characteristics of a component or subsystem that affect overall performance of the system by replacing one or multiple state transition models. Figure 6 is a flowchart of the system with prognostics of the motor undergoing faster degradation. Corresponding state transition models are listed in Table 4. Because of the higher state transition probability from "healthy" to "in danger"—which implies a fast motor degradation—of $\Pr\left(i_{\text{Health}}^{(t+1)} | i_{\text{Health}}^{(t)}, m_{\text{BoP}}^{(t)}, i_{\text{Mode}}^{(t+1)}\right)$, when the system operates at the 100% power level, the optimal action scenario begins transitioning to 80% power. At $t = 3$, the system can operate at the 100% power level if the motor health state remains in State 1 (i.e., healthy state). Different transition probabilities of $\Pr\left(i_{\text{Health}}^{(t+1)} | i_{\text{Health}}^{(t)}, m_{\text{BoP}}^{(t)}, i_{\text{Mode}}^{(t+1)}\right)$ change the state values and optimal operational actions at each time step.

4. Conclusion

This study extended IARF capabilities in decision-making. In IARF, we used a physics-based approach that converts system information into state transition models for MDP to find traceable and explainable optimal solutions. By extending the conditional probability distribution of the system state transition to the multiplication of multiple conditional probability distributions of subsystems, the effect of process variables having physical relations was encapsulated in state transition models. The key benefits of the approach are:

- It helps process large amounts of system information into state transition models, which can then support explainable decision-making.
- By substituting one or more component state transition models, it aids in verifying the component or subsystem characteristics that impact the overall system performance.

The case study showed the benefits of using the state transition and reward models of MDP knowing component health characteristics. This MDP implementation supports operators making decisions that maximize profit by providing a clear understanding of probabilistic system state trajectories, which was unavailable with conventional control methods based on logical statements. The MDP agent successfully found the optimal scenario that can maximize the electricity generation while maintaining the integrity of system components. More applications of our proposed approach for the nuclear power industry will be presented in Part II[19]: Applications, including a prototype sodium fast reactor model and two real-world operating applications.

Data availability

Zenodo: Code, data, and models used in this and Part II of the work can be obtained at the Repository IARF_Models_Code_Data, https://doi.org/10.5281/zenodo.10680506.[20]

This contains the MDP code utilized in Parts I and II, the resultant data for the case studies in Part II, and the system simulation models used in Part II.

Acknowledgements

We thank Elizabeth Kirby for her assistance in the editing and formatting of this paper and are thankful to Pradeep Ramuhalli and William Gurecky from Oak Ridge National Laboratory for their technical feedback.

References

1. Mazyavkina N, Sviridov S, Ivanov S, *et al.*: **Reinforcement learning for combinatorial optimization: A survey.** *Comput. Oper. Res.* 2021; **134**: 105400.
Publisher Full Text

2. Sarker IH: **Machine learning: Algorithms, real-world applications and research directions.** *SN. Comput. Sci.* 2021; **2**(3): 160.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Antoniadi AM, Du Y, Guendouz Y, *et al.*: **Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review.** *Appl. Sci.* 2021; **11**(11): 5088.
Publisher Full Text

4. Sifakis J, Harel D: **Trustworthy Autonomous System Development.** *Embed. Comput. Syst.* 2023; **22**(3): 1–24.
Publisher Full Text

5. Bellman RE: *Dynamic Programming.* Princeton, NJ, USA: Princeton University Press; 1957.

6. Hsu CS: *Cell-to-cell mapping: a method of global analysis for nonlinear systems.* Springer Science & Business Media; 2013; vol. 64.

7. Aldemir T: **Computer-assisted Markov failure modeling of process control systems.** *IEEE Trans. Reliab.* 1987; **R-36**(1): 133–144.
Publisher Full Text

8. Kotsiantis S, Kanellopoulos D: **Discretization techniques: A recent survey.** *Computer Science and Engineering.* 2006; **32**(1): 47–58.

9. Kim J, Shah AUA, Kang HG: **Dynamic risk assessment with Bayesian network and clustering analysis.** *Reliab. Eng. Syst. Saf.* 2020; **201**: 106959.
Publisher Full Text

10. Yang X-S: *Introduction to algorithms for data mining and machine learning.* Academic Press; 2019.

11. Kim J, Kang HG: **Quantitative Reasoning and Risk Assessment with Dynamic Bayesian Network.** *2020 American Nuclear Society (ANS) Virtual Winter Meeting.* 2000.

