MOX-Report No. 41/2025

Active digital twins via active inference

Torzoni, M.; Maisto, D.; Manzoni, A.; Donnarumma, F.; Pezzulo, G.;
Corigliano, A.

Active Digital Twins via Active Inference

Matteo Torzoni[a], Domenico Maisto[b], Andrea Manzoni[c],
Francesco Donnarumma[b], Giovanni Pezzulo[b], Alberto Corigliano[a]

[a]*Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32,
I-20133, Milano, Italy*
[b]*Institute of Cognitive Sciences and Technologies, National Research Council, via S. Martino della Battaglia 44,
I-00185 Roma, Italy*
[c]*MOX – Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133, Milano, Italy*

Abstract

Digital twins are transforming engineering and applied sciences by enabling real-time monitoring, simulation, and predictive analysis of physical systems and processes. However, conventional digital twins rely primarily on passive data assimilation, which limits their adaptability in uncertain and dynamic environments. This paper introduces the *active digital twin* paradigm, based on active inference. Active inference is a neuroscience-inspired, Bayesian framework for probabilistic reasoning and predictive modeling that unifies inference, decision-making, and learning under a unique, free energy minimization objective. By formulating the evolution of the active digital twin as a partially observable Markov decision process, the active inference agent continuously refines its generative model through Bayesian updates and forecasts future states and observations. Decision-making emerges from an optimization process that balances pragmatic exploitation (maximizing goal-directed utility) and epistemic exploration or information gain (actively resolving uncertainty). Actions are dynamically planned to minimize *expected free energy*, which quantifies both the divergence between predicted and preferred future observations, and the epistemic value of expected information gain about hidden states. This approach enables a new level of autonomy and resilience in digital twins, offering superior spontaneous exploration capabilities. The proposed framework is assessed on the health monitoring and predictive maintenance of a railway bridge.

1. Introduction

Over the past decade, the digital twin (DT) paradigm has emerged as a transformative approach for monitoring, control, and decision support, enabling diagnostic and predictive capabilities that surpass those of traditional computational models. As outlined in the 2024 report by the National Academies of Engineering, Science, and Medicine [1], DTs differ from both forward digital models and digital shadows [2]. The former are designed to simulate how input parameters and internal states influence system behavior to generate observable outputs, while the latter focus on data assimilation and model updating. A DT is a tailored virtual representation that captures key attributes of a physical system or process [3]. This digital representation dynamically synchronizes with its physical counterpart by continuously assimilating sensor data and providing predictive capabilities. Specifically, DTs enable the simulation of what-if scenarios, supporting predictive decision-making aimed at maximizing utility. This paper proposes active inference (AIF) [4] as a new paradigm for DTs. By modeling the twin's evolution as a partially observable Markov decision process (POMDP) [5], the AIF agent achieves intelligent automation under the *free-energy principle* [6, 7]. This results in a unified mathematical framework for a new class of *active digital twins* (ADT), equipped with spontaneous exploration capabilities.

Emerging from aeronautical and aerospace engineering [8, 9], DT applications nowadays expand across several domains. These include structural health monitoring and predictive maintenance [10–12], additive manufacturing [13], smart cities [14], energy transition [15], urban sustainability [16], geotechnical engineering [17], subduction zone modeling [18], railway infrastructure management [19], aerial vehicles monitoring

---

and control [20, 21], spacecraft operations in orbit [22], personalized medicine [23, 24], and climate science [25]. Despite the growing interest in DTs, their implementation remains highly customized, typically tailored on the specific application, and often hard to deploy. The need for a widely accepted framework for DTs is therefore increasingly recognized in both research and industry. In [26], Kapteyn et al. proposed an application-agnostic formulation for describing coupled physical-digital systems that evolve dynamically over time and interact via observed data and control inputs. A key contribution of their work is the abstraction of the coupled dynamical system into a generalized representation, which serves as the foundation for a mathematical description of DTs. This abstraction is consistent with agent-based representations in POMDPs, typically formalized using probabilistic graphical models [27].

We introduce ADTs with enhanced exploratory capabilities, employing AIF agents based on discrete generative models to leverage and significantly extend the abstraction of physical-digital systems by Kapteyn et al. [26]. Active inference is a theoretical framework integrating perception, decision-making, and learning within the unified objective of *free energy minimization* [4]. An AIF agent maintains an internal generative model of its environment, continuously updating its beliefs in response to sensory inputs. By minimizing variational free energy, the agent simultaneously fulfills two objectives: reducing the divergence between predicted and preferred future observations, and resolving expected uncertainty about hidden states through action. This dual mechanism naturally balances *exploitation* of existing knowledge to achieve specific goals with *exploration*, i.e., the acquisition of new information. These two imperatives can be referred to using interchangeable terminology. The exploitative or pragmatic behavior is associated with terms such as goal-directed behavior or utility maximization, while the exploratory or epistemic behavior is described using terms such as information seeking, information gain, or uncertainty resolution. The AIF framework has been applied in diverse domains, from neuroscience [28–32] – for modeling decision-making under uncertainty – to reinforcement learning [33, 34], collective behavior [35, 36], and robotics [37–40], demonstrating its versatility in modeling dynamic systems.

The generative model of an AIF agent functions as a self-updating engine that unifies the key aspects underpinning ADTs – namely, data assimilation, state estimation, prediction, planning, and learning – under a Bayesian framework that generalizes across applications. Furthermore, as demonstrated in the following sections, AIF agents naturally provide a mechanism for active information seeking, thereby unlocking the full potential of ADTs. When combined with goal-directed (pragmatic) behavior and possibly enhanced with learning capabilities, this information-seeking (epistemic) drive enables ADTs to engage in spontaneous exploration in response to (potentially critical) uncertainty, ultimately maximizing pragmatic utility.

The limitations of conventional DTs, which are typically restricted to passive observation and open-loop simulation, have already been recognized in [41], although the challenge of actively seeking information to enhance perception or learning remains largely unaddressed. Similar problems have long been studied in fields such as active vision, where perception is not limited to passively acquired images, but involves actively steering the sensing process to resolve uncertainty about the environment [42, 43]. Likewise, robotic systems dynamically adjust their sensing devices to enhance environmental exploration [44, 45], as seen in simultaneous localization and mapping (SLAM) tasks [46]. Similar perception mechanisms are embedded in autonomous driving systems, where view and sensor attention can be dynamically adjusted based on environmental conditions and contextual priorities [47]. In the same spirit, pan-tilt-zoom cameras can actively track objects or events of interest in real time for surveillance purposes [48]. Active digital twins rely on the analogous principle of closing the loop between perception and action, enabling them to autonomously improve situational awareness, refine their internal models through active exploration, and proactively manage the environment evolution.

Compared to alternative approaches for developing DTs, AIF offers remarkable advantages. Unlike reinforcement learning, which relies on trial-and-error exploration (often infeasible in real-world applications) and typically requires extensive datasets, AIF enables ADTs to infer hidden states and optimize future behavior using a compact generative model. Moreover, the free energy minimization imperative of AIF balances information-seeking (epistemic) and goal-directed (pragmatic) behaviors, without the need for manually tuned reward functions or random exploration. These features make AIF particularly well-suited for adaptive and robust ADTs.

We present the ADT paradigm through an application in structural health monitoring and predictive maintenance of engineering structures. Given the potentially high life-cycle costs – economic, social, and safety – associated with such systems, adopting a DT perspective is crucial to enable condition-based or predictive maintenance practices, replacing traditionally employed time-based methods [49, 50]. To this end, non-destructive tests and in-situ inspections are inadequate for continuous and global monitoring. Conversely, by assimilating sensor data from permanent data collection systems, vibration-based structural health monitoring techniques enable automated damage identification and evolution tracking [51, 52]. This paradigm shift has the potential to unlock personalized monitoring, management, and maintenance programs [53, 54], offering numerous benefits throughout the system life-cycle – including more informed structural safety assessments, better resource allocation, and increased system availability [55].

A graphical abstraction of the computational flow is illustrated in Fig. 1. The end-to-end loop spans from the physical to the digital domain through data assimilation and inference, and then back to the asset through action and observation, while explicitly accounting for uncertainty quantification, propagation, and resolution. We refer to the monitored asset, whose physical state is hidden to the AIF agent and only indirectly accessible via the sensed structural response, as the *external generative process*. The asset state evolves over time according to physical laws influenced by both its internal properties and external factors. These external factors might encompass long-term degradation mechanisms caused by chemical, physical, or mechanical aging, as well as sudden changes, such as discrete damage events or maintenance interventions [56].

The digital counterpart (AIF agent) is defined by an *internal generative model*, implemented as a probabilistic graphical model in the form of a dynamic Bayesian network (DBN) [5, 27]. This factored representation provides a systematic way to maintain a posterior belief about latent variables that characterize the (hidden) structural health of the asset, such as damage presence, location, and severity, by continuously integrating new observations within a sequential Bayesian inference scheme. Belief updating is achieved by minimizing variational *free energy*, which measures the discrepancy between the model's predicted observations and the actual sensor data.

In parallel, the internal generative model supports the forward simulation of future. This enables the ADT to evaluate "what-if" trajectories for structural health evolution, conditioned on its current beliefs. This forecasting step involves modeling not only the asset's physical dynamics but also the agent's control policies, represented as latent variables encoding sequences of future actions, usually termed policies [57]. Policy selection is then framed as an optimization problem, where the agent seeks to minimize the *expected free energy* – a quantity that balances (*i*) selecting policies that align future observations with (pragmatic) goal-directed prior preferences, and (*ii*) resolving uncertainty about hidden states through (epistemic) information-seeking. This formalism unifies inference and control: posterior beliefs are updated via free energy minimization, while action sequences are selected to minimize expected free energy, converting the problem of decision-making into a problem of inference under the generative model. Once an action is executed, the generative process evolves, and the bidirectional perception-action cycle restarts.

The paper is organized as follows. Section 2 describes the POMDP that encodes the coupled dynamics of the physical-digital system. Section 3 illustrates the use of AIF agents to realize ADTs. Section 4 assesses the proposed procedure on the simulated monitoring, management, and maintenance of a railway bridge, providing comparative results for different AIF agents featuring increasingly rich behavior. Conclusions and future developments are finally outlined in Sec. 5.

2. Partially observable Markov decision process for digital twins

Figure 2 illustrates the probabilistic graphical model – adapted from [26] – that represents the dynamic interaction between the physical and virtual domains. This abstraction is inspired by classical POMDP formulations [5]. POMDPs are state-space models for decision-making in stochastic, partially observable environments, where system dynamics are typically described by Markov transition models. Unlike standard Markov decision processes, where a policy directly maps observable states to actions, POMDPs define the policy as a mapping from belief states – probabilistic representations of hidden states inferred from observations – to actions.
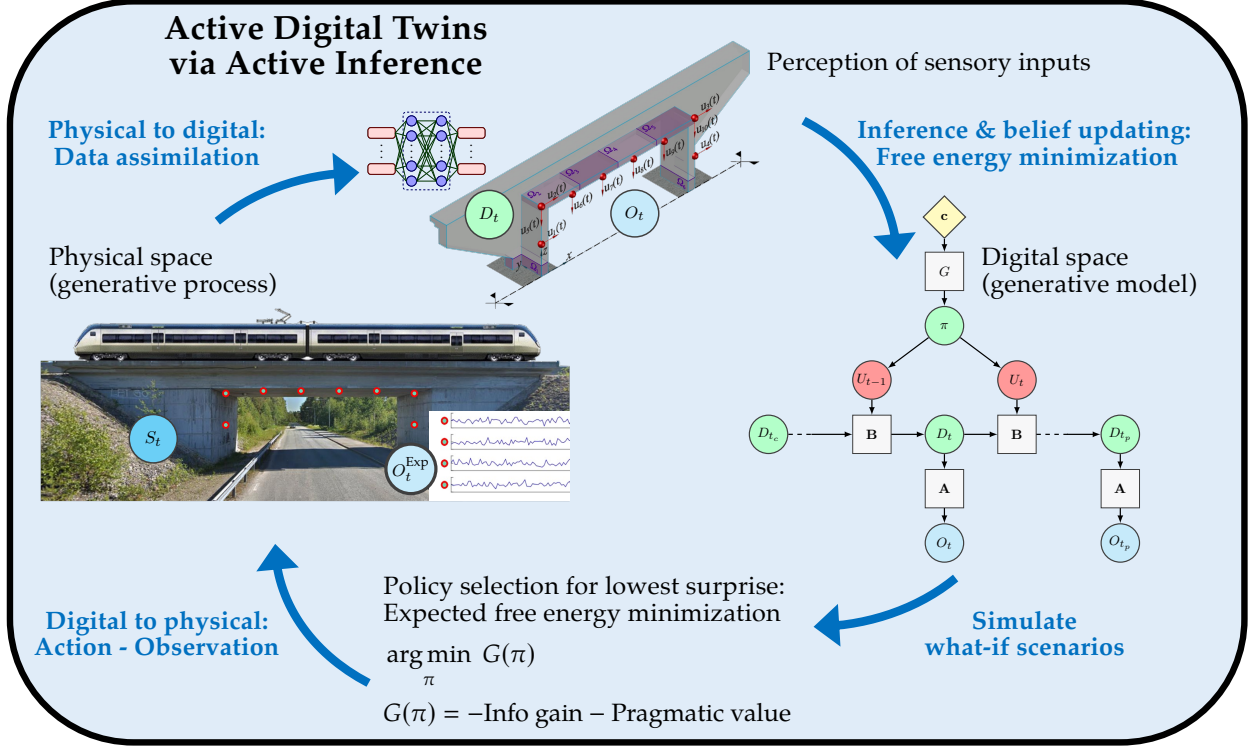
3

Figure 1: Active digital twins via active inference – Graphical abstraction of the end-to-end information flow. The dichotomy between the external physical process generating observational data (i.e., the *generative process*) and the agent's internal model (i.e., the *generative model*) is evident by the symmetry along the vertical axis. Meanwhile, the two forms of inference – digital state estimation and policy selection – exhibit a symmetry along the horizontal axis. A detailed schematic of generative models for both digital state and policy inference is presented in Fig. 4.

The graph in Fig. 2 is a DBN, in which circular nodes represent random variables, square nodes denote taken actions, and diamond-shaped nodes symbolize the objective function. All variables are defined at discrete time steps. Each time the DT is updated through the assimilation of new observational data, the DBN advances by one time step, with $t \in \{0, \ldots, T\}$, where $t = 0$ marks the moment the DT enters operation, and $t = T$ defines its lifetime horizon. Nodes with bold outlines indicate observed quantities, while those with thin outlines correspond to latent variables that must be inferred. The DBN is sparsely connected, with edges encoding conditional dependencies among the variables. For an overview of the fundamentals of DBNs, the reader is referred to [5, 27].

Capital letters denote random variables associated with the quantities in our abstraction, the corresponding lowercase letters refer to their specific realizations, and subscripts indicate their time index. Calligraphic letters denote the set of possible values each quantity can assume. For instance, the hidden physical state is denoted as $S_t \sim p(s_t)$, where s_t represents a particular realization at time t, and $p(s_t)$ defines the probability that $S_t = s_t$ for any possible state $s_t \in \mathcal{S}$.

The digital state $D_t \sim p(d_t)$ is designed to capture the essential features of the (hidden) physical state that are relevant for diagnosis, prediction, and decision-making [3]. The digital state space $\mathcal{D}$ can represent a variety of information, including initial and/or boundary conditions, material properties, and other key characteristics to describe the asset under consideration.

The physical-to-digital information flow from S_t to D_t is mediated by the assimilation of observational data $O_t \sim p(o_t)$, enabling the inference of D_t. The observation space $\mathcal{O}$ may include sensor measurements, inspection results, or diagnostic reports. Since the physical state S_t is only partially and indirectly observable, the digital state D_t encodes posterior beliefs over possible system configurations at time t, reflecting the evidence provided by the available observations [58, 59]. This perceptual process is realized through observation models – one for each observation modality – which relate digital states and observations in a
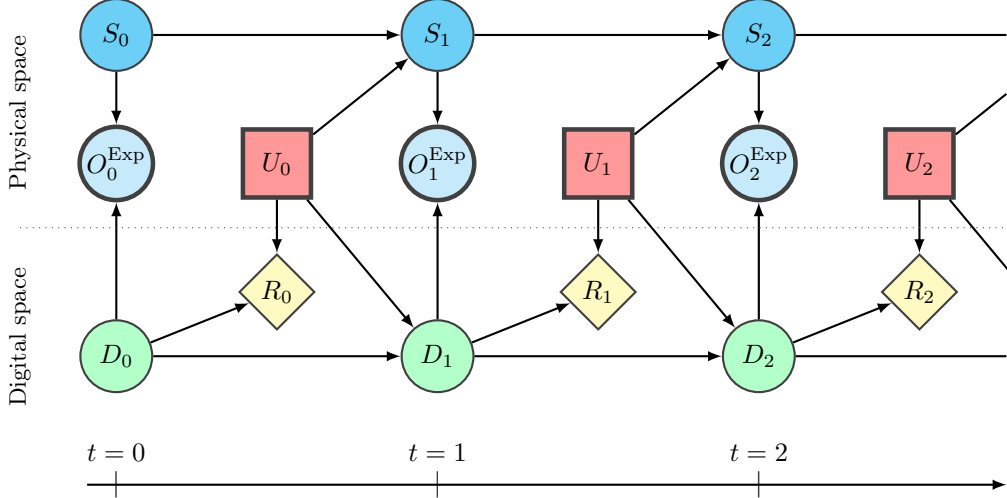
Figure 2: Dynamic Bayesian network encoding the asset-twin dynamical system. Circular nodes represent random variables, square nodes denote taken actions, and diamond-shaped nodes symbolize the objective function. Nodes with bold outlines indicate observed quantities, while those with thin outlines represent latent variables to be inferred. Directed edges encode conditional dependencies between variables.

probabilistic manner, such that O_t is modeled as stochastically generated from D_t. Throughout the paper, we will use both O_t and O_t^{Exp} to represent observations: O_t refers to predicted (expected) observations under the generative model, while O_t^{Exp} denotes actual sensor data. Belief updates are driven by minimizing the discrepancy between predicted and actual observations.

The updated digital state D_t informs the digital-to-physical information flow by guiding the selection of control actions to influence future physical states. In Fig. 2, $U_t \sim p(u_t)$ denotes a decision variable representing the action taken. The action space $\mathcal{U}$ may include interventions that directly modify the physical state, adjustments to the operational conditions, or modifications to the observational process. Each action is associated with its own transition model – one for each digital state factor – across the digital state space, which serves as a control-dependent predictor that propagates the digital state beliefs forward in time.

Finally, the reward node $R_t \sim p(r_t)$ quantifies the performance of the asset-twin system within a reward space $\mathcal{R}$. These rewards assess the expected "quality" of DBN trajectories to guide action selection toward optimal outcomes. In general, reward values may represent real costs associated with states and actions, or abstract metrics tuned to steer the system toward the desired behavior.

Formally, a POMDP can be defined as a seven-tuple $\langle \mathcal{D}, \mathcal{O}, \mathcal{U}, \mathcal{R}, \mathbf{A}, \mathbf{B}, \boldsymbol{\phi} \rangle$, where: $\mathcal{D}$ denotes the space of beliefs over hidden states; $\mathcal{O}$ is the space of possible observations; $\mathcal{U}$ is the space of available actions; $\mathcal{R} : \mathcal{D} \times \mathcal{U} \mapsto \mathbb{R}$ defines the reward function, which assigns a numerical value to beliefs-action pairs; $\mathbf{A} : \mathcal{O} \times \mathcal{D} \mapsto [0, 1]$ is the observation model, encoding the conditional observation likelihood $p(O_t \mid D_t; \boldsymbol{\phi})$, which represents beliefs about how hidden states give rise to observations; $\mathbf{B} : \mathcal{D} \times \mathcal{D} \times \mathcal{U} \mapsto [0, 1]$ is the transition model, encoding the conditional probability $p(D_t \mid D_{t-1}, U_{t-1}; \boldsymbol{\phi})$, which represents beliefs about the temporal evolution of hidden states conditioned on control actions; finally, $\boldsymbol{\phi}$ is a vector of hyperparameters of the POMDP model.

In the following, we assume that digital states, observational data, and control actions are defined over discrete and finite spaces. This implies that these variables can only take value on a finite set of discrete levels. Consequently, categorical distributions give a natural choice for representing the corresponding probability distributions. These latter assign a probability value between 0 and 1 to each discrete outcome, under the constraint that probabilities across all levels must sum to one, as they represent a complete and mutually exclusive set of realizations.

The joint probability distribution $p(O_t, D_t, U_t, R_t, \boldsymbol{\phi})$ over the POMDP factorizes – according to the chain rule of probability – into a product of categorical distributions (representing conditional likelihoods) and Dirichlet distributions (serving as priors). Numerically, these discrete distributions are organized as multidimensional arrays known as conditional probability tables (CPTs). The leading dimensions (rows) of

a CPT correspond to the support of the random variable, while the lagging dimensions (columns) represent the conditioning variables. Each column specifies the probability distribution of a random variable given a particular configuration of its parent nodes, and the entries within each column sum to one, as they represent a complete set of mutually exclusive and exhaustive outcomes. If a node has no parents, its CPT reduces to a single column representing the prior probabilities of its possible values. The contents of these CPTs can be controlled through the hyperparameters included in ϕ.

The complete set of possible realizations of the unobserved variables – conditioned on observational data $O_{0:t_c}^{\text{Exp}} = o_{0:t_c}^{\text{Exp}}$ and control actions $U_{0:t_c} = u_{0:t_c}$ – from the initial time step $t = 0$ up to the current time t_c, with $t = 0, \ldots, t_c$, can be extracted by leveraging the conditional independence assumptions implied by the graph structure in Fig. 2. The joint belief state can then be factorized according to the following sequential Bayesian inference formulation:

$$
\begin{aligned}
p(D_{0:t_c}, R_{0:t_c}, \phi \mid O_{0:t_c}^{\text{Exp}} = o_{0:t_c}^{\text{Exp}}, U_{0:t_c} = u_{0:t_c}) = \\
p(\phi)p(D_0; \phi) \prod_{t=1}^{t_c} p(D_t \mid D_{t-1}, u_{t-1}; \phi) \prod_{t=0}^{t_c} p(o_t^{\text{Exp}} \mid D_t; \phi)p(R_t \mid D_t, u_t).
\end{aligned}
\tag{1}
$$

In Eq. (1), the term $p(\phi)$ represents the prior distribution over the hyperparameters ϕ; inference over them typically evolves on a slower timescale than the inference of hidden states and control actions. $p(D_0; \phi)$ denotes the prior over the initial hidden states, representing the digital state belief at $t = 0$, before any observation is incorporated. The term $p(o_t^{\text{Exp}} \mid D_t; \phi)$ represents the sensory likelihood encoded in **A**. Similarly, $p(D_t \mid D_{t-1}, u_{t-1}; \phi)$ defines the transition likelihood encoded in **B**. Finally, $p(R_t \mid D_t, u_t)$ represents the likelihood of receiving a given reward, encapsulating the objective function evaluation. Note that selecting actions $U_t = u_t$ underpins solving the planning problem induced by the probabilistic graphical model. After forming a belief that measures the desirability of actions, such as $p(U_t \mid D_t)$, the actual action can be selected either as the best-point estimate or by sampling from this posterior, converting probabilistic control into a decision.

3. Active inference for digital twins

An attractive feature of AIF is that perception, learning, and action emerge as distinct manifestations of variational Bayesian inference [4]. Perception, or state estimation, is accomplished through inference over dynamically evolving hidden states, conditioned on assimilated observations and past actions. Learning corresponds to the gradual inference of hyperparameters that capture the statistical regularities of the environment. Action, in turn, is realized by inferring a posterior distribution over policies and sampling actions accordingly.

In the following, we describe the use of AIF agents to "navigate" the POMDP underlying the DT problem, enabling the full potential of ADTs. Section 3.1 introduces the AIF generative model, which encodes the probabilistic assumptions about the underlying environment. Section 3.2 addresses digital state inference via variational free energy minimization. Section 3.3 covers policy inference and action selection through expected free energy minimization. Section 3.4 describes the slow-scale learning of the hyperparameters that define the AIF generative model. Finally, Sec. 3.5 discusses the active information-seeking (epistemic) behavior that characterizes ADTs.

3.1. Active inference generative model

In AIF, the set of probabilistic assumptions about how the environment (or *generative process*) produces observations (via the observation model **A**) and how actions influence the environment evolution (via the transition model **B**) is referred to as the POMDP generative model. This model is used to represent the joint distribution in Eq. (1), from current time t_c to a prediction horizon $t_p > t_c$. Specifically, for time- and space-discretized POMDPs, probabilistic estimates of future digital states and observations over the prediction time steps $t = t_c, \ldots, t_p$ are computed as:

$$
p(O_{t_c:t_p}, D_{t_c:t_p}, \phi \mid \pi) = p(\phi)p(D_{t_c}; \phi) \prod_{t=t_c+1}^{t_p} p(D_t \mid D_{t-1}, \pi; \phi) \prod_{t=t_c}^{t_p} p(O_t \mid D_t; \phi),
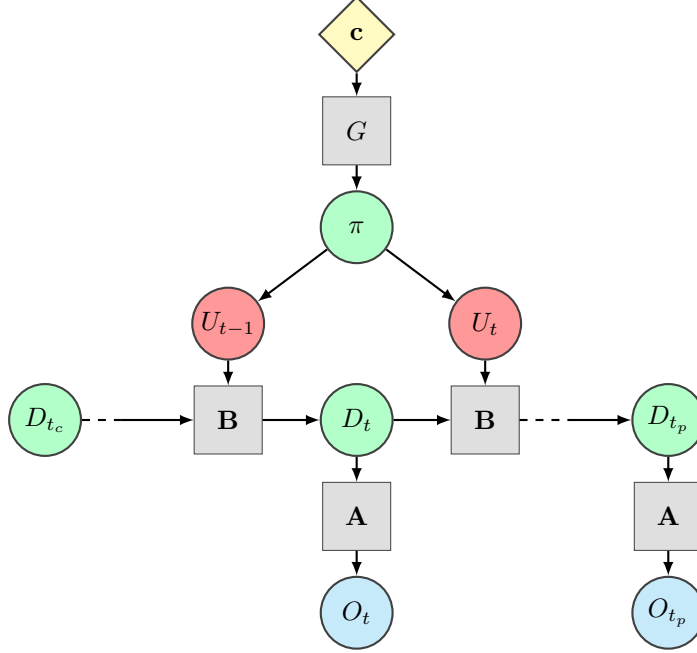\tag{2}
$$

Figure 3: Dynamic Bayesian network encoding the active inference generative model used to predict future digital states and observations under each policy. Circular nodes represent random variables, while the diamond-shaped node denotes prior preferences that reflect a goal-directed (pragmatic) objective. Gray square nodes represent parametrized operators of the generative model. Directed edges encode conditional dependencies between variables.

which reflects unrolling the AIF generative model of Fig. 3 over $t = t_c, \ldots, t_p$. Compared to Eq. (1), the factorization in Eq. (2) introduces several modifications to align with the AIF framework. First, the control variable U is replaced by a policy π, defined as a sequence of control states $\pi = \{u_{t_c}, \ldots, u_{t_p}\}$. The generative model in Eq. (2) is conditioned on a fixed policy π, which is how it is used for inference purposes. Policies are treated as latent variables to be inferred: the posterior over policies represents the agent beliefs about its intended actions, while single actions are realizations sampled from the posterior over control states. The policy-to-control mapping $p(U_t \mid \pi)$ assigns the control state at each time-step based on the selected policy.

The second modification concerns the omission of the reward variable R. In AIF, utility-maximization goals are encoded as a prior distribution $\widetilde{p}(O_{t_c:t_p})$ over future observations. These preferences are specified through an unconditional CPT $\mathbf{c} : \mathcal{O} \mapsto [0,1]$. Such prior preferences guide policy selection toward goal-directed (pragmatic) behavior by favoring actions expected to produce preferred observations. This formal equivalence between rewards and priors eliminates the need for explicit cost functions. Further, it enables optimal control to be cast as an inference problem: the joint probability of observations, digital states, control states, and model parameters is maximized when the system samples from preferred observations. The square node G in the graph represents the expected free energy, which quantifies the desirability of each policy by incorporating both pragmatic and information-seeking (epistemic) components, as explained in Sec. 3.3. Finally, the initial prior $p(D_{t_c}; \boldsymbol{\phi})$ is typically represented by an unconditional CPT denoted as $\mathbf{d} : \mathcal{D} \mapsto [0,1]$.

Graphically, the generative model illustrated in Fig. 3 shows several differences compared to the DBN in Fig. 2. This formulation focuses on predicting future digital states $D_{t_c:t_p}$ and sampling (or generating) sequences of potential observations $O_{t_c:t_p}$ based on the probabilistic structure encoded in $\mathbf{A}$ and $\mathbf{B}$, conditioned on control actions $U_{t_c:t_p}$ that have not yet been executed. Accordingly, actions U_t are modeled as (circular) random variables rather than (square) decision nodes, since they represent what-if scenarios beyond data assimilation. Moreover, the same color is used to represent both digital states and control policies in the graph, as both are latent variables of the generative model. Equipped with this generative model – specified by the four-tuple $\langle \mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{d} \rangle$ – AIF supports the inference over D_t, π, and ϕ, as described in the following sections.

3.2. Digital state inference via variational free energy minimization

Given an observation $O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}}$, the underlying digital state D_{t_c} can be inferred by estimating a posterior distribution $p(D_{t_c} \mid O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}})$, using Bayes' Rule:

$$p(D_{t_c} \mid O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}}) = \frac{p(o_{t_c}^{\text{Exp}}, D_{t_c})}{p(o_{t_c}^{\text{Exp}})} = \frac{p(o_{t_c}^{\text{Exp}} \mid D_{t_c})p(D_{t_c})}{\sum_{d_{t_c} \in \mathcal{D}} p(o_{t_c}^{\text{Exp}}, D_{t_c} = d_{t_c})}, \tag{3}$$

where the (generative model) joint distribution $p(o_{t_c}^{\text{Exp}}, D_{t_c})$ is factorized into a likelihood term $p(o_{t_c}^{\text{Exp}} \mid D_{t_c})$ and a prior $p(D_{t_c})$. The denominator $p(o_{t_c}^{\text{Exp}})$ is the marginal likelihood or model evidence, which captures the probability of observing $O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}}$ under the generative model.

Since Bayesian inversion to estimate hidden states from observations is generally intractable, AIF employs variational inference [60] as an approximate Bayesian method, trading exactness for computational tractability. Specifically, we define a tractable variational distribution $Q(D_{t_c}; \boldsymbol{\theta}) : \mathcal{D} \mapsto [0, 1]$, parametrized by $\boldsymbol{\theta}$, and optimize this surrogate distribution to make it as close as possible to the true posterior $p(D_{t_c} \mid O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}})$. In our discrete POMDP setting, the variational parameters $\boldsymbol{\theta}$ correspond to the relative frequencies of each category in the support of a random variable. This leads to the following optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \ \mathrm{D}_{\text{KL}} \left[Q(D_{t_c}; \boldsymbol{\theta}) \,||\, p(D_{t_c} \mid o_{t_c}^{\text{Exp}}) \right], \tag{4}$$

where $\mathrm{D}_{\text{KL}} \left[Q(X) \,||\, P(X \mid Y) \right] = \mathbb{E}_Q \left[\ln Q(X) - \ln P(X \mid Y) \right]$ denotes the Kullback-Leibler (KL) divergence between the approximate posterior $Q(X)$ and the true posterior $P(X \mid Y)$, for two generic random variables X and Y. Here, $\mathbb{E}_Q$ denotes the expectation with respect to the variational posterior. However, this objective remains intractable because it depends on the true posterior $p(D_{t_c} \mid o_{t_c}^{\text{Exp}})$ that we seek to approximate. To circumvent this, we reformulate the objective as the variational free energy (VFE):

$$\begin{aligned}
\mathcal{F}_{t_c}(\boldsymbol{\theta}) &= \mathrm{D}_{\text{KL}} \left[Q(D_{t_c}; \boldsymbol{\theta}) \,||\, p(D_{t_c} \mid o_{t_c}^{\text{Exp}}) \right] - \ln p(o_{t_c}^{\text{Exp}}) \\
&= \sum_{\mathcal{D}} Q(D_{t_c}; \boldsymbol{\theta}) \left[\ln \frac{Q(D_{t_c}; \boldsymbol{\theta})}{p(D_{t_c} \mid o_{t_c}^{\text{Exp}})} - \ln p(o_{t_c}^{\text{Exp}}) \right] \\
&= \mathbb{E}_Q \left[\ln Q(D_{t_c}; \boldsymbol{\theta}) - \ln p(o_{t_c}^{\text{Exp}}, D_{t_c}) \right],
\end{aligned} \tag{5}$$

which serves as an upper bound on the negative log marginal likelihood $(-\ln p(o_{t_c}^{\text{Exp}}))$, also known as the Bayesian surprise. Minimizing VFE thus brings the variational posterior closer to the true posterior while simultaneously increasing the marginal likelihood of the observation. The VFE objective leads to the following final form of the optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \ \mathcal{F}_{t_c}(\boldsymbol{\theta}). \tag{6}$$

At convergence, if $Q^*(D_{t_c}; \boldsymbol{\theta}^*)$ exactly matches the true posterior, the KL divergence vanishes, i.e., $\mathrm{D}_{\text{KL}} = 0$, and the VFE equals surprise: $\mathcal{F}_{t_c} = -\ln p(o_{t_c}^{\text{Exp}})$. By further minimizing surprise, the VFE then provides a useful objective not only for inference but also for learning the parameters of the generative model. The underlying rationale is that AIF agents aim to avoid surprising observations, and minimizing surprise is equivalent to maximizing model evidence.

With reference to the generative model formulation (2), instantaneous inference over digital states involves approximating the true posterior $p(D_{t_c} \mid O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}}, D_{t_c-1}, U_{t_c-1} = u_{t_c-1})$. This inference is conditioned on the current observation $O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}}$, the previous (posterior) distribution over digital states D_{t_c-1}, and the previously executed action $U_{t_c-1} = u_{t_c-1}$, as:

$$\begin{aligned}
\boldsymbol{\theta}^* &= \arg\min_{\boldsymbol{\theta}} \ \mathcal{F}_{t_c}(\boldsymbol{\theta}) \\
&= \arg\min_{\boldsymbol{\theta}} \ \mathbb{E}_Q \left[\ln Q(D_{t_c}; \boldsymbol{\theta}) - \ln p(o_{t_c}^{\text{Exp}}, D_{t_c} \mid D_{t_c-1}, u_{t_c-1}; \boldsymbol{\phi}) \right] \\
&= \arg\min_{\boldsymbol{\theta}} \ \mathbb{E}_Q \left[\ln Q(D_{t_c}; \boldsymbol{\theta}) - \ln \left(p(o_{t_c}^{\text{Exp}} \mid D_{t_c}; \boldsymbol{\phi}) p(D_{t_c} \mid D_{t_c-1}, u_{t_c-1}; \boldsymbol{\phi}) \right) \right].
\end{aligned} \tag{7}$$

8

The optimization problem in Eq. (7) is solved using fixed-point iteration [61], under the assumption of temporal factorization, where variational posteriors at different time steps are conditionally independent. As a result, the full VFE across trajectories decomposes into a sum of single-time-step free energies, enabling independent optimization at each time point. Moreover, the posterior $Q(D_t; \boldsymbol{\theta})$ at a given time step t can be further factorized across F independent hidden state factors $D = \{D^1, \ldots, D^F\}$, following the mean-field approximation [62]:

$$Q(D_t; \boldsymbol{\theta}) = \prod_{f=1}^{F} Q(D_t^f; \boldsymbol{\theta}), \tag{8}$$

where $Q(D_t^f; \boldsymbol{\theta})$ denotes the posterior over the fth hidden state factor, $f = 1, \ldots, F$.

These factors may represent distinct aspects of the generative process, potentially varying in dimensionality, transition dynamics, and association with specific observation modalities. Similarly, observations can be structured into M distinct modalities $O = \{O^1, \ldots, O^M\}$, where each O^m, $m = 1, \ldots, M$, corresponds to a separate sensory channel used by the agent at each time step. For example, in a DT application for the human health, one hidden state factor may represent a patient's metabolic state, while another factor could encode cardiovascular function. Correspondingly, observation modalities may include blood glucose readings and heart rate measurements, each providing information about different latent physiological processes.

In this multi-modal, multi-factor setup, the observation likelihood array $\mathbf{A}$ becomes a collection of M sub-arrays $\mathbf{A} = \{\mathbf{A}^1, \ldots, \mathbf{A}^M\}$, with each $\mathbf{A}^m$, $m = 1, \ldots, M$, representing the observation model for the mth modality. Each sub-array encodes the likelihood $p(O^m \mid D^1, \ldots, D^F; \boldsymbol{\phi})$, capturing the dependency of that observation modality on the hidden state factors. Similarly, the transition model $\mathbf{B}$ is represented as a collection of F sub-arrays $\mathbf{B} = \{\mathbf{B}^1, \ldots, \mathbf{B}^F\}$, under the assumption that hidden state factors evolve independently without influencing each other. Each $\mathbf{B}^f$, $f = 1, \ldots, F$, encodes the dynamics $p(D_t^f \mid D_{t-1}^f, u_{t-1}^f; \boldsymbol{\phi})$, conditioned on the previous state and action for that factor. Note that control states are factorized analogously to hidden states, such that $U = \{U^1, \ldots, U^F\}$. Each control factor U^f governs the transitions of the corresponding digital state factor D^f, with a dimensionality matching the number of possible control actions applicable to that aspect of the system.

This factored structure enables the encoding of complex conditional dependencies while significantly reducing memory requirements. For instance, if the model employs two separate hidden state factors to represent the location and identity of a phenomenon, the memory requirements for the factored representation scale linearly with the dimensionality of the two factors. An additional advantage of this factorization lies in its interpretability: by explicitly designing digital state factors to reflect intuitive features of the environment, the resulting generative model becomes more transparent and modular. In contrast, explicitly enumerating all possible combinations of "where" and "what" would incur polynomial memory complexity.

The marginal variational posteriors for each hidden state factor at the current time t_c are computed analytically via mean-field fixed-point iteration [61]. The algorithm proceeds by setting the gradient of the VFE $\mathcal{F}_{t_c}(\boldsymbol{\theta})$ to zero, and iteratively solving for each factorized component $Q(D_{t_c}^f; \boldsymbol{\theta})$, for $f = 1, \ldots, F$. A detailed derivation of this procedure can be found in [63].

Note that in the AIF framework, there is no need to introduce an explicit node for representing quantities of interest, unlike the abstraction of physical-digital systems proposed in [26]. In their probabilistic graphical model, these variables are represented by a dedicated node and predicted from the updated digital state via the computational models comprising the DT. In contrast, under the AIF framework, such a node is redundant, as quantities of interest are naturally embedded within the observational data node. When observational evidence is unavailable for a particular modality, inference simply remains uninformed in that dimension of the observation space. Nevertheless, the updated digital state can still be used to predict expected values across any observation channel – whether observed or unobserved – via the corresponding observation model. The models may, in principle, incorporate arbitrarily complex forward mappings, ranging from high-fidelity physics-based simulators to purely data-driven surrogates or hybrid combinations of the two.