12. Kim J, Zhao X, Shah AUA, *et al.*: **System risk quantification and decision making support using functional modeling and dynamic Bayesian network.** *Reliab. Eng. Syst. Saf.* 2021; **215**: 107880.
Publisher Full Text

13. Kim J, Zhao X, Shah AUA, *et al.*: **Physics-Informed Machine Learning-Aided System Space Discretization.** *12th International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC&HMIT 2021).* 2021.

14. Lind M: **An introduction to multilevel flow modeling.** *Journal of Nuclear Safety and Simulation.* 2011; **2**: 22–32.

15. Lind M: **An introduction to multilevel flow modeling.** *Nuclear Safety and Simulation.* 2011; **2**(1): 22–32.

16. Jonsson A, Barto A: **Active learning of dynamic Bayesian networks in Markov decision processes.** *International Symposium on Abstraction, Reformulation, and Approximation.* 2007.

17. Bellman R: **On the theory of dynamic programming.** *Proc. Natl. Acad. Sci.* 1952; **38**: 716–719.
PubMed Abstract | Publisher Full Text | Free Full Text

18. Larson FR, Miller J: **Time-Temperature relationship for rupture and creep stresses.** *Trans. ASME.* 1952; **74**: 765–771.
Publisher Full Text

19. Warns K, Kim J, Wang X, *et al.*: **Decision-making based on Markov decision process in integrated artificial reasoning framework–Part 2: Applications [version 1; peer review: awaiting peer review].** *Nucl. Sci. Technol. Open Res.* 2024; **2**: 50.
Publisher Full Text

20. KyleWarns: KyleWarns/IARF_Models_Code_Data: IARF_Models_Code_Data (1.0). [Dataset]. *Zenodo.* 2024.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 11 November 2024

https://doi.org/10.21956/nuclscitechnolopenres.18766.r27735

?

Elvan Sahin
Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

This manuscript presents a novel theoretical framework that integrates MDPs with IARF, employing DBNs and MFM. The proposed framework addresses challenges in operational decision-making for complex systems by providing traceable, explainable, and optimal decision pathways.

INTRODUCTION:

- The citations seem sparse when discussing "traditional approaches" and ML limitations
- The review of "traditional approaches" (Reference 1) needs expansion with contemporary examples
- The discussion of ML limitations in nuclear applications requires additional citations, particularly regarding: a) Recent developments in explainable AI for critical systems and b) Current applications of ML in nuclear operations
- The transition from general ML to RL could be more technically precise by: a) Explaining why RL is particularly suitable for this application and b) Discussing specific advantages over other ML approaches

SECTION 3: Decision-making Example

- The section describes state transition probabilities but could better explain how these probabilities are derived or validated. Were they derived from empirical data, expert input, or simulations?
- How scalable is the framework when applied to more complex systems with numerous interdependent components? Are there computational limitations?
- Can the framework accommodate real-time updates to state transitions or probabilities based on new data or changing operational conditions?

CONCLUSION:

- While the paper focuses on strengths, it does not explicitly acknowledge limitations. Briefly discuss limitations, such as computational complexity, reliance on accurate probability distributions, or challenges in real-time implementation.

- The benefits are described qualitatively, but their practical impact could be quantified or exemplified. Please include metrics or examples that illustrate how this approach improves decision-making efficiency or system reliability compared to conventional methods.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Probabilistic Risk Analysis (PRA), Bayesian Networks, Markov Chain Analysis, Machine Learning Applications in Nuclear Safety

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 23 September 2024

https://doi.org/10.21956/nuclscitechnolopenres.18766.r27660

?

Muhammad Zubair

University of Sharjah, Sharjah, United Arab Emirates

The main purpose of the paper could be more explicitly stated at the beginning. Consider briefly mentioning the specific problem or challenge the framework addresses. Example: "This paper aims to enhance decision-making in complex systems by integrating artificial reasoning with Markov decision processes."

Some technical terms (e.g., MFM, DBN) may not be familiar to all readers. Consider adding very brief explanations for readers outside your immediate field. For example, you could include a phrase like "Multilevel Flow Modeling (MFM), a method based on energy and mass conservation laws."

It could benefit from a brief mention of key findings or results in the abstract section.
The connection between the problems presented (complexity, uncertainty, consequences) and your proposed solution (ML, RL, MDP, DBN) could be more explicit.

The paper discusses existing challenges, but it would be helpful to clearly indicate how the approach in your paper directly addresses these challenges.

It would be beneficial to highlight the gaps or limitations in the current literature. This will help to better justify the need for your approach.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Safety and Reliability of NPP

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.