3.3. Policy inference-action selection via expected free energy minimization

Given the updated variational posterior over the digital state $Q^*(D_{t_c}; \boldsymbol{\theta}^*)$, policy inference involves evaluating the quality of each admissible policy comprising future actions over a prediction horizon $t = t_c, \ldots, t_p$.
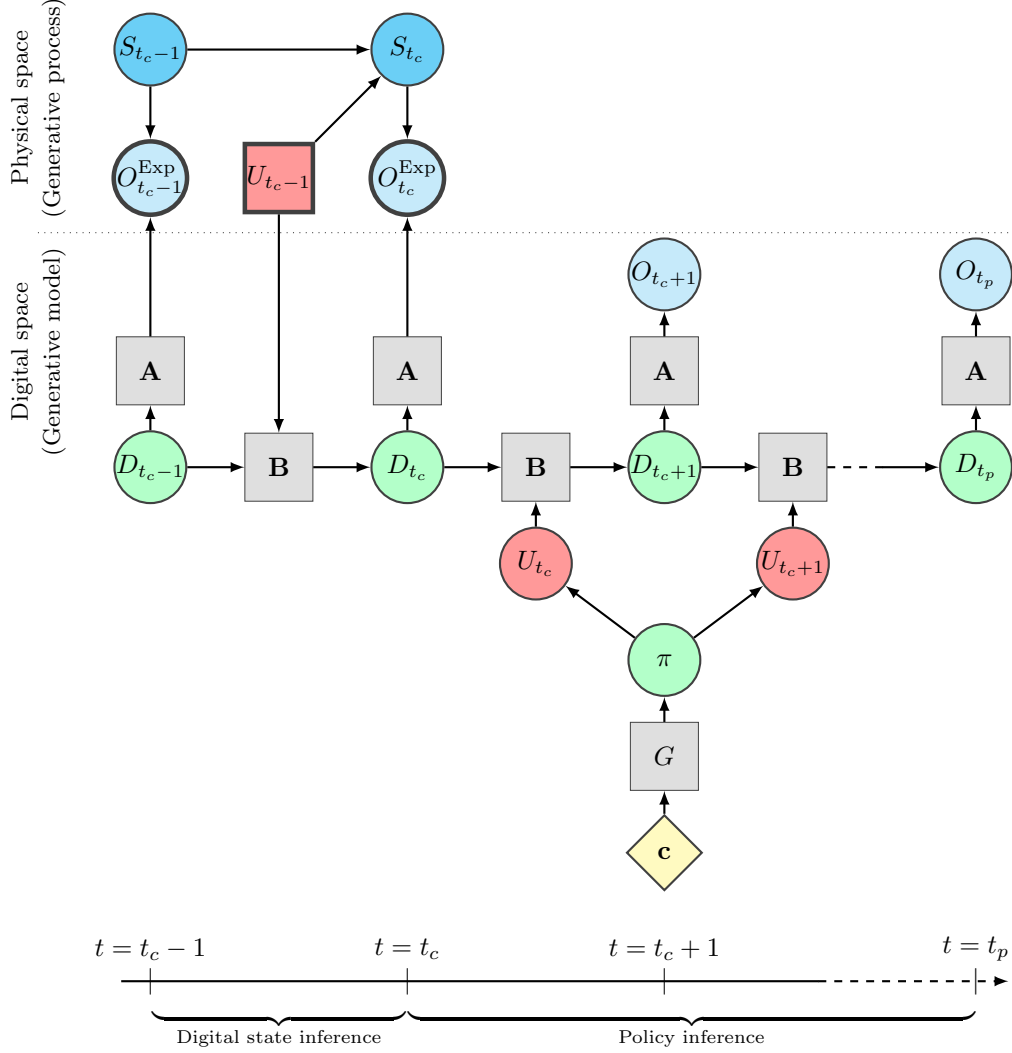
Figure 4: A dynamic Bayesian network illustrating the use of active inference generative models to navigate the partially observable Markov decision process underlying the digital twin problem. Circular nodes represent random variables, red square nodes denote taken actions, gray square nodes represent parametrized operators of the generative model, and the diamond-shaped node symbolizes prior preferences that reflect a goal-directed (pragmatic) objective. Nodes with bold outlines indicate observed quantities, while those with thin outlines represent latent variables to be inferred. Directed edges encode conditional dependencies between variables. The upper left-to-right path represents the evolution of the physical space, while the lower path depicts the evolution of the digital space. Digital state inference is performed at the current time t_c, whereas policy inference involves propagating the updated digital state from t_c to the prediction time t_p.

In AIF, the desirability of (or preference for) each policy is quantified through the expected free energy (EFE). The EFE is the central quantity driving the behavior of ADTs and is formulated to evaluate sequences of actions (or policies) both on goal-directed (pragmatic) and information-seeking (epistemic) behaviors. Like the VFE, the EFE is a function of observations, hidden states, and policies. However, different from the VFE, it pertains to sequences of *future* actions, where no actual observations are yet available, and it includes expectations over future digital states and future observations generated by the generative model.

The use of AIF generative models for digital state inference and policy inference is graphically summarized in Fig. 4. Digital state inference integrates the prior belief at time $t_c - 1$ with the observational data assimilated at t_c. In contrast, policy inference entails predictive modeling over the horizon $t = t_c, \ldots, t_p$, where the generative model operates without access to future sensory data or executed actions from the interfacing generative process.

The EFE associated to a generic policy π is defined as:

$$G^\pi = \mathbb{E}_{Q(O_{t_c:t_p}, D_{t_c:t_p}|\pi)} \left[\ln Q(D_{t_c:t_p} \mid \pi) - \ln \widetilde{p}(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi) \right], \tag{9}$$

where, for simplicity, we omit the explicit dependence of the variational posterior on the variational parameters $\boldsymbol{\theta}$, denoting it simply as $Q(D_t)$. Similarly, we omit the dependency of the generative model on the hyperparameters $\boldsymbol{\phi}$. In Eq. (9), $\widetilde{p}(O_t, D_t \mid \pi) = p(D_t \mid O_t, \pi)\widetilde{p}(O_t)$ defines a generative model biased by the predictive prior over observations $\widetilde{p}(O_t)$. This construction integrates the prior preferences encoded in $\mathbf{c}$ into the inference process (described below), enabling the AIF agent to act in ways that maximize the likelihood of preferred outcomes.

Given the assumed conditional independence of variational posteriors across time, the EFE at a generic time step $t \in \{t_c, \ldots, t_p\}$ for policy π is given by:

$$\begin{aligned}
G_t^\pi &= \mathbb{E}_{Q(O_t, D_t|\pi)} \left[\ln Q(D_t \mid \pi) - \ln \widetilde{p}(O_t, D_t \mid \pi) \right] \\
&= - \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\mathrm{D}_{\mathrm{KL}} \left[Q(D_t \mid O_t, \pi) \| Q(D_t \mid \pi) \right] \right]}_{\text{Epistemic value (information gain)}} - \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\ln \widetilde{p}(O_t) \right]}_{\text{Pragmatic value (utility)}} \\
&+ \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\mathrm{D}_{\mathrm{KL}} \left[Q(D_t \mid O_t, \pi) \| p(D_t \mid O_t, \pi) \right] \right]}_{\text{Expected variational approximation error } (\geq 0)},
\end{aligned} \tag{10}$$

with the complete derivation provided in Appendix A, as adapted to the ADT framework from [63]. In Eq. (10), the first term denotes the epistemic value [28], which promotes information-seeking behavior. It favors policies under which the agent is expected to *explore* states that yield high information gain about the digital state. This gain is quantified as the divergence between predicted digital states conditioned and unconditioned on observations under the same policy. The second term corresponds to the pragmatic value, which reflects goal-directed behavior. It favors policies that lead the agent to states expected to generate outcomes aligned with prior preferences $\widetilde{p}(O_t)$. The final term captures the expected approximation error – the divergence between the true digital state posterior and its variational approximation – which is typically assumed to be negligible.

The epistemic drive in Eq. (10) is a crucial component that enables ADTs to exhibit spontaneous exploratory behavior. Epistemic actions in ADTs encompass decisions that gather information or improve the digital state observability. These may include, for instance, installing new sensors, scheduling targeted inspections, or testing model predictions. For example, in a manufacturing ADT, the agent might deliberately vary process parameters within safe limits to resolve uncertainty about machine wear dynamics. In a personalized medicine context, the ADT might recommend a low-risk diagnostic test to disambiguate between competing hypotheses about a patient's physiological condition. In both cases, the primary objective of these actions is not immediate (pragmatic) utility maximization, but rather to refine the generative model and enhance the understanding of the environment.

The EFE of temporally deep policies is given by the sum of time step-specific contributions:

$$G^\pi = \sum_{t=t_c}^{t_p} G_t^\pi, \tag{11}$$

where each term is evaluated based on the agent's predictive beliefs over future digital states and observations. The computation begins from the current posterior belief $Q^*(D_{t_c})$, which is then propagated over the prediction horizon $t = t_c, \ldots, t_p$ using the policy-specific transition and observation models. This process generates the posterior predictive densities $Q(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi)$, which are subsequently used to evaluate the goal-directed (pragmatic) and information-seeking (epistemic) values at each time step.

Let $\Pi = \{\pi_1, \ldots, \pi_P\}$ denote the set of P feasible policies, constructed through the combinatorial enumeration of sequences of actions from the action space $\mathcal{U}$ over the time horizon $t = t_c, \ldots, t_p$. The EFE vector $\mathbf{G} = [G^{\pi_1}, \ldots, G^{\pi_P}]^\top \in \mathbb{R}^P$, which assigns a scalar EFE to each policy, defines a prior over policies according to:

$$p(\pi) = \sigma(-\gamma \mathbf{G}), \tag{12}$$

where $\sigma(x) = \frac{\exp(x)}{\sum_x \exp(x)}$ is the Softmax function, and $\gamma \in \mathbb{R}^+$ is an inverse temperature parameter that modulates the precision over policies. Higher γ values yield more deterministic preferences.

Under the prior (12), AIF agents perform policy inference by optimizing a variational posterior over policies $Q(\pi)$ [63], to minimize the following VFE expansion over the prediction horizon $t = t_c, \ldots, t_p$:

$$
\begin{aligned}
\mathcal{F}_{t_c:t_p} &= \mathbb{E}_{Q(D_{t_c:t_p}, \pi)} \left[\ln Q(D_{t_c:t_p}, \pi) - \ln p(O_{t_c:t_p}, D_{t_c:t_p}, \pi) \right] \\
&= \mathbb{E}_{Q(D_{t_c:t_p}, \pi)} \left[\ln Q(D_{t_c:t_p} \mid \pi) + \ln Q(\pi) - \ln p(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi) - \ln p(\pi) \right] \\
&= \mathbb{E}_{Q(\pi)} \left[\ln Q(\pi) - \ln p(\pi) \right] \\
&\quad + \mathbb{E}_{Q(\pi)} \left[\mathbb{E}_{Q(D_{t_c:t_p} \mid \pi)} \left[\ln Q(D_{t_c:t_p} \mid \pi) - \ln p(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi) \right] \right] \\
&= \mathrm{D_{KL}} \left[Q(\pi) \,\|\, p(\pi) \right] + \mathbb{E}_{Q(\pi)} \left[\mathcal{F}_{t_c:t_p}^{\pi} \right],
\end{aligned}
\tag{13}
$$

which measures the KL divergence between the approximate posterior $Q(D_{t_c:t_p}, \pi)$ and the generative model $p(O_{t_c:t_p}, D_{t_c:t_p}, \pi)$ as a sum of two contributions. The first is the KL divergence between the variational posterior over policies and the corresponding prior (12), thereby incorporating the EFE into the inference process. The second term is a policy-weighted average of the free energy across all policies, where $\mathcal{F}_{t_c:t_p}^{\pi}$ denotes the free energy associated with a single policy π:

$$
\begin{aligned}
\mathcal{F}_{t_c:t_p}^{\pi} &= -\mathbb{E}_{Q(D_{t_c:t_p} \mid \pi)} \left[\ln p(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi) - \ln Q(D_{t_c:t_p} \mid \pi) \right] \\
&= -\mathbb{E}_{Q(D_{t_c:t_p} \mid \pi)} \left[\ln p(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi) \right] - H \left[Q(D_{t_c:t_p} \mid \pi) \right],
\end{aligned}
\tag{14}
$$

with $H \left[Q(D_{t_c:t_p} \mid \pi) \right] = \mathbb{E}_{Q(D_{t_c:t_p} \mid \pi)} \left[-\ln Q(D_{t_c:t_p} \mid \pi) \right]$ being the variational posterior entropy, which quantifies the uncertainty in the beliefs about future digital states under policy π.

By evaluating each policy independently and computing its associated free energy, the optimal posterior $Q^*(\pi)$ is obtained by minimizing the total VFE $\mathcal{F}_{t_c:t_p}$ with respect to $Q(\pi)$. This is achieved by enforcing the stationarity of $\mathcal{F}_{t_c:t_p}$ with respect to $Q(\pi)$, leading to a Softmax distribution through the following update rule:

$$
Q^*(\pi) = \underset{Q(\pi)}{\arg\min} \, \mathcal{F}_{t_c:t_p} = \sigma(\ln p(\pi) - \mathcal{F}_{t_c:t_p}^{\pi}),
\tag{15}
$$

assigning higher probability to policies with lower free energy while remaining close to the prior.

The posterior over policies can be further biased by incorporating a policy prior $p(\pi_0)$, which encodes habitual tendencies. For example, $p(\pi_0)$ could represent the standard policy implemented by decision-makers, such as state agencies and companies, depending on the context, or standard protocols in medical settings. By expanding the prior (12) as:

$$
p(\pi) = \sigma(\ln p(\pi_0) - \gamma \mathbf{G}),
\tag{16}
$$

the resulting update rule for belief estimation becomes:

$$
Q^*(\pi) = \underset{Q(\pi)}{\arg\min} \, \mathcal{F}_{t_c:t_p} = \sigma(\ln p(\pi_0) - \gamma \mathbf{G} - \mathcal{F}_{t_c:t_p}^{\pi}).
\tag{17}
$$

The posterior over control states $Q^*(U_t)$ is formed by marginalizing over policies as follows:

$$
Q^*(U_t) = \sum_{\pi \in \Pi} p(U_t \mid \pi) Q^*(\pi),
\tag{18}
$$

where $p(U_t \mid \pi)$ defines a deterministic mapping from policies to control states. The actual action $U_{t_c} = u_{t_c}$ to be executed on the system can eventually be selected either as the maximum a-posteriori estimate or by sampling from $Q^*(U_{t_c})$.

3.4. Learning of the generative model via parameter inference

In this section, we describe the learning of the parameters ϕ that define the AIF generative model, based on the outcomes of inference. "Learning" ϕ is a generative model's parameter updating occurring at a slower timescale than the faster inference processes for digital states and policies. Nevertheless, the update equations for ϕ follow the same variational principles of digital state inference, where a variational posterior over ϕ is optimized through VFE minimization.

In our discrete setting, posterior inference over ϕ is performed by parametrizing the likelihood and prior distributions of the generative model with Dirichlet distributions, following an approach similar to [11, 21]. The choice to treat hyperparameters ϕ as the parameters of Dirichlet distributions is motivated by their conjugacy to the categorical distribution. This formulation enables online learning via closed-form Bayesian updates, allowing evidence about the system response to actions to be incorporated efficiently, while ensuring that the posterior remains within the Dirichlet family. The approach is computationally scalable and supports continual refinement of ADTs, even when initialized with potentially inaccurate or uncertain priors and likelihoods.

In the following, we refer to the generative model (2) by decomposing ϕ into subsets corresponding to the categorical and Dirichlet parameters associated with the arrays $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{d}$. Specifically, we write $\phi = \{\mathfrak{A}, \mathfrak{a}, \mathfrak{B}, \mathfrak{b}, \mathfrak{D}, \mathfrak{d}\}$ to explicitly highlight the stochastic parametrization of each likelihood and prior distribution, as defined below:

1. The observation model $\mathbf{A} \in \mathbb{R}^{|\mathcal{O}| \times |\mathcal{D}|}$, which encodes the observation likelihood $p(O_t \mid D_t; \mathfrak{A})$, is parametrized by the matrix of categorical probabilities $\mathfrak{A} \in \mathbb{R}^{|\mathcal{O}| \times |\mathcal{D}|}$, as follows:

$$O_t \mid D_t; \mathfrak{A} \sim \mathrm{Cat}(\mathfrak{A}), \tag{19}$$

$$p(\mathfrak{A}) = \prod_{d \in \mathcal{D}} p(\mathfrak{A}_{\bullet,d}), \quad \mathfrak{A}_{\bullet,d} \sim \mathrm{Dir}(\mathfrak{a}_{\bullet,d}), \tag{20}$$

where $|\mathcal{X}| \in \mathbb{N}$ denotes the cardinality of a generic set $\mathcal{X}$; $\mathrm{Cat}(X)$ and $\mathrm{Dir}(X)$ denote categorical and Dirichlet distributions over a generic random variable X, respectively; the notation $\mathbf{X}_{\bullet,j}$ refers to the jth column of a matrix $\mathbf{X}$; and $\mathfrak{a} \in \mathbb{R}^{|\mathcal{O}| \times |\mathcal{D}|}$ is the matrix of (positive) concentration parameters defining the Dirichlet prior over $\mathfrak{A}$. These parameters encode prior beliefs over categorical probabilities and can be interpreted as pseudo-counts representing the expected frequency of each possible realization – here, the frequency of each observation given a digital state. For notational simplicity, we assume the generative model is not factorized into multiple digital state factors or observation modalities. However, the formulation can be easily extended to a multi-modal, multi-factor setup via additional parametrized dimensions.

2. The transition model $\mathbf{B} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}| \times |\mathcal{U}|}$, which encodes the control-dependent forward-time predictor $p(D_t \mid D_{t-1}, u_{t-1}; \mathfrak{B})$, is parametrized by the tensor of categorical parameters $\mathfrak{B} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}| \times |\mathcal{U}|}$, as follows:

$$D_t \mid D_{t-1}, u_{t-1}; \mathfrak{B} \sim \mathrm{Cat}(\mathfrak{B}), \tag{21}$$

$$p(\mathfrak{B}) = \prod_{d \in \mathcal{D}} \prod_{u \in \mathcal{U}} p(\mathfrak{B}_{\bullet,d,u}), \quad \mathfrak{B}_{\bullet,d,u} \sim \mathrm{Dir}(\mathfrak{b}_{\bullet,d,u}), \tag{22}$$

where $\mathfrak{b} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}| \times |\mathcal{U}|}$ is the tensor of parameters for the Dirichlet prior over $\mathfrak{B}$.

3. the initial state model $\mathbf{d} \in \mathbb{R}^{|\mathcal{D}|}$, which encodes the prior over initial digital states $p(D_0; \mathfrak{D})$, is parametrized by the vector of categorical parameters $\mathfrak{d} \in \mathbb{R}^{|\mathcal{D}|}$, as follows:

$$D_0; \mathfrak{D} \sim \mathrm{Cat}(\mathfrak{D}), \tag{23}$$

$$\mathfrak{D} \sim \mathrm{Dir}(\mathfrak{d}), \tag{24}$$

where $\mathfrak{d} \in \mathbb{R}^{|\mathcal{D}|}$ is the vector of parameters defining the Dirichlet prior over $\mathfrak{D}$.

Given the split of ϕ into the individual parametrizations for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{d}$, the generative model (2) can be expressed as:

$$p(O_{t_c:t_p}, D_{t_c:t_p}, \mathfrak{A}, \mathfrak{B}, \mathfrak{D}, \pi) =$$

$$p(\mathfrak{A})p(\mathfrak{B})p(\mathfrak{D})p(\pi)p(D_{t_c}; \mathfrak{D}) \prod_{t=t_c+1}^{t_p} p(D_t \mid D_{t-1}, \pi; \mathfrak{B}) \prod_{t=t_c}^{t_p} p(O_t \mid D_t; \mathfrak{A}). \tag{25}$$

Learning is thus formulated as the approximate inference of $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{D}$ by minimizing the VFE with respect to their corresponding approximate posteriors $Q(\mathfrak{A})$, $Q(\mathfrak{B})$, and $Q(\mathfrak{D})$. The full variational posterior is assumed to factorize as:

$$Q(D_{t_c:t_p}, \mathfrak{A}, \mathfrak{B}, \mathfrak{D}, \pi) = Q(\mathfrak{A})Q(\mathfrak{B})Q(\mathfrak{D})Q(\pi) \prod_{t=t_c}^{t_p} Q(D_t \mid \pi), \tag{26}$$

where the variational distributions $Q(\mathfrak{A})$, $Q(\mathfrak{B})$, and $Q(\mathfrak{D})$ are modeled as Dirichlet distributions:

$$Q(\mathfrak{A}) = \prod_{d \in \mathcal{D}} Q(\mathfrak{A}_{\bullet,d}), \qquad\qquad Q(\mathfrak{A}_{\bullet,d}) = \text{Dir}(\widehat{\mathfrak{a}}_{\bullet,d}), \tag{27}$$

$$Q(\mathfrak{B}) = \prod_{d \in \mathcal{D}} \prod_{u \in \mathcal{U}} Q(\mathfrak{B}_{\bullet,d,u}), \qquad\qquad Q(\mathfrak{B}_{\bullet,d,u}) = \text{Dir}(\widehat{\mathfrak{b}}_{\bullet,d,u}), \tag{28}$$

$$Q(\mathfrak{D}) = \text{Dir}(\widehat{\mathfrak{d}}), \tag{29}$$

where $\widehat{\mathfrak{a}} \in \mathbb{R}^{|\mathcal{O}| \times |\mathcal{D}|}$, $\widehat{\mathfrak{b}} \in \mathbb{R}^{|\mathcal{D}| \times |\mathcal{D}| \times |\mathcal{U}|}$, and $\widehat{\mathfrak{d}} \in \mathbb{R}^{|\mathcal{D}|}$ serve the same role as $\mathfrak{a}$, $\mathfrak{b}$, and $\mathfrak{d}$ in defining Dirichlet distributions, while being treated as variational parameters to be optimized. Accordingly, the full VFE objective for the generative model (25) is given by:

$$\mathcal{F}_{t_c:t_p} = \mathbb{E}_{Q(D_{t_c:t_p}, \mathfrak{A}, \mathfrak{B}, \mathfrak{D}, \pi)} \left[\ln Q(D_{t_c:t_p}, \mathfrak{A}, \mathfrak{B}, \mathfrak{D}, \pi) - \ln p(O_{t_c:t_p}, D_{t_c:t_p}, \mathfrak{A}, \mathfrak{B}, \mathfrak{D}, \pi) \right], \tag{30}$$

which, using Eq. (25) and Eq. (26), can be factorized as:

$$\begin{aligned} \mathcal{F}_{t_c:t_p} = \mathbb{E}_{Q(D_{t_c:t_p}, \mathfrak{A}, \mathfrak{B}, \mathfrak{D}, \pi)} \big[& \ln Q(\mathfrak{A}) - \ln p(\mathfrak{A}) + \ln Q(\mathfrak{B}) - \ln p(\mathfrak{B}) + \ln Q(\mathfrak{D}) - \ln p(\mathfrak{D}) \\ & + \ln Q(\pi) - \ln p(\pi) - \ln p(D_{t_c}; \mathfrak{D}) + \ln Q(D_{t_c:t_p} \mid \pi) \\ & - \ln p(D_{t_c+1:t_p} \mid D_{t_c:t_p-1}, \pi; \mathfrak{B}) - \ln p(O_{t_c:t_p} \mid D_{t_c:t_p}; \mathfrak{A}) \big]. \end{aligned} \tag{31}$$

The update rules for the Dirichlet parameters $\mathfrak{a}$, $\mathfrak{b}$, and $\mathfrak{d}$ are derived by independently setting the gradients of the VFE (31) to zero with respect to each parameter direction. Specifically, when the ADT is being updated at the current time step t_c, learning proceeds as follows:

- **A** array: Given the Dirichlet prior parameters $\mathfrak{a}$ over the generative model, the observation $O_{t_c}^{\text{Exp}} = o_{t_c}^{\text{Exp}}$, and the digital state posterior $Q^*(D_{t_c})$, the fixed-point update rule for the variational posterior Dirichlet parameters $\widehat{\mathfrak{a}}$ over $Q(\mathfrak{A})$ is:

$$\widehat{\mathfrak{a}}^* = \mathfrak{a} + \eta^{\mathbf{A}} (o_{t_c}^{\text{Exp}} \otimes Q^*(D_{t_c})), \tag{32}$$

 where $\otimes$ denotes the outer product, and $\eta^{\mathbf{A}} \in \mathbb{R}$, with $0 \le \eta^{\mathbf{A}} \le 1$, is a learning rate parameter that scales the update step.

- **B** array: Given the Dirichlet prior parameters $\mathfrak{b}$ over the generative model, the digital state posterior $Q^*(D_{t_c})$, the previous digital state posterior $Q^*(D_{t_c-1})$, and the action $U_{t_c-1} = u_{t_c-1}$ taken at the previous time step, the fixed-point update rule for the variational posterior Dirichlet parameters $\widehat{\mathfrak{b}}$ over $Q(\mathfrak{B})$ is:

$$\widehat{\mathfrak{b}}^*_{\bullet,\bullet,u_{t_c-1}} = \mathfrak{b}_{\bullet,\bullet,u_{t_c-1}} + \eta^{\mathbf{B}}(Q^*(D_{t_c}) \otimes Q^*(D_{t_c-1})), \tag{33}$$

 which corresponds to an update applied to the u_{t_c-1}th slice of $\widehat{\mathfrak{b}}$.

- **d** array: Given the Dirichlet prior parameters $\mathfrak{d}$ and the digital state posterior $Q^*(D_{t_c})$, the fixed-point update rule for the variational posterior Dirichlet parameters $\widehat{\mathfrak{d}}$ over $Q(\mathfrak{D})$ is:

$$\widehat{\mathfrak{d}}^* = \mathfrak{d} + \eta^{\mathbf{d}} Q^*(D_{t_c}). \tag{34}$$

3.5. Epistemic behavior of active digital twins

If the agent also maintains a variational posterior over the model hyperparameters $Q(\boldsymbol{\phi})$, as discussed in Sec. 3.4, the EFE expression (10) can be extended to capture the epistemic value associated with the expected information gain not only over digital states but also over $\boldsymbol{\phi}$:

$$
\begin{aligned}
G_t^{\pi} &= \mathbb{E}_{Q(O_t, D_t, \boldsymbol{\phi}|\pi)} \left[\ln Q(D_t, \boldsymbol{\phi} \mid \pi) - \ln \widetilde{p}(O_t, D_t, \boldsymbol{\phi} \mid \pi) \right] \\
&= - \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\mathrm{D}_{\mathrm{KL}} \left[Q(D_t \mid O_t, \pi) \mid\mid Q(D_t \mid \pi) \right] \right]}_{\text{Epistemic value (digital state information gain)}} - \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\mathrm{D}_{\mathrm{KL}} \left[Q(\boldsymbol{\phi} \mid O_t, \pi) \mid\mid Q(\boldsymbol{\phi} \mid \pi) \right] \right]}_{\text{Epistemic value (model parameters information gain)}} \\
&\quad - \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\ln \widetilde{p}(O_t) \right]}_{\text{Pragmatic value (utility)}} + \underbrace{\mathbb{E}_{Q(O_t|\pi)} \left[\mathrm{D}_{\mathrm{KL}} \left[Q(D_t, \boldsymbol{\phi} \mid O_t, \pi) \mid\mid p(D_t, \boldsymbol{\phi} \mid O_t, \pi) \right] \right]}_{\text{Expected variational approximation error } (\geq 0)} .
\end{aligned}
\tag{35}
$$

The full derivation is provided in Appendix A, adapted to the ADT framework from [63]. The second epistemic term quantifies the value of resolving uncertainty over the Dirichlet parameters that govern the prior and posterior distributions of the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{d}$ arrays. When the AIF agent maintains and updates beliefs over these model parameters, this term steers policy inference toward action-observation trajectories expected to yield informative updates to the generative model. We point out that the epistemic value over $\boldsymbol{\phi}$ is not exploited in the numerical demonstrations presented in Sec. 4. Nevertheless, it is retained in the formulation of the ADT framework, as this capability may enable essential functionalities depending on the context and specific application objectives.

Epistemic actions aimed at refining the generative model can be regarded as forms of autonomous calibration, wherein the ADT steers its operation into underexplored regimes or perturbs its environment to test and improve its generative model. For instance, a sensor might be temporarily activated solely to evaluate its reliability while updating a likelihood model deemed unreliable. This behavior underscores the distinction between passive and active learning: while passive learning entails assimilating externally provided or randomly encountered data, active learning reflects the strategic initiation of data acquisition to accelerate model refinement and enhance future decision-making. The information-seeking (epistemic) behavior of ADTs thus emerges from their capacity for self-adaptive inference and learning, pursued alongside goal-directed (pragmatic) objectives. This dual optimization is embedded in policy inference through EFE minimization, which unifies goal-directed exploration and utility maximization within a single computational framework that moves beyond the passive replication of physical systems.

3.6. Algorithmic description

An algorithmic description of a single step of the AIF loop for ADTs is provided in Algorithm 1. Given the generative model, an observation sampled from the generative process, the posterior over digital states from the previous time step, and the action taken at the previous time step, one step of the loop involves: (1) performing inference over digital states based on the new observation; (2) using the posterior belief over digital states to perform policy inference and select the next action; (3) updating the generative model through learning informed by inference results.

4. Numerical demonstrations

This section demonstrates the proposed methodology through the simulated monitoring, management, and maintenance planning of the Hörnefors railway bridge [64]. Although this case study focuses specifically on structural health monitoring (SHM), the underlying framework broadly applies to a wide range of systems or domains.

Section 4.1 introduces the monitored physical asset. Section 4.2 describes the composition of the handled vibration data and the numerical models used to generate labeled examples under various damage scenarios. Section 4.3 outlines the assimilation of observational data for structural health identification using artificial neural networks. Section 4.4 details the step-by-step construction of the AIF generative model, namely the four-tuple $\langle \mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{d} \rangle$. Section 4.5 presents the results of ADT simulations under purely goal-directed behavior, serving as a baseline for comparison with the simulations involving mixed pragmatic-epistemic behavior, subsequently discussed in Sec. 4.6.

Algorithm 1 Active inference loop for active digital twins.

---

input: generative model $\langle \mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{d} \rangle$
  assimilated observation $O_{t_c}^{\mathrm{Exp}} = o_{t_c}^{\mathrm{Exp}}$
  digital state posterior $Q^*(D_{t_c-1})$ at previous time step
  action $U_{t_c-1} = u_{t_c-1}$ executed at previous time step

  ▷ digital state inference by minimizing variational free energy $\mathcal{F}_{t_c}$
1: infer digital state posterior $Q^*(D_{t_c})$

  ▷ policy inference and action selection by minimizing future variational free energy $\mathcal{F}_{t_c:t_p}$
2: compute posterior predictive distributions $Q(O_{t_c:t_p}, D_{t_c:t_p} \mid \pi)$
3: evaluate epistemic and pragmatic values over $t = t_c, \ldots t_p$ under each policy
4: infer control policies posterior $Q^*(\pi)$
5: select action $U_{t_c} = u_{t_c}$ by taking the best-point estimate or sampling from $Q^*(U_{t_c})$

  ▷ learning by minimizing variational free energy $\mathcal{F}_{t_c}$
6: update observation model $\mathbf{A}$ by computing the variational posterior Dirichlet parameters $\widehat{\mathfrak{a}}$
7: update transition model $\mathbf{B}$ by computing the variational posterior Dirichlet parameters $\widehat{\mathfrak{b}}$
8: update initial prior $\mathbf{d}$ by computing the variational posterior Dirichlet parameters $\widehat{\mathfrak{d}}$

  return updated generative model $\langle \mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{d} \rangle$
    updated posterior distribution over control policies $Q^*(\pi)$
    control action to be executed $U_{t_c} = u_{t_c}$
    posterior predictive density over digital states $Q(D_{t_c:t_p})$
    posterior predictive density over actions $Q(U_{t_c:t_p})$

The AIF agents based on discrete, Markovian generative models have been simulated using the open-source `Python` package `pymdp` library [63]. Compared to other AIF libraries, such as the `MATLAB` toolbox `DEM` [65] and the `C++` library `cpp-AIF` [66], `pymdp` offers notable advantages in terms of user-friendliness, flexibility, and customizability, although featuring lower process representation (`DEM`) and less computational efficiency (`cpp-AIF`). The simulations have been run on a PC featuring an `Intel®` `Core™` `i9-14900KF` CPU @ 3.2 GHz and 64 GB RAM.

4.1. Physical asset

The Hörnefors railway bridge, shown in Fig. 5(a), is an integral reinforced concrete structure along the Swedish Bothnia line. It spans 15.7 m, with a clearance height of 4.7 m and a width of 5.9 m (excluding edge beams). The main structural elements have a thickness of 0.5 m for the deck, 0.7 m for the frame walls, and 0.8 m for the wing walls. The foundation system comprises two slabs connected by stay beams, supported by pile groups. The concrete is of grade C35/45, characterized by the following material properties: Young's modulus $E = 34$ GPa, Poisson's ratio $\nu = 0.2$, and density $\rho = 2500$ kg/m^3. The bridge supports a single railway track with sleepers spaced at 0.65 m intervals, resting on a ballast layer that is 0.6 m deep and 4.3 m wide, with a density of $\rho_B = 1800$ kg/m^3. The structure is subjected to dynamic loading from *Gröna Tåget* trains operating at speeds between $v \in [160, 215]$ km/h. We specifically consider configurations involving two-car trainsets, totaling eight axles, with each axle bearing a mass of $\psi \in [16, 22]$ ton. The geometrical and mechanical parameters, as well as the moving load model, are adapted from [67]. The physical state space $\mathcal{S}$ represents the ground-truth variability in the bridge structural health.

4.2. Offline data assembly

The bridge monitoring system provides displacement data in the form of multivariate time series, denoted as $\mathbf{U}(\boldsymbol{\mu}) = [\mathbf{u}_1(\boldsymbol{\mu}), \ldots, \mathbf{u}_{N_s}(\boldsymbol{\mu})] \in \mathbb{R}^{L \times N_s}$. These consist of $N_s = 10$ individual time series corresponding to the degrees of freedom (dofs) indicated in Fig. 5(b). Each series contains L samples equally spaced over the time interval $[0, 1.5 \text{ s}]$, acquired with a sampling frequency of 400 Hz. The vector $\boldsymbol{\mu} \in \mathbb{R}^{N_{\mathrm{Par}}}$ collects N_{par}
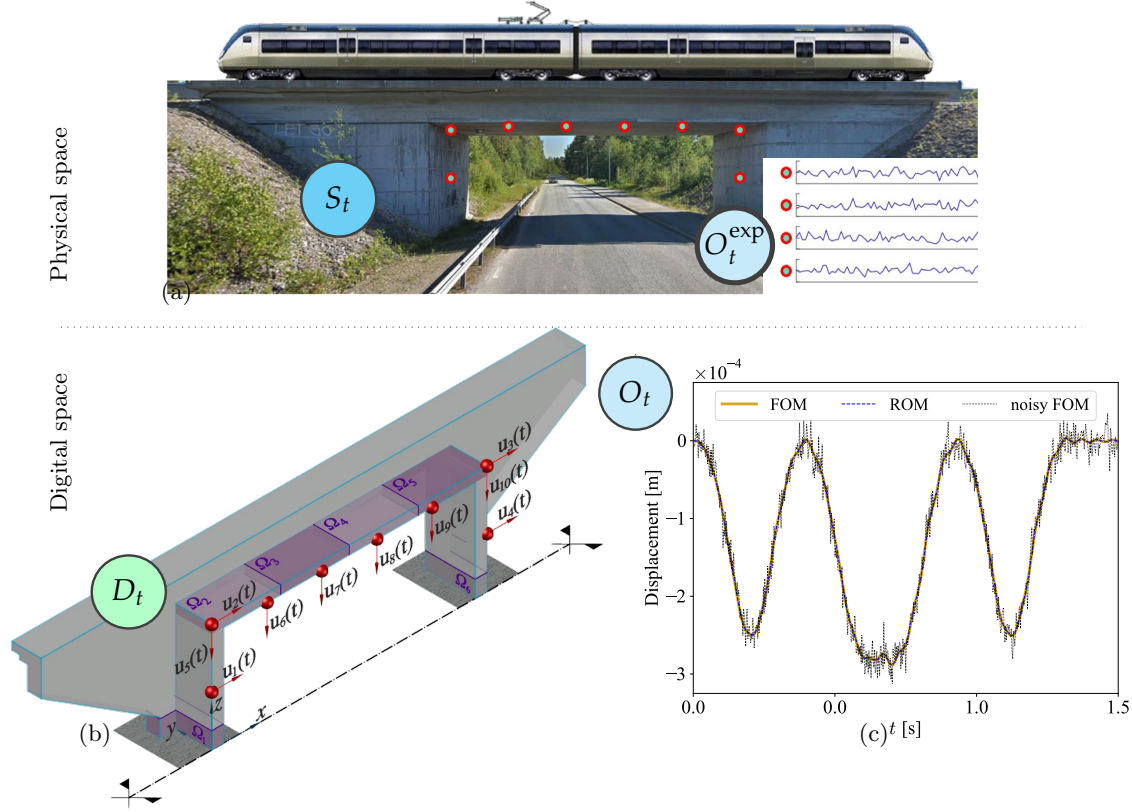
Figure 5: Physical asset and its digital twin. (a) The physical space corresponds to the Hörnefors bridge. (b) The digital space represents a structural health monitoring schematization, including details of synthetic recordings related to displacements $u_1(t), \ldots, u_{10}(t)$, and predefined damage regions $\Omega_1, \ldots, \Omega_6$. (c) Exemplary vertical displacement time history at midspan, comparing full-order model (FOM), reduced-order model (ROM), and noisy FOM approximations.

control parameters, which are assumed to represent the operational and damage conditions. For the problem settings we consider, each observation spans a relatively short time interval, within which these conditions are regarded as constant.

We simulate the monitored asset using a physics-based computational model. Specifically, the structure is modeled as a linear-elastic continuum under the assumption of linearized kinematics, and the equations of elasto-dynamics describe its dynamic response to train transits. The model is spatially discretized using linear tetrahedral finite elements, and its solution is advanced in time to generate synthetic observational data, controlled by the parameter vector $\boldsymbol{\mu}$.

The full-order model (FOM) is described in detail in [10]; here, we summarize its key features. The finite element mesh consists of elements with a nominal size of 0.8 m, refined to 0.15 m along the deck, resulting in a total of 17,292 dofs. The ballast layer is accounted for by increasing the density of the deck and edge beams to represent an equivalent mass. Embankment effects are captured using distributed springs applied along the surfaces in contact with the ground, implemented via a Robin-type boundary condition with an elastic coefficient of 10^8 N/m^3. Structural damping is introduced using Rayleigh damping, calibrated to yield a 5% damping ratio in the first two vibrational modes. The dynamic response is computed over the time interval [0, 1.5 s], uniformly partitioned into $L = 600$ time steps, using an implicit Newmark time integration scheme [68].

Damage-induced variations in the structural dynamic response are modeled as localized reductions in effective stiffness. Assuming that each observation spans a time window short enough compared to the timescale of damage progression, the structural behavior can be treated as linear within that interval. This enables a separation of timescales between the slow evolution of damage and the structural health assessment [69]. While the precise damage mechanisms are typically confirmed through on-site inspections following early detection, the degradation patterns in integral bridges that can be described in this way include:

cracking in concrete due to thermal gradients, freeze-thaw cycles, or overloading; progressive deterioration from alkali-silica reactions, which may lead to cracking and spalling; cracking from stress concentrations caused by differential settlements; and surface erosion from prolonged environmental exposure.

The digital state space $\mathcal{D}$ includes a set of predefined configurations of damage presence, location, and severity. These are modeled by parametrizing the stiffness matrix using two variables $y \in \mathbb{N}$ and $\delta \in \mathbb{R}$, both included in the parameter vector $\boldsymbol{\mu}$. The discrete variable $y \in \{0, \ldots, 6\}$ designates the damage region, with $y = 0$ denoting the undamaged baseline. For the damage cases $y = 1, \ldots, 6$, we consider $N_\Omega = 6$ predefined subdomains Ω_m, for $m = 1, \ldots, 6$, each representing a potential damage location as shown in Fig. 5. Within each subdomain, the material stiffness may be reduced by a factor $\delta \in [30\%, 80\%]$, which remains constant throughout the passage of a train.

To reduce the computational cost of solving the FOM for arbitrary values of $\boldsymbol{\mu}$, we employ a projection-based reduced-order model (ROM). The reduction is performed using a Galerkin reduced basis method [70, 71], relying on a low-dimensional set of basis functions computed through proper orthogonal decomposition. Following the method of snapshots [72], the ROM is constructed upon 400 FOM solutions for different configurations of the input parameters $\boldsymbol{\mu} = (v, \psi, y, \delta)^\top$, which are taken as uniformly distributed and sampled via the Latin hypercube rule. The dimension of the reduced-order expansion is determined based on an energy retention criterion. By setting a tolerance of 10^{-3} for the fraction of discarded energy, the number of dofs is reduced to 133. Both the FOM and ROM have been implemented in the `Matlab` environment, using the `redbKIT` library [73]. For a more detailed description, the reader is referred to [10].

A representative example of displacement time histories is reported in Fig. 5(c), showing the vertical displacement at midspan obtained from both the FOM and ROM. To emulate measurement noise and assess its potential impact on the handled structural response, signals are corrupted with additive Gaussian noise, yielding a signal-to-noise ratio of 120.

4.3. Data assimilation via artificial neural networks

The vibration recordings are assimilated for structural health diagnostics by leveraging the flexibility of deep learning (DL) models for SHM applications, as demonstrated in [74–76].

Data-driven approaches to SHM follow a pattern recognition paradigm [62], in which damage is assessed by comparing measurements with data previously collected under known structural conditions. This process relies on two key components: (i) feature selection and extraction, and (ii) statistical modeling to associate these features with specific damage patterns [77]. A major challenge lies in identifying damage-sensitive features that remain robust under varying operational and environmental conditions. DL offers an automated alternative for selecting and extracting optimized features by capturing temporal correlations within and across time series data [78, 79].

In our framework, each time a train crosses the bridge, the vibration recordings $\mathbf{U}$ are initially processed by a DL classifier, which outputs confidence scores indicating the likelihood that $\mathbf{U}$ corresponds to each damage class defined by the y parameter. The class with the highest confidence is selected as the best-point estimate for categorizing the measurements. Whenever damage is detected and localized within a region Ω_m, $m = 1, \ldots, 6$, the vibration recordings $\mathbf{U}$ are further processed by a dedicated regression model – one for each damageable region – to estimate the severity of damage δ. These initial estimates are then incorporated into the AIF framework as assimilated observation $O_{t_c}^{\mathrm{Exp}} = o_{t_c}^{\mathrm{Exp}}$, as detailed below.

The DL architectures have been implemented through the `Tensorflow`-based `Keras` API [80], and trained on a single `Nvidia GeForce RTX`™ 3080 GPU card. The training has been performed in a supervised fashion using $10,000$ noisy data instances generated from ROM simulations. For a comprehensive description the reader is referred to [10].

4.4. Active digital twin framework

The outcomes from the DL models are integrated into our POMDP framework by discretizing the range of δ into $N_\delta = 6$ intervals: $\{[30\%, 35\%], [35\%, 45\%], [45\%, 55\%], [55\%, 65\%], [65\%, 75\%], [75\%, 80\%]\}$. This discretization results in a total of $N_\Omega N_\delta + 1 = 37$ possible damage scenarios, each specifying a combination of damage location and severity. By ordering them first by location and then by severity, this post-processed output constitutes the first observation modality $O^{\Omega\delta}$.

The observation space $\mathcal{O}$ is completed with a second observation modality O^u corresponding to the action taken prior to data assimilation, such that $O = \{O^{\Omega\delta}, O^u\}$. Including this additional perceptual channel

provides two key benefits. First, it enables prior preferences (via the **c** array) to account not only for the costs associated with structural health states but also for those linked to actions. Second, as detailed below, it naturally supports the formulation of an action-conditioned observation model, introducing an inductive bias that facilitates digital state identification.

The digital state space $\mathcal{D}$ is structured into three factors $D = \{D^\Omega, D^\delta, D^{\mathrm{Epi}}\}$, corresponding to: the damage location $D^\Omega = \{\Omega_1, \ldots, \Omega_6\}$; the discretized percentage reduction in material stiffness $D^\delta = \{0\%,$ $[30\%, 35\%], [35\%, 45\%], [45\%, 55\%], [55\%, 65\%], [65\%, 75\%], [75\%, 80\%]\}$; and an epistemic switch $D^{\mathrm{Epi}} = \{\texttt{Epi}, \texttt{Non-Epi}\}$, which indicates whether the AIF agent is likely to engage in information-seeking (epistemic) behavior. This third factor allows the agent to autonomously switch between acting as an active information seeker or as a utility maximizer that is confident in its beliefs. It is worth noting that this factorization is neither the only viable option nor necessarily the most appropriate. This reflects the inherent subjectivity involved in shaping the digital state space. For example, D^Ω and D^δ could have been merged into a single enumerated representation, similar to the one used for $O^{\Omega\delta}$, at the expense of increased computational complexity and reduced interpretability. Alternatively, a more expressive but computationally demanding option would involve defining six separate D^δ factors, one for each of the N_Ω damageable regions. Finally, note that including D^{Epi} is essential to enable epistemic behavior, as this factor leads to distinct observation models associated with the `Epi` and `Non-Epi` states, as discussed further below.

The action space $\mathcal{U}$ comprises four control actions, each producing specific effects:

1. *Do nothing* (DN): the structural health state evolves according to a stochastic deterioration process, while regular revenue is maintained.

2. *Maintenance* (MA): a high-cost maintenance intervention is executed to mitigate existing damage. Although this action improves the structural condition, it may not fully restore the system to a pristine (damage-free) state.

3. *Restrict operations* (RO): traffic is limited to lightweight trains with axle load below 18 ton, thereby reducing the rate of structural degradation. However, this also leads to a reduction in the revenue generated by the infrastructure.

4. *Read sensors* (RE): a moderate-cost, high-fidelity sensing action is performed to resolve uncertainty in the structural health state. This action provides high epistemic value by decreasing the entropy of the digital state posterior, thus increasing the mutual information between latent states and expected observations. This effect reflects the use of high-quality sensors, controlled forced vibration tests, or in-situ inspection. From the perspective of the generative model, performing an inspection is equivalent to reading vibration recordings from sensors, albeit with significantly higher information content and a corresponding higher cost.

The observation likelihood array $\mathbf{A} = \{\mathbf{A}^{\Omega\delta}, \mathbf{A}^u\}$ comprises two observations models: $\mathbf{A}^{\Omega\delta} \in \mathbb{R}^{|O^{\Omega\delta}| \times |D^\Omega| \times |D^\delta| \times |D^{\mathrm{Epi}}|}$ and $\mathbf{A}^u \in \mathbb{R}^{|O^u| \times |D^\Omega| \times |D^\delta| \times |D^{\mathrm{Epi}}|}$, respectively encoding the conditional sensory likelihoods $p(O^{\Omega\delta} \mid D^\Omega, D^\delta, D^{\mathrm{Epi}})$ and $p(O^u \mid D^\Omega, D^\delta, D^{\mathrm{Epi}})$ for the first and second observation modalities. Conceptually, these tensors are designed to answer two distinct questions: (i) what might the agent believe about the pre-classified signals? and (ii) what might the agent infer about its previous action?

The slice of $\mathbf{A}^{\Omega\delta}$ for the epistemic state, i.e., $p(O^{\Omega\delta} \mid D^\Omega, D^\delta, D^{\mathrm{Epi}} = \texttt{Epi})$, is denoted by $\mathbf{A}^{\Omega\delta}_{\mathrm{Epi}} \in \mathbb{R}^{|O^{\Omega\delta}| \times |D^\Omega| \times |D^\delta|}$. This observation model is derived from a confusion matrix that quantifies the offline (expected) performance of the DL models in identifying the digital state factors D^Ω and D^δ. The confusion matrix is interpreted as a CPT, where rows correspond to ground-truth responses and columns to predicted outcomes. The offline evaluation has been performed using 4000 noisy FOM solutions, achieving a classification accuracy of 91.39%. To mitigate the risk of inconsistencies due to zero-likelihood observations, i.e., evidence contradicting the confusion matrix, a small positive perturbation 10^{-5} is added to all entries of $\mathbf{A}^{\Omega\delta}_{\mathrm{Epi}}$ prior to normalization. An exemplary slice of $\mathbf{A}^{\Omega\delta}_{\mathrm{Epi}}$ associated with $p(O^{\Omega\delta} \mid D^\Omega = \Omega_4, D^\delta, D^{\mathrm{Epi}} = \texttt{Epi})$ is shown in Fig. 6a.

While $\mathbf{A}^{\Omega\delta}_{\mathrm{Epi}}$ serves as a relatively informative sensory likelihood, a higher-entropy likelihood is used to model the slice of $\mathbf{A}^{\Omega\delta}$ under the non-epistemic state, i.e., $p(O^{\Omega\delta} \mid D^\Omega, D^\delta, D^{\mathrm{Epi}} = \texttt{Non-Epi})$, denoted by $\mathbf{A}^{\Omega\delta}_{\mathrm{Non-Epi}} \in \mathbb{R}^{|O^{\Omega\delta}| \times |D^\Omega| \times |D^\delta|}$. This non-epistemic model is obtained via uniform random perturbation of
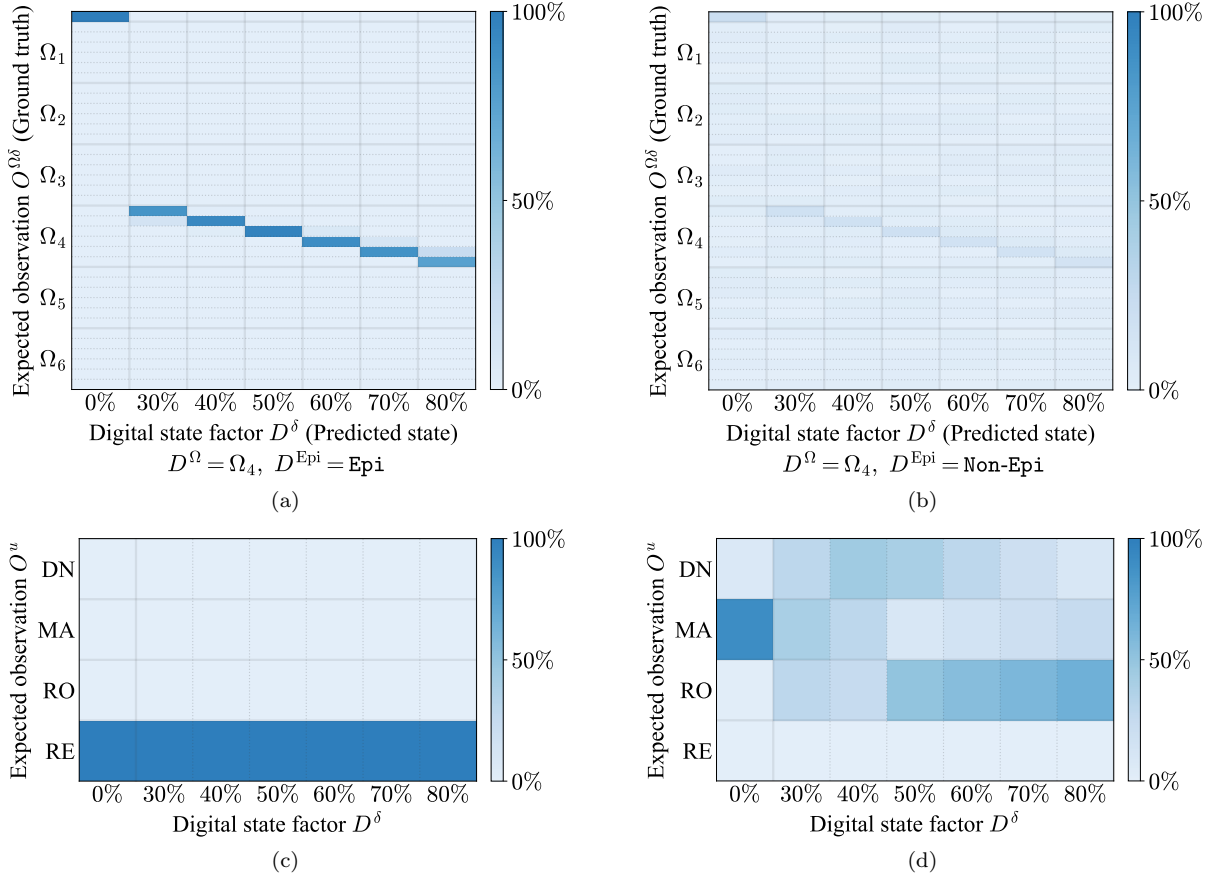
Figure 6: Visualization of the observation models: Panels (a) and (b) show slices of $\mathbf{A}^{\Omega\delta}$, corresponding to the sensory likelihoods (a) $p(O^{\Omega\delta} \mid D^{\Omega} = \Omega_4, D^{\delta}, D^{\mathrm{Epi}} = \mathtt{Epi})$ and (b) $p(O^{\Omega\delta} \mid D^{\Omega} = \Omega_4, D^{\delta}, D^{\mathrm{Epi}} = \mathtt{Non\text{-}Epi})$ for $\alpha = 0.2$. Panels (c) and (d) show slices of $\mathbf{A}^{u}$, corresponding to the sensory likelihoods (c) $p(O^{u} \mid D^{\Omega} = \Omega_{1:6}, D^{\delta}, D^{\mathrm{Epi}} = \mathtt{Epi})$ and (d) $p(O^{u} \mid D^{\Omega} = \Omega_{1:6}, D^{\delta}, D^{\mathrm{Epi}} = \mathtt{Non\text{-}Epi})$.

$\mathbf{A}^{\Omega\delta}_{\mathrm{Epi}}$ as the following linear combination:

$$\mathbf{A}^{\Omega\delta}_{\mathrm{Non\text{-}Epi}} = (1 - \alpha)\mathbf{A}^{\Omega\delta}_{\mathrm{Epi}} + \alpha\mathbf{A}^{\Omega\delta}_{\mathrm{Entropic}}, \tag{36}$$

which is then properly renormalized. Here, $\mathbf{A}^{\Omega\delta}_{\mathrm{Entropic}}$ is a purely entropic observation model sampled from a uniform distribution over $[0, 1]$, and $0 \leq \alpha \leq 1$ is a weighting coefficient controlling the degree of entropy introduced. Figure 6b shows an exemplary slice corresponding to $p(O^{\Omega\delta} \mid D^{\Omega} = \Omega_4, D^{\delta}, D^{\mathrm{Epi}} = \mathtt{Non\text{-}Epi})$ for $\alpha = 0.2$. It is worth noting that modulating α can also be interpreted as a simple yet effective mechanism to account both for potential errors in, and for the decision-maker confidence about, the use of DL models to assimilate real-world data.

The slice of $\mathbf{A}^{u}$ for the epistemic state, i.e., encoding $p(O^{u} \mid D^{\Omega}, D^{\delta}, D^{\mathrm{Epi}} = \mathtt{Epi})$, is denoted as $\mathbf{A}^{u}_{\mathrm{Epi}} \in \mathbb{R}^{|O^{u}| \times |D^{\Omega}| \times |D^{\delta}|}$. It is populated with Dirac delta distributions centered at the RE action for all possible combinations of D^{Ω} and D^{δ} (see also Fig. 6c). This design reflects the assumption that if the agent is in the state $D^{\mathrm{Epi}} = \mathtt{Epi}$, it knows with certainty that the previously taken action was the (epistemic) RE action, regardless of the values of the other digital state factors. From a data assimilation point of view, receiving $O^{u} = \mathrm{RE}$ provides no informative cues for inferring D^{Ω} or D^{δ}, but it deterministically sets $D^{\mathrm{Epi}} = \mathtt{Epi}$. In contrast, under the non-epistemic state, the corresponding observation model $\mathbf{A}^{u}_{\mathrm{Non\text{-}Epi}} \in \mathbb{R}^{|O^{u}| \times |D^{\Omega}| \times |D^{\delta}|}$ is filled with entries that reflect a plausible causality for what the agent can infer about the previous action given D^{Ω} and D^{δ}. This prior CPT (see also Fig. 6d) is modeled consistently across the D^{Ω} factor, as follows:

$$p(O^u \mid D^\Omega = \Omega_{1:6}, D^\delta, D^{\mathrm{Epi}} = \texttt{Non-Epi}) = \begin{pmatrix} 0.08 & 0.3 & 0.45 & 0.4 & 0.3 & 0.2 & 0.1 \\ 0.9 & 0.4 & 0.3 & 0.1 & 0.15 & 0.2 & 0.25 \\ 0.02 & 0.3 & 0.25 & 0.5 & 0.55 & 0.6 & 0.65 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{37}$$

For data assimilation, observing $O^u \neq \mathrm{RE}$ has two implications: first, it deterministically sets $D^{\mathrm{Epi}} = \texttt{Non-Epi}$; second, it introduces an inductive bias by leveraging the structure of $\mathbf{A}^u_{\mathrm{Non-Epi}}$ to condition inference on the previous action, similar to the influence of the transition model.

The transition array $\mathbf{B} = \{\mathbf{B}^\Omega, \mathbf{B}^\delta, \mathbf{B}^{\mathrm{Epi}}\}$ comprises three sub-arrays $\mathbf{B}^f \in \mathbb{R}^{|D^f| \times |D^f| \times |U^f|}$, each encoding the transition dynamics $p(D^f_t \mid D^f_{t-1}, u^f_{t-1}; \mathfrak{B}^f)$ of a specific digital state factor $D^f \in \{D^\Omega, D^\delta, D^{\mathrm{Epi}}\}$, conditioned on its previous state and the corresponding control factor $u^f \in \{u^\Omega, u^\delta, u^{\mathrm{Epi}}\}$. Starting with initial priors over the transition probabilities defined by $\mathfrak{B}^f$, these are iteratively refined by assimilating evidence from the system response to actions, as described in Sec. 3.4. A graphical visualization of the initial transition models for each digital state and control factor is shown in Fig. 7. Note that these internal models do not replicate the ground-truth evolution, which remains unknown to the ADT. Moreover, assuming digital state factors evolve independently, the control space is factorized as $U = \{U^\Omega = \varnothing, U^\delta = \overline{U}, U^{\mathrm{Epi}} = \overline{U}\}$. This reflects that D^Ω is an uncontrollable factor, with a control dimensionality of 1, while D^δ and D^{Epi} are both influenced by the same control variable $\overline{U} \in \mathcal{U} = \{\mathrm{DN}, \mathrm{MA}, \mathrm{RO}, \mathrm{RE}\}$. The set of feasible policies Π is constructed by combinatorially enumerating all possible sequences of actions from the action space $\mathcal{U}$ over the prediction horizon $t = t_c, \ldots, t_p$, resulting in a total of $4^{t_p - t_c}$ policies.

The initial Dirichlet parameters $\mathfrak{b}^\Omega$ over the categorical distribution $\mathfrak{B}^\Omega$ for the uncontrollable $\mathbf{B}^\Omega$ are selected to yield a 0.8 probability that damage stays in the same subdomain Ω_m, for $m = 1, \ldots, 6$. The remaining 0.2 probability is evenly distributed across the other subdomains, reflecting a strong prior belief that damage is unlikely to move between different regions.

For the action-conditioned $\mathbf{B}^\delta$, each action-specific slice encodes the probability of transitioning between discrete δ intervals. The diagonal entries represent the probability of remaining in the same damage state, while the lower-left and upper-right triangles denote the probabilities of deterioration and improvement, respectively. Under the DN action, the initial Dirichlet parameters $\mathfrak{b}^\delta$ for the categorical distribution $\mathfrak{B}^\delta$ are configured to yield transition probabilities of 0.85, 0.1, and 0.05 for degradation of zero, one, or two δ intervals, respectively. For the RO action, the corresponding probabilities are set to 0.92, 0.05, and 0.03, reflecting a slower rate of deterioration due to reduced structural load. The slice associated with the RE action is designed to reflect improved damage tracking. It assigns probabilities of 0.9 and 0.1 for degradation of zero and one δ intervals, capturing the higher confidence associated with epistemic control actions. In contrast, the MA action slice is designed to support transitions across up to six δ intervals, with probabilities 0.05, 0.15, 0.20, 0.20, 0.20, and 0.20, for improvements of zero to five intervals, respectively. To mitigate the risk of numerical inconsistencies caused by evidence that contradicts the assumed transition dynamics, a small perturbation of 10^{-3} is added to all entries of $\mathbf{B}^\delta$ prior to normalization.

The sub-array $\mathbf{B}^{\mathrm{Epi}}$ serves as an epistemic switch, enabling deterministic transitions between the states $D^{\mathrm{Epi}} = \texttt{Epi}$ and $D^{\mathrm{Epi}} = \texttt{Non-Epi}$. This mechanism is implemented through Boolean matrices that enforce $D^{\mathrm{Epi}} = \texttt{Non-Epi}$ – regardless of its previous value – whenever the ADT selects DN, MA, or RO actions. Conversely, selecting the RE action triggers a transition to the epistemic state $D^{\mathrm{Epi}} = \texttt{Epi}$. This transition model is not subject to learning updates, as its structure is predefined and not expected to benefit from interaction with the generative process.

At each time step, the ADT selects a control action $u_t \in \mathcal{U}$ whose effects on the generative process are uncertain and may lead to unexpected outcomes. The costs associated with both the structural health state and the control actions are modeled as prior preferences via the array $\mathbf{c} = \{\mathbf{c}^{\Omega\delta}, \mathbf{c}^u\}$. The components $\mathbf{c}^{\Omega\delta} \in \mathbb{R}^{|O^{\Omega\delta}|}$ and $\mathbf{c}^u \in \mathbb{R}^{|O^u|}$ assign relative log-probabilities to each outcome of the two observation modalities, respectively:

$$\mathbf{c}^{\Omega\delta} \leftarrow \ln \widetilde{p}(O_t^{\Omega\delta}) = \begin{cases} 0 & \text{if } y = 0, \\ -\exp(\delta) & \text{if } 30\% < \delta < 80\%, \\ -10 & \text{if } \delta = 80\%, \end{cases} \quad \mathbf{c}^u \leftarrow \ln \widetilde{p}(O_t^u) = \begin{cases} +5.5 & \text{if } u_t = \mathrm{DN}, \\ -5 & \text{if } u_t = \mathrm{MA}, \\ +2.5 & \text{if } u_t = \mathrm{RO}, \\ -0.5 & \text{if } u_t = \mathrm{RE}. \end{cases} \tag{38}$$
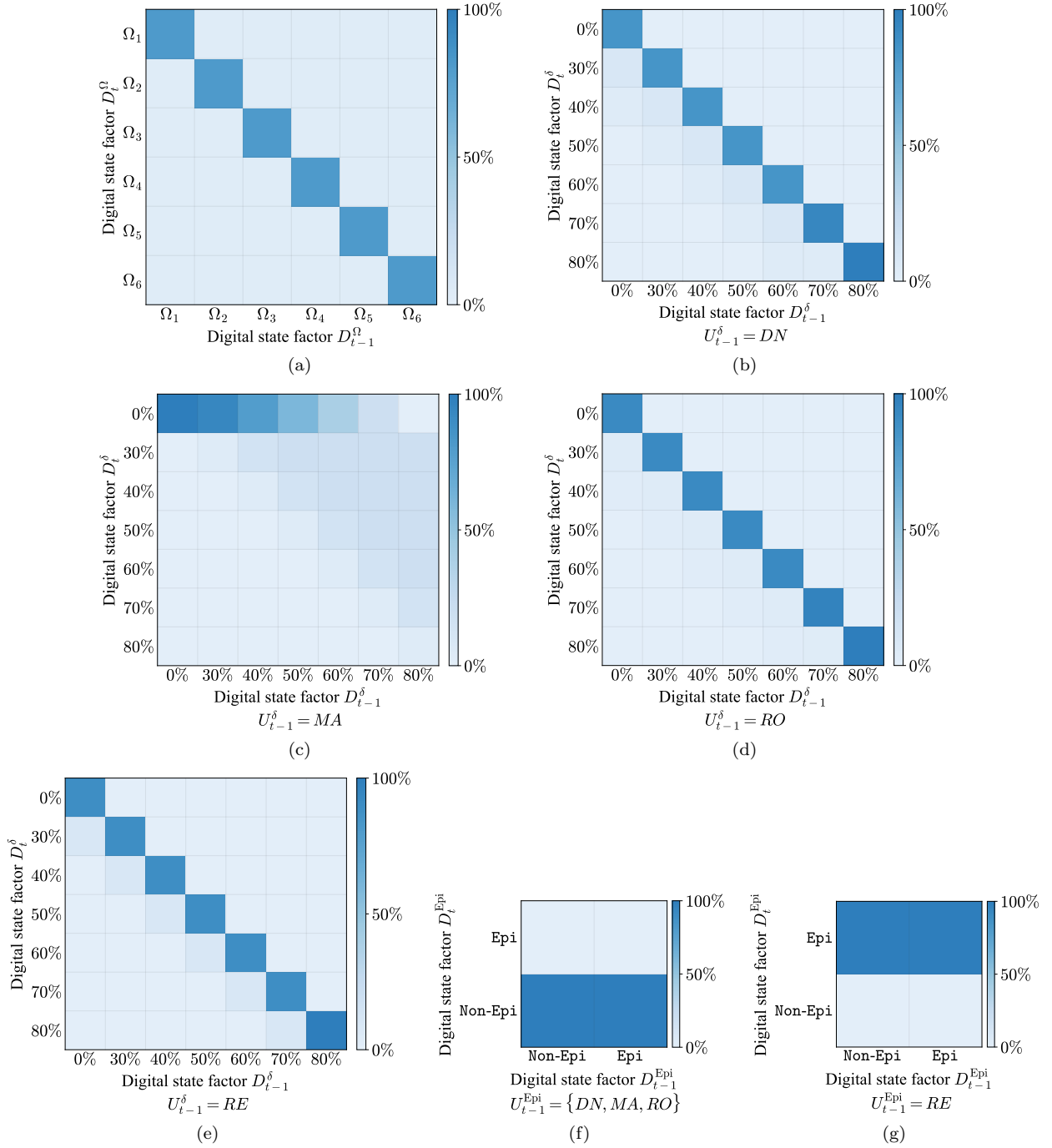
Figure 7: Visualization of the transition models: Panel (a) shows the uncontrollable $\mathbf{B}^{\Omega}$, corresponding to the transition likelihood $p(D_t^{\Omega} \mid D_{t-1}^{\Omega})$. Panels (b-e) show the action-specific slices of $\mathbf{B}^{\delta}$, corresponding to the transition likelihoods (b) $p(D_t^{\delta} \mid D_{t-1}^{\delta}, U_{t-1}^{\delta} = \text{DN})$, (c) $p(D_t^{\delta} \mid D_{t-1}^{\delta}, U_{t-1}^{\delta} = \text{MA})$, (d) $p(D_t^{\delta} \mid D_{t-1}^{\delta}, U_{t-1}^{\delta} = \text{RO})$, and (e) $p(D_t^{\delta} \mid D_{t-1}^{\delta}, U_{t-1}^{\delta} = \text{RE})$. Panels (f) and (g) shows the action-specific slices of $\mathbf{B}^{\text{Epi}}$, corresponding to the transition likelihoods (f) $p(D_t^{\text{Epi}} \mid D_{t-1}^{\text{Epi}}, U_{t-1}^{\text{Epi}} = \{\text{DN,MA,RO}\})$ and (g) $p(D_t^{\text{Epi}} \mid D_{t-1}^{\text{Epi}}, U_{t-1}^{\text{Epi}} = \text{RE})$.

These log-probability vectors are passed through a Softmax function to produce valid probability distributions $\widetilde{p}(O_t^{\Omega\delta})$ and $\widetilde{p}(O_t^u)$, which are then used to compute the expected utility term in the EFE. The structural health preferences penalize deterioration in proportion to the exponential of δ, with a steep penalty for severely compromised states. The control action preferences reflect trade-offs between epistemic value of

expected information gain and operational cost: DN and RO actions yield positive rewards but carry the risk of structural deterioration; RE similarly allows for deterioration, yet it is expected to reduce the entropy of the digital state posterior at the cost of a moderately negative reward. MA mitigates deterioration but carries a significantly negative reward due to its high cost. While these values are expressed in non-dimensional form, they represent indicative costs charged to the decision maker. Actual values may be derived from service and cost catalogs issued by governmental agencies or infrastructure operators. In particular, the health-related preference distribution $\widetilde{p}(O_t^{\Omega\delta})$ should reflect a prioritization analysis that accounts for both the likelihood and consequences of different damage scenarios – such as loss of serviceability, increased accident risk, or structural failure – as well as the risk tolerance of the decision-maker.

The array defining the initial state model $\mathbf{d} = \{\mathbf{d}^\Omega, \mathbf{d}^\delta, \mathbf{d}^{\mathrm{Epi}}\}$ consists of three sub-arrays $\mathbf{d}^f \in \mathbb{R}^{|D^f|}$, each specifying the initial prior distribution $p(D_{t_c}^f)$ over a digital state factor $D^f \in \{D^\Omega, D^\delta, D^{\mathrm{Epi}}\}$. Uniform probability distributions are adopted for D^Ω and D^{Epi} to reflect initial uncertainty. In contrast, D^δ is initialized as a Dirac delta distribution centered at 0%, consistent with the assumption of undamaged structure when the ADT enters into operation.

The (unknown) ground-truth generative process evolves conditionally on the most recent control action. In particular, we assume that damage can develop in any predefined region, without propagating across different damageable subdomains. The evolution follows the degradation (or improvement) stochastic models described below. Under the DN, RO, and RE actions, structural health is assumed to degrade monotonically. For the DN action, the damage class y is sampled from a categorical distribution $y \sim \mathrm{Cat}(\frac{1}{2}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12})$, which assigns half of the probability mass to the undamaged state $y = 0$, and distributes the remaining half uniformly among the six damage classes $y = 1, \ldots, 6$. When damage first initiates, the magnitude δ is sampled uniformly within the range of the first damage interval $\delta_t \mid y_t \neq 0, y_{t-1} = 0 \sim \mathrm{Uniform}(0.3\%, 0.35\%)$. Subsequent damage progression is modeled by sampling δ increments from a truncated normal distribution centered at 1.5% with a standard deviation of 1%, $\delta_t - \delta_{t-1} \mid y_{t-1} \neq 0 \sim \mathrm{Normal}_{\geq 0}(1.5\%, 1\%)$, with any increments below 0% rounded up to 0%. For the RO action, a similar model is employed, but with a lower probability of damage initiation and slower deterioration. In this case, the damage class is sampled as $y \sim \mathrm{Cat}(\frac{3}{4}, \frac{1}{24}, \frac{1}{24}, \frac{1}{24}, \frac{1}{24}, \frac{1}{24}, \frac{1}{24})$, and the damage magnitude evolves as $\delta_t - \delta_{t-1} \mid y_{t-1} \neq 0 \sim \mathrm{Normal}_{\geq 0}(0.95\%, 0.5\%)$. For the RE action, the generative process is the same as that under the DN or RO actions, respectively, depending on whether the system was previously in a restricted or unrestricted condition before engaging in information-seeking (epistemic) behavior. In contrast, the MA action is modeled as a healing process. If $y = 0$, the system remains undamaged. If $y \neq 0$, the damage magnitude decreases according to $\delta_t - \delta_{t-1} \mid y_t \neq 0 \sim \mathrm{Normal}_{\leq 10\%}(-25\%, 15\%)$, with any decrement below 10% rounded up to 10%. The system is assumed to return to an undamaged condition $(y = 0)$ if the resulting damage magnitude satisfies $\delta < 30\%$, reflecting a minimal detectable deterioration threshold.

4.5. Results: Purely goal-directed behavior

In this section and the next, we present the results of several ADT simulations, each spanning 60 time steps. At every time step, new observational data are generated based on the (unknown) ground-truth generative process. The ADT assimilates these data to infer the variational posterior $Q^*(D_{t_c})$ over the current digital state, and performs policy inference by computing the posterior $Q^*(\pi)$ over policies. Control actions are subsequently selected as the best-point estimate from the posterior $Q^*(U_{t_c})$ over control states, and the generative model is eventually learned by updating the variational posterior Dirichlet parameters $\widehat{\mathfrak{b}}^*$.

We adopt a policy horizon of $t_p - t_c = 4$ and begin by analyzing a baseline scenario where the ADT operates under a purely goal-directed (pragmatic) behavior. This is achieved by retaining only the utility term associated with pragmatic value in the EFE formulation, excluding any contributions from information-seeking (epistemic) value and removing the epistemic RE action from the available action set. The entropy level in the observation model $\mathbf{A}_{\mathrm{Non\text{-}Epi}}^{\Omega\delta}$ is set using $\alpha = 0.5$. The inverse temperature parameter controlling the precision of policy selection is left to its default value of $\gamma = 16$. Furthermore, learning updates to the generative model are disabled in this baseline setting.

Figure 8 illustrates a representative ADT simulation. Results are reported in terms of both the ground-truth physical state and the corresponding ADT estimates obtained after assimilating observational data. The evolution of the digital state is shown only for regions that experience damage, although all damageable
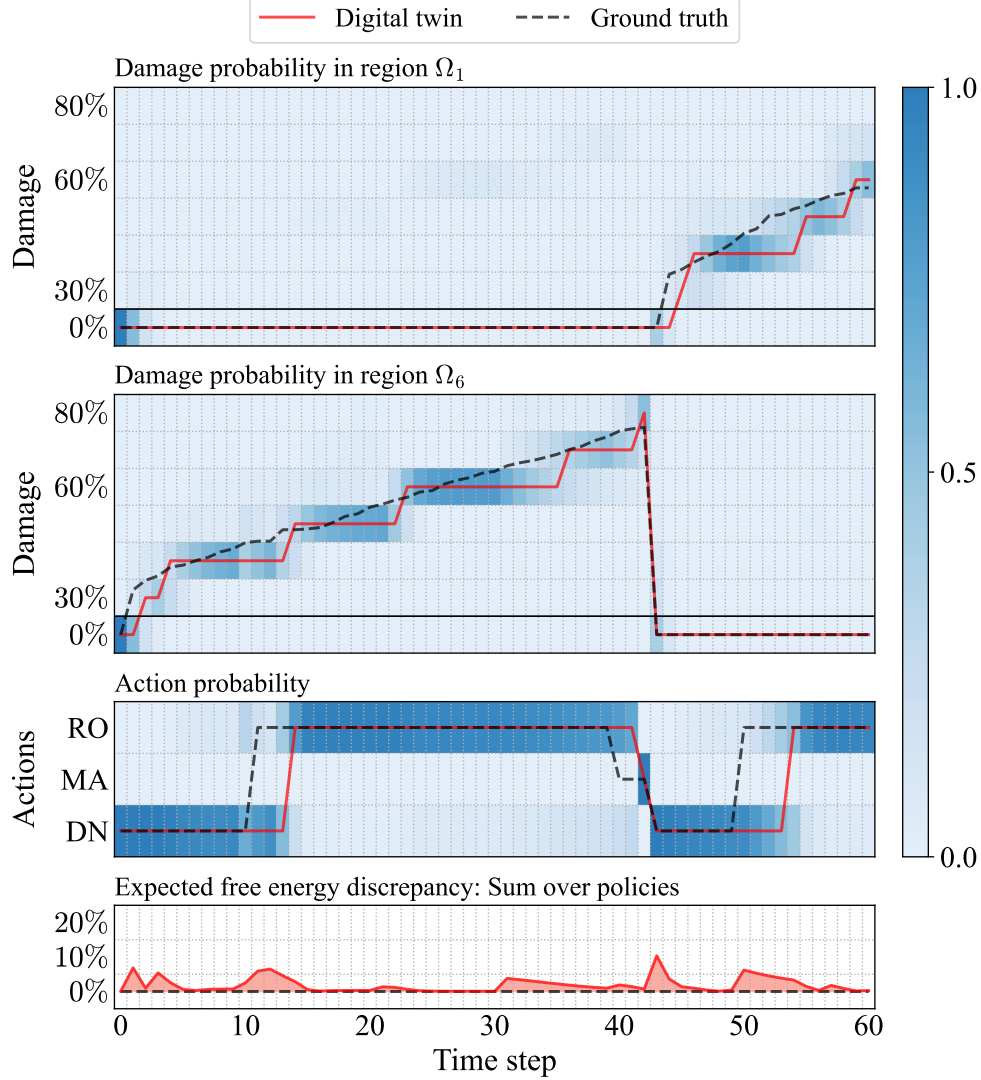
Figure 8: Active digital twin using purely goal-directed (pragmatic) behavior. Probabilistic and best-point estimates of: (top two panels) digital state evolution compared to the ground-truth physical state; (penultimate panel) control actions recommended by the digital twin versus the optimal action under the ground-truth generative process. In the top panels, background colors represent the belief distribution over the digital state at each time step. In the penultimate panel, background colors indicate the belief distribution over the control actions. The bottom panel quantifies simulation quality in terms of the percentage absolute discrepancy between the sum of the policy-specific expected free energies computed by the digital twin and those obtained under the ground truth.

regions $\Omega_1, \ldots, \Omega_6$ are susceptible to degradation. Initially, damage develops in Ω_1, and the posterior $Q^*(D_{t_c})$ reveals relatively high uncertainty, primarily due to the entropy in the observation model. Nevertheless, despite the severely corrupted observation model $\mathbf{A}_{\text{Non-Epi}}^{\Omega\delta}$, the ADT successfully follows the ground-truth evolution by leveraging prior information from the forward-time predictor $\mathbf{B}$. The corresponding sequence of control action estimates $Q^*(U_{t_c})$ is shown in the penultimate panel. The ADT initially recommends DN actions, aligned with the prior preferences over the two observation modalities encoded in $\mathbf{c}$, i.e., to maximize utility. Once a substantial probability mass in $Q^*(D_{t_c})$ is assigned to $D^\delta \geq 45\%$, RO actions begin to be selected, enabling the ADT to continue monitoring degradation, which now evolves at a reduced rate. Eventually, an MA action is selected when the structural state becomes critically compromised, as indicated by a consistent probability mass over $D^\delta \geq 75\%$ in $Q^*(D_{t_c})$. A similar behavior is shown for the subsequent damage event in Ω_6.

24

For comparison, control actions under the ground-truth generative process are computed using a second AIF agent that mirrors the ADT architecture but has access to the true physical state. The ADT selects the appropriate control action from $Q^*(U_{t_c})$ with a delay of at most five time steps relative to the ground-truth-informed agent. This delay is mainly attributed to the continued use of a highly entropic observation model, which limits fast and accurate inference, along with the need to recursively update prior beliefs from earlier time steps. Note that including the RE action in the available action set would not affect the results in this case, as the ADT is driven solely by (pragmatic) utility maximization and does not engage in information-seeking (epistemic) behavior, i.e., it does not seek to reduce the entropy of $Q(D_{t_c:t_p})$ through exploratory actions. The bottom panel of the figure assesses simulation quality by tracking the evolution of the percentage absolute discrepancy between the sum of the policy-specific EFEs computed by the ADT and those obtained under the ground-truth-informed agent:

$$\Delta^{\mathbf{G}} = \left| \frac{\sum_{\pi \in \Pi} (G^\pi - \widehat{G}^\pi)}{\sum_{\pi \in \Pi} \widehat{G}^\pi} \right| \cdot 100, \tag{39}$$

where $\widehat{G}^\pi$ denotes the EFE associated with policy π under the ground-truth-informed AIF agent.

A second representative ADT simulation is shown in Fig. 9, exemplifying the same purely goal-directed (pragmatic) behavior but with a different random seed. In this case, damage begins to develop in region Ω_3, and the ADT initially behaves consistently with the previous results, tracking the generative process with relatively high-entropy estimates propagated forward in time. However, starting from time step $t = 33$, the ADT begins to diverge from the ground truth, and the digital state posterior $Q^*(D_{t_c})$ progressively loses synchronization with the physical state. The probability mass in $Q^*(D_{t_c})$ gradually shifts from $D^\Omega = \Omega_3$ to $D^\Omega = \Omega_6$, where D^δ is consistently underestimated as lying within the range $[65\%, 75\%]$, while the actual value is in the range $[75\%, 80\%]$. As a result, the ADT fails to select an MA action for more than ten time steps, despite its necessity. The simulation eventually terminates at time step $t = 46$, due to a digital failure at $t = 47$, caused by $D^\delta > 80\%$, and symbolizing structural collapse. The ADT inability to recover accurate tracking is attributed to the interplay between the poorly informative sensory likelihood and the recursive propagation of outdated prior beliefs, which degrade over time.

By running a cluster of 100 simulations, each spanning 60 time steps and initialized with a different random seed for both the observation model $\mathbf{A}^{\Omega\delta}_{\text{Non-Epi}}$ and the ground-truth generative process, the ADT operating under a purely goal-directed (pragmatic) behavior fails in 47 out of 100 cases. In this baseline setting, the failure rate is strongly driven by the high level of entropy introduced in the observation model $\mathbf{A}^{\Omega\delta}_{\text{Non-Epi}}$. Although highly corrupted observations realistically reflect many real-world conditions, the results in the following section show that equipping the ADT with both goal-directed and information-seeking (epistemic) components enables active exploration in response to critical uncertainty, resulting in a significant performance improvement over the purely pragmatic baseline.

The behavior described above can also be illustrated using a simplified scenario in which the ADT relies on two sensors, each providing partial observations to update its beliefs about the evolving damage state. If one sensor becomes faulty but the transition model closely approximates the actual dynamics of damage progression, the ADT may still track the system accurately, as the predictive power of the prior compensates for the degraded sensory evidence. However, in the more typical case where the transition model does not fully capture the actual system evolution, outdated priors dominate the inference process, and the compromised likelihood is unable to correct them. In such conditions, an information-seeking (epistemic) action should ideally be triggered to resolve ambiguity and restore confidence in the likelihood model – for instance, by querying a redundant sensor, activating a dormant one, or scheduling a targeted diagnostic procedure.

It is interesting to note how the EFE discrepancy shown in the bottom panel of Fig. 9 does not indicate any critical issue. This is because the EFE is a subjective metric, reflecting how the ADT evaluates its own performance rather than measuring the correctness of the digital state estimates. Indeed, the $\Delta^{\mathbf{G}}$ indicator quantifies the misalignment in belief-driven action planning between the ADT and an idealized agent with access to the true physical state. As a result, $\Delta^{\mathbf{G}}$ remains low simply because the ADT is (mistakenly) confident in its estimated behavior, just as the ground-truth-informed agent is confident in its own. However, the resulting control actions differ. To address this limitation, one could instead employ objective performance indicators derived from the generative process. Examples include utility scores evaluated on realized system outcomes, or the survival time before reaching a critical condition.
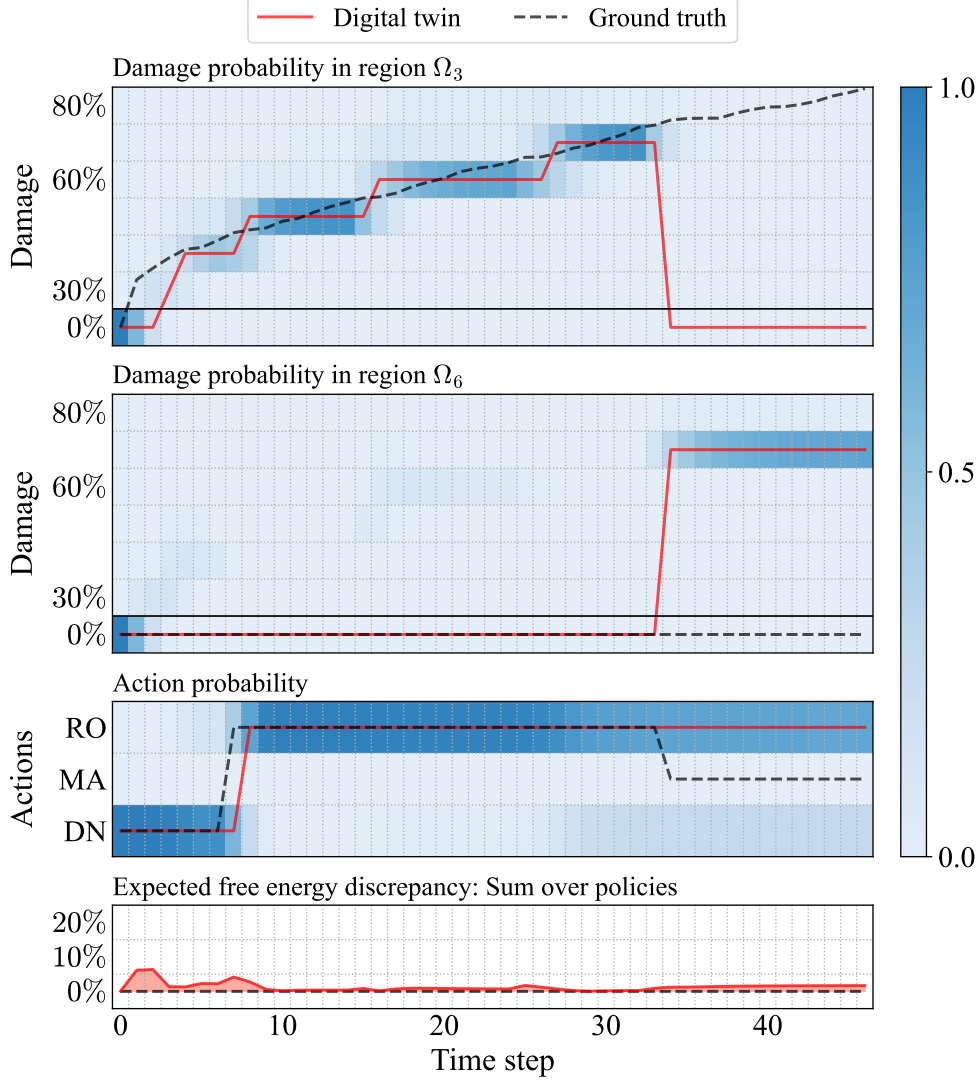
Figure 9: Failed active digital twin using purely goal-directed (pragmatic) behavior. Probabilistic and best-point estimates of: (top two panels) digital state evolution compared to the ground-truth physical state; (penultimate panel) control actions recommended by the digital twin versus the optimal action under the ground-truth generative process. In the top panels, background colors represent the belief distribution over the digital state at each time step. In the penultimate panel, background colors indicate the belief distribution over the control actions. The bottom panel quantifies simulation quality in terms of the percentage absolute discrepancy between the sum of the policy-specific expected free energies computed by the digital twin and those obtained under the ground truth.

4.6. Results: Combining goal-directed and information-seeking behaviors

In this section, we present the results of ADT simulations combining goal-directed (pragmatic) and information-seeking (epistemic) behaviors. For this, we adopt the complete EFE formulation including both pragmatic and epistemic terms; furthermore, we include the epistemic RE action in the available actions. All other settings remain unchanged from the simulations presented earlier.

Figure 10 illustrates a representative simulation. The ADT initially exhibits information-seeking (epistemic) behavior, executing a sequence of RE actions to gather information about damage onset. Once the posterior $Q^*(D_{t_c})$ identifies evolving damage within Ω_1 with relatively low uncertainty, the ADT shifts to DN actions aimed at (pragmatic) utility maximization. When a significant portion of $Q^*(D_{t_c})$ supports $D^\delta \geq 45\%$, RO actions begin to emerge. An MA action is eventually selected when the risk of structural failure becomes substantial, i.e., for a significant probability mass over $D^\delta \geq 65\%$. A similar pattern is observed during the subsequent damage event in Ω_6. In this case, sporadic RE actions are also triggered
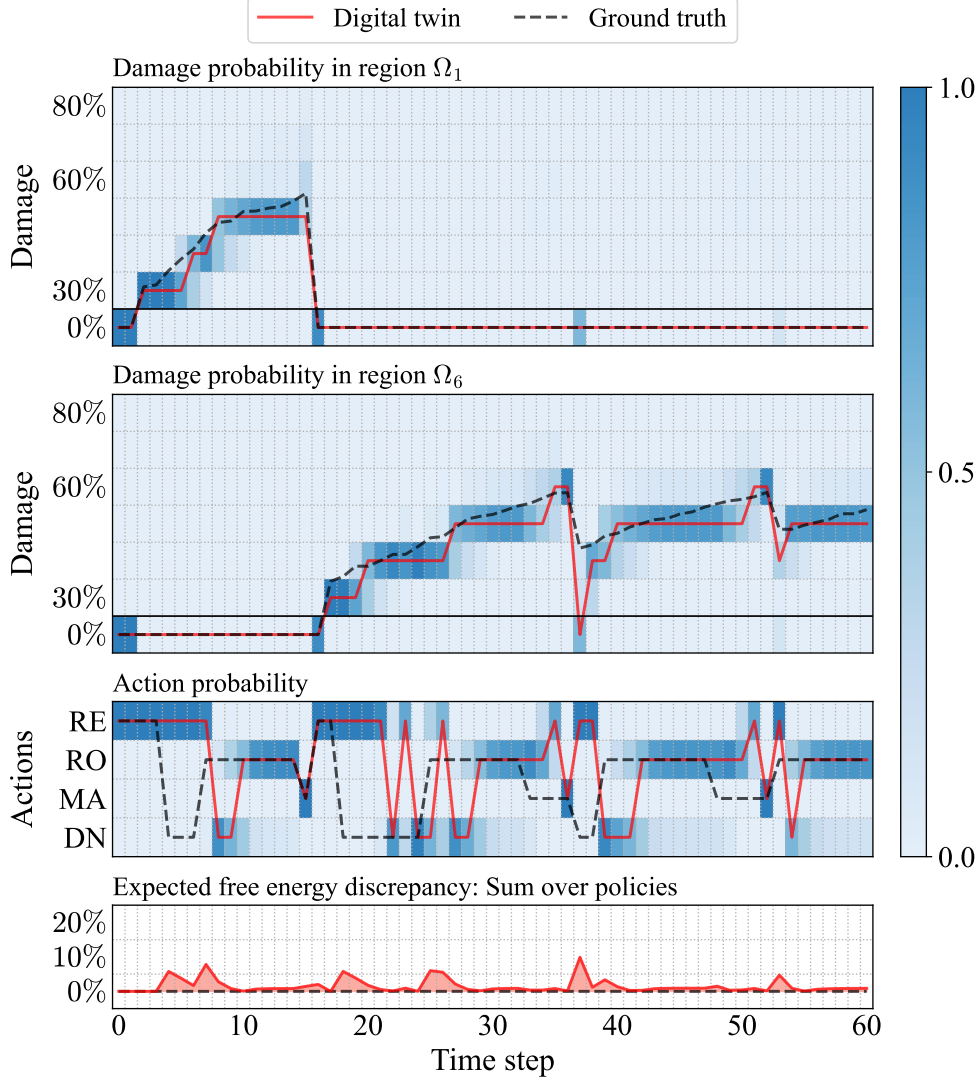
Figure 10: Active digital twin using a combination of goal-directed (pragmatic) and information-seeking (epistemic) behaviors. Probabilistic and best-point estimates of: (top two panels) digital state evolution compared to the ground-truth physical state; (penultimate panel) control actions recommended by the digital twin versus the optimal action under the ground-truth generative process. In the top panels, background colors represent the belief distribution over the digital state at each time step. In the penultimate panel, background colors indicate the belief distribution over the control actions. The bottom panel scores simulation quality in terms of the percentage absolute discrepancy between the sum of the policy-specific expected free energies computed by the digital twin and those obtained under the ground truth.

whenever the entropy of $Q^*(D_{t_c})$ increases, to prevent desynchronization from the physical state. These RE actions are interleaved with extended sequences of DN and RO decisions, depending on the evolving health state and the interaction between the sensory likelihood and the transition model. For instance, RE actions at $t = 35$ and $t = 51$ are deployed to disambiguate the digital state just before executing costly MA interventions. In contrast, the first MA action at $t = 15$ is not preceded by RE behavior, as the ADT maintains a confident, low-entropy belief at that point. Note that the epistemic RE action carries a lower prior preference $\widetilde{p}(O^u)$ than DN or RO actions and does not directly affect damage progression. As a result, it is employed only under epistemic-driven behavior, where its role is to support future goal-directed (pragmatic) decisions. Similarly, RE actions occurring immediately after MA interventions reflect the ADT effort to resolve uncertainty via active exploration of maintenance outcomes. In contrast, under the ground-truth generative process, RE actions are triggered exclusively to gather initial evidence about damage onset. Corrective inference is unnecessary in this setting due to perfect, uncertainty-free access to the physical state.
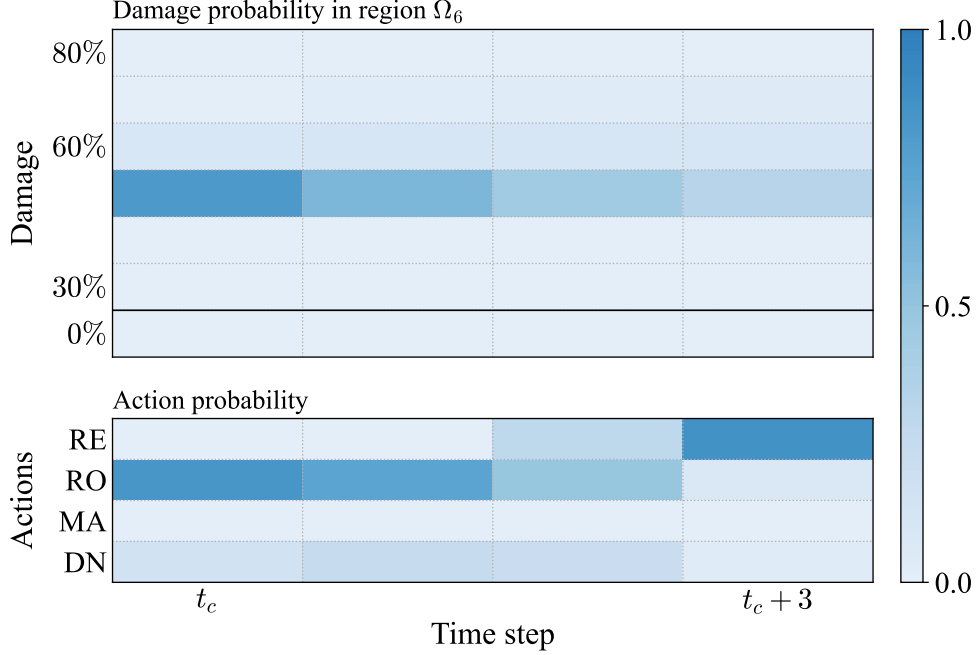
27

Figure 11: Active digital twin using a combination of goal-directed (pragmatic) and information-seeking (epistemic) behaviors. Posterior predictive densities beyond data assimilation over (future) digital states and control states, starting at $t_c = 60$. In the top panel, background colors represent the belief distribution over the digital state at each time step. In the bottom panel, background colors indicate the belief distribution over the control actions.

Figure 11 shows the posterior predictive densities for the future digital states $Q^*(D_{t_c:t_p})$ and the corresponding control states $Q^*(U_{t_c:t_p})$, starting at $t_c = 60$ and spanning four time steps. These predictions capture the expected progression of structural health, conditioned on the posterior over policies $Q^*(\pi)$, thereby supporting the planning of preventive interventions. When belief propagation leads to an overly flat digital state distribution, the likelihood of selecting an RE action increases, mitigating the risk of decisions based on unreliable or uncertain belief states.

By running a second cluster of 100 simulations, each initialized with a different random seed, the ADT operating under combined goal-directed (pragmatic) and information-seeking (epistemic) behaviors consistently succeeds, with zero failures observed. This result underscores the potential of fully equipped ADTs compared to the purely pragmatic baseline discussed in Sec. 4.5. In this configuration, the ADT is able to autonomously cope with highly corrupted observations by actively exploring its environment in response to potentially critical uncertainty.

The results of a complete ADT simulation, combining goal-directed (pragmatic) and information-seeking (epistemic) behaviors and additionally incorporating learning updates to the generative model, are shown in Fig. 12 for the same initialization seed as in Fig. 10. This scenario spans 80 time steps and introduces learning via updates to the transition model array $\mathbf{B}$, with a learning rate of $\eta^{\mathbf{B}} = 0.1$. Learning demonstrates beneficial in several aspects. First, it reduces the frequency of incorrect digital state inferences, thereby shortening the average response delay relative to the ground-truth agent. Second, as the transition dynamics become progressively tailored to the (unknown) generative process, the ADT gains confidence in its predictions, resulting in a reduced need for (corrective) RE actions. Third, the gradual reduction of uncertainty in the transition model enables the ADT to safely delay maintenance toward the end of the simulation. For example, the third maintenance action, previously triggered at $t = 52$, is now postponed to $t = 74$. Moreover, this intervention is no longer based on a maximum a-posteriori estimate of D^δ within the [55%, 65%] range, but instead within the higher [65%, 75%] range – highlighting the potential for resource savings across the system operational lifespan. The runtime for this simulation is about 130 s, averaging 1.6 s per time slice.
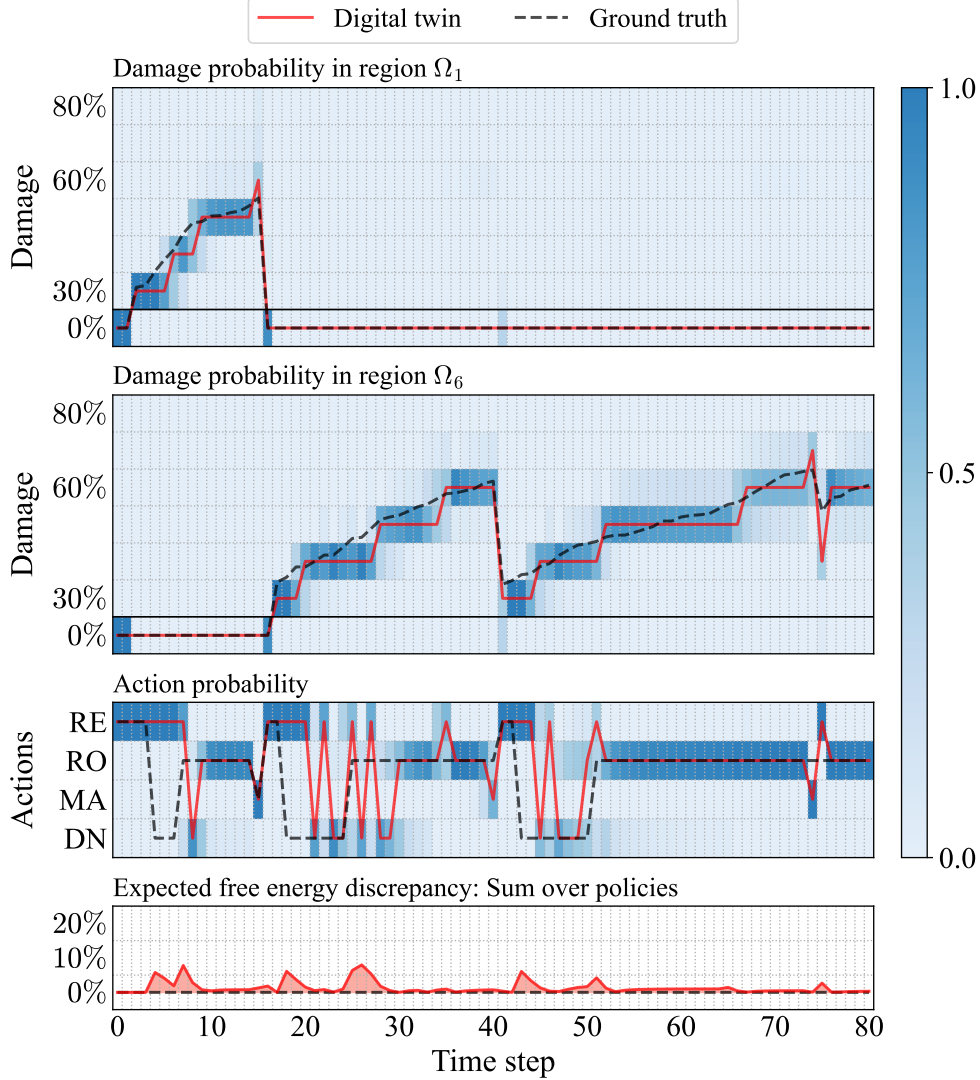
28

Figure 12: Active digital twin using a combination of goal-directed (pragmatic) and information-seeking (epistemic) behaviors and additionally incorporating learning updates to the generative model. Probabilistic and best-point estimates of: (top two panels) digital state evolution compared to the ground-truth physical state; (penultimate panel) control actions recommended by the digital twin versus the optimal action under the ground-truth generative process. In the top panels, background colors represent the belief distribution over the digital state at each time step. In the penultimate panel, background colors indicate the belief distribution over the control actions. The bottom panel quantifies simulation quality in terms of the percentage absolute discrepancy between the sum of the policy-specific expected free energies computed by the digital twin and those obtained under the ground truth.

5. Conclusions

This paper introduces active digital twins based on the active inference paradigm. By unifying probabilistic modeling and inference with perception, learning, and decision-making under uncertainty, the proposed approach enables digital twins to autonomously balance both goal-directed (pragmatic or utility-maximization) and information-seeking (epistemic or uncertainty-resolving) behaviors. This is because both factors are included in the variational free energy minimization process that drives active inference agents [4, 6]. Active inference endows active digital twins with the capacity to adaptively monitor, interact with, and learn from uncertain and dynamic environments. We have presented a case study in the context of structural health monitoring for a railway bridge, which demonstrates that active digital twins can achieve a new level of autonomy and resilience, surpassing traditional (more passive) approaches.

The active digital twin paradigm has been realized by leveraging active inference agents [4] to navigate the partially observable Markov decision process underlying the abstraction of physical-digital systems proposed by Kapteyn et al. [26]. At the core of this framework lies a self-updating generative model that enables intelligent automation through data assimilation, state estimation, prediction, planning, and learning – all with quantified uncertainty. Digital state estimation is performed via inference over dynamically evolving latent states, using variational free energy minimization. Action selection is framed as an inference process, wherein the agent infers a distribution over control policies that minimizes expected free energy, effectively closing the loop between perception and action. Finally, learning emerges from the (slow-timescale) inference of hyperparameters that define the generative model itself.

The considered case study is among the few active inference applications in engineering, focusing on the step-by-step construction of a physics-based generative model that enables bidirectional perception-action interaction. Simulations of both purely goal-directed (pragmatic) and combined goal-directed and information-seeking (pragmatic-epistemic) behaviors have demonstrated that the proposed framework supports a wide range of adaptive responses, where actions emerge from internal beliefs and their entropy. By incorporating all components of the expected free energy, the active twin effectively balances pragmatic goals and epistemic drives. Active exploration has proven essential for maintaining synchronization between the digital and physical states, particularly when the generative model must uncover key system features or when belief states become outdated. Moreover, including learning updates has shown that personalized generative models can improve inference accuracy, reduce the need for corrective epistemic actions, and enable the safe postponement of costly interventions. Crucially, the behavior of the twin is not fixed but remains customizable. There is no universally optimal behavior; rather, the performance and role of an active digital twin should be assessed in light of the objectives encoded in its generative model. For example, prior preferences on the structural health can be adjusted to reflect specific operational or safety requirements, and even treated as sweep parameters to derive meta-decision curves providing insight into how risk sensitivity can be modulated by the decision-maker.

Beyond the structural health monitoring application presented here, the proposed framework offers a generalizable methodology applicable across a wide range of domains. Active digital twins are envisioned as key enablers of autonomous agents in the development of smart structures and systems, as well as in fields such as medicine and neuroscience [81]. Future extensions will target self-healing structures and multi-agent physical systems capable of self-organization and continual self-learning. In addition, future work will investigate the scalability of the approach to higher-dimensional continuous state spaces, the multi-agent coordination of digital twin networks, and online learning from real-world data. This online learning will be complemented by offline updates via Bayesian model reduction [82], allowing the generative model to be periodically simplified by pruning uninformative concentration parameters. This process will enable an optimal trade-off between model complexity (quantifying the magnitude of belief updates required to maintain predictive performance) and model accuracy (reflecting the model fit to observed data). Competitive generative models will be compared to select the most parsimonious explanation of recent experience, retaining only those latent causes that meaningfully contribute to accurate inference.

References

[1] National Academy of Engineering and National Academies of Sciences, Engineering, and Medicine, Foundational Research Gaps and Future Directions for Digital Twins, The National Academies Press, Washington, DC, 2024. `doi:10.17226/26894`.

[2] W. Kritzinger, M. Karner, G. Traar, J. Henjes, W. Sihn, Digital Twin in manufacturing: A categorical literature review and classification, IFAC-PapersOnLine 51 (11) (2018) 1016–1022, 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018. `doi:10.1016/j.ifacol.2018.08.474`.

[3] A. Ferrari, K. Willcox, Digital twins in mechanical and aerospace engineering, Nature Computational Science 4 (2024) 178–183. `doi:10.1038/s43588-024-00613-8`.

[4] T. Parr, G. Pezzulo, K. J. Friston, Active Inference: The Free Energy Principle in Mind, Brain, and Behavior, MIT Press, Cambridge, MA, 2022. `doi:10.7551/mitpress/12441.001.0001`.

[5] S. J. Russell, Artificial Intelligence: A Modern Approach, Pearson Education, London, United Kingdom, 2020.

[6] K. Friston, The free-energy principle: a unified brain theory?, Nature Reviews Neuroscience 11 (2) (2010) 127–138. `doi:10.1038/nrn2787`.

[7] K. J. Friston, T. Parr, B. de Vries, The graphical brain: Belief propagation and active inference, Network Neuroscience 1 (4) (2017) 381–414. `doi:10.1162/NETN\_a\_00018`.

[8] Shafto, Mike and Conroy, Mike and Doyle, Rich and Glaessgen, Ed and Kemp, Chris and LeMoigne, Jacqueline and Wang, Lui, Technology Area 11: Modeling, Simulation, Information Technology & Processing, Tech. rep., National Aeronautics and Space Administration, Washington, District of Columbia (2020). `doi:10.17226/13354`.

[9] Digital Twin: Definition & Value — An AIAA and AIA Position Paper, AIAA Digital Engineering Integration Committee and others, Tech. rep., AIAA: Reston, Virginia (2020).

[10] M. Torzoni, M. Tezzele, S. Mariani, A. Manzoni, K. E. Willcox, A digital twin framework for civil engineering structures, Computer Methods in Applied Mechanics and Engineering 418 (2024) 116584. `doi:10.1016/j.cma.2023.116584`.

[11] E. Varetti, M. Torzoni, M. Tezzele, A. Manzoni, Adaptive digital twins for predictive decision-making through hierarchical Bayesian learning of transition dynamics, na (2025). `doi:na`.

[12] C. Li, S. Mahadevan, Y. Ling, S. Choze, L. Wang, Dynamic Bayesian Network for Aircraft Wing Health Monitoring Digital Twin, AIAA Journal 55 (3) (2017) 930–941. `doi:10.2514/1.J055201`.

[13] A. Phua, C. Davies, G. Delaney, A digital twin hierarchy for metal additive manufacturing, Computers in Industry 140 (2022) 103667. `doi:10.1016/j.compind.2022.103667`.

[14] M. Jans-Singh, K. Leeming, R. Choudhary, M. Girolami, Digital twin of an urban-integrated hydroponic farm, Data-Centric Engineering 1 (2020) 20. `doi:10.1017/dce.2020.21`.

[15] A. L. Reis, A. Andrade-Campos, P. Matos, C. Henggeler Antunes, M. A. Lopes, An energy and cost efficiency Model Predictive Control framework to optimize Water Supply Systems operation, Applied Energy 384 (2025) 125478. `doi:10.1016/j.apenergy.2025.125478`.

[16] A. Tzachor, S. Sabri, C. E. Richards, A. Rajabifard, M. Acuto, Potential and limitations of digital twins to achieve the sustainable development goals, Nature Sustainability 5 (10) (2022) 822–829. `doi: 10.1038/s41893-022-00923-7`.

[17] D. Cotoarbă, D. Straub, I. F. Smith, Probabilistic digital twins for geotechnical design and construction, arXiv preprint arXiv:2412.09432v1 (2024). `doi:10.48550/arXiv.2412.09432`.

[18] S. Henneking, S. Venkat, O. Ghattas, Goal-Oriented Real-Time Bayesian Inference for Linear Autonomous Dynamical Systems With Application to Digital Twins for Tsunami Early Warning, arXiv preprint arXiv:2501.14911v1 (2025). `doi:10.48550/arXiv.2501.14911`.

[19] G. Arcieri, C. Hoelzl, O. Schwery, D. Straub, K. Papakonstantinou, E. Chatzi, POMDP inference and robust solution via deep reinforcement learning: An application to railway optimal maintenance, Machine Learning (2024). `doi:10.1007/s10994-024-06559-2`.

[20] A. McClellan, J. Lorenzetti, M. Pavone, C. Farhat, A Physics-Based Digital Twin for Model Predictive Control of Autonomous Unmanned Aerial Vehicle Landing, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 380 (2229) (2022) 20210204. `doi:10.1098/rsta.2021.020417`.

[21] M. Tezzele, S. Carr, U. Topcu, K. E. Willcox, Adaptive planning for risk-aware predictive digital twins, arXiv preprint arXiv:2407.20490v2 (2024). `doi:10.48550/arXiv.2407.20490`.

[22] S. Henao-Garcia, M. Kapteyn, K. E. Willcox, M. Tezzele, M. Castroviejo-Fernandez, T. Kim, M. Ambrosino, I. Kolmanovsky, H. Basu, P. Jirwankar, R. Sanfelice, Digital-Twin-Enabled Multi-Spacecraft On-Orbit Operations, in: AIAA SCITECH 2025 Forum, 2025, p. 1432. `doi:10.2514/6.2025-1432`.

[23] J. Corral-Acero, F. Margara, M. Marciniak, C. Rodero, F. Loncaric, Y. Feng, A. Gilbert, J. F. Fernandes, H. A. Bukhari, A. Wajdan, et al., The 'Digital Twin' to enable the vision of precision cardiology, European Heart Journal 41 (48) (2020) 4556–4564. `doi:10.1093/eurheartj/ehaa159`.

[24] A. Chaudhuri, G. Pash, D. A. Hormuth, G. Lorenzo, M. Kapteyn, C. Wu, E. A. Lima, T. E. Yankeelov, K. Willcox, et al., Predictive digital twin for optimizing patient-specific radiotherapy regimens under uncertainty in high-grade gliomas, Frontiers in Artificial Intelligence 6 (2023). `doi:10.3389/frai.2023.1222612`.

[25] P. Bauer, B. Stevens, W. Hazeleger, A digital twin of Earth for the green transition, Nature Climate Change 11 (2) (2021) 80–83. `doi:10.1038/s41558-021-00986-y`.

[26] M. G. Kapteyn, J. V. Pretorius, K. E. Willcox, A probabilistic graphical model foundation for enabling predictive digital twins at scale, Nature Computational Science 1 (5) (2021) 337–347. `doi:10.1038/s43588-021-00069-0`.

[27] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, Cambridge, Massachusetts, 2009.

[28] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, G. Pezzulo, Active inference and epistemic value, Cognitive Neuroscience 6 (4) (2015) 187–214. `doi:10.1080/17588928.2015.1020053`.

[29] T. H. FitzGerald, R. J. Dolan, K. Friston, Dopamine, reward learning, and active inference, Frontiers in Computational Neuroscience 9 (2015) 136. `doi:10.3389/fncom.2015.00136`.

[30] T. Parr, K. J. Friston, Uncertainty, epistemics and active inference, Journal of the Royal Society Interface 14 (136) (2027) 20170376. `doi:10.1098/rsif.2017.0376`.

[31] G. Pezzulo, F. Rigoli, K. J. Friston, Hierarchical Active Inference: A Theory of Motivated Control, Trends in Cognitive Sciences 22 (4) (2018) 294–306. `doi:10.1016/j.tics.2018.01.009`.

[32] T. Van de Maele, B. Dhoedt, T. Verbelen, G. Pezzulo, A hierarchical active inference model of spatial alternation tasks and the hippocampal-prefrontal circuit, Nature Communications 15 (1) (2024) 9892. `doi:10.1038/s41467-024-54257-3`.

[33] Z. Fountas, N. Sajid, P. Mediano, K. Friston, Deep active inference agents using Monte-Carlo methods, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, New York, 2020, pp. 11662–11675.

[34] P. Mazzaglia, T. Verbelen, B. Dhoedt, Contrastive active inference, in: Proceedings of the 35th International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, New York, 2021, pp. 13870–13882.

[35] D. Maisto, F. Donnarumma, G. Pezzulo, Interactive inference: A multi-agent model of cooperative joint actions, IEEE Transactions on Systems, Man, and Cybernetics: Systems 54 (2) (2024) 704–715. `doi:10.1109/TSMC.2023.3312585`.

[36] C. Heins, B. Millidge, L. Da Costa, R. P. Mann, K. J. Friston, I. D. Couzin, Collective behavior from surprise minimization, Proceedings of the National Academy of Sciences 121 (17) (2024) e2320239121. `doi:10.1073/pnas.2320239121`.

[37] C. L. Buckley, C. S. Kim, S. McGregor, A. K. Seth, The free energy principle for action and perception: A mathematical review, Journal of mathematical psychology 81 (2017) 55–79. `doi:10.1016/j.jmp.2017.09.004`.

[38] P. Lanillos, C. Meo, C. Pezzato, A. A. Meera, M. Baioumy, W. Ohata, A. Tschantz, B. Millidge, M. Wisse, C. L. Buckley, J. Tani, Active inference in robotics and artificial agents: Survey and challenges, arXiv preprint arXiv:2112.01871 (2021). `doi:10.48550/arXiv.2112.01871`.

[39] T. Taniguchi, S. Murata, M. Suzuki, D. Ognibene, P. Lanillos, E. Ugur, L. Jamone, T. Nakamura, A. Ciria, B. Lara, et al., World models and predictive coding for cognitive and developmental robotics: frontiers and challenges, Advanced Robotics 37 (13) (2023) 780–806. `doi:10.1080/01691864.2023.2225232`.

[40] P. Vijayaraghavan, J. F. Queißer, S. V. Flores, J. Tani, Development of compositionality through interactive learning of language and action of robots, Science Robotics 10 (98) (2025) eadp0751. `doi:10.1126/scirobotics.adp0751`.

[41] A. Bounceur, M. Kara, Intelligent Acting Digital Twins (IADT), IEEE Access 13 (2025) 15201–15214. `doi:10.1109/ACCESS.2025.3532545`.

[42] R. Bajcsy, Active perception, Proceedings of the IEEE 76 (8) (1988) 966–1005. `doi:10.1109/5.5968`.

[43] J. Aloimonos, I. Weiss, A. Bandyopadhyay, Active vision, International Journal of Computer Vision 1 (4) (1988) 333–356. `doi:10.1007/BF00133571`.

[44] S. Thrun, W. Burgard, D. Fox, Probabilistic Robotics (Intelligent Robotics and Autonomous Agents), MIT Press, Cambridge, MA, 2005. `doi:10.7551/mitpress/12441.001.0001`.

[45] L. P. Kaelbling, M. L. Littman, A. R. Cassandra, Planning and acting in partially observable stochastic domains, Artificial Intelligence 101 (1-2) (1998) 99–134. `doi:10.1016/S0004-3702(98)00023-X`.

[46] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J. J. Leonard, Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age, IEEE Transactions on Robotics 32 (6) (2016) 1309–1332. `doi:10.1109/TRO.2016.2624754`.

[47] S. Grigorescu, B. Trasnea, T. Cocias, G. Macesanu, A survey of deep learning techniques for autonomous driving, Journal of Field Robotics 37 (3) (2020) 362–386. `doi:10.1002/rob.21918`.

[48] H. Niu, J. Liu, Z. Yu, D. Zheng, P. He, F. Wang, Real-time object tracking system using PTZ camera, in: 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology, 2022, pp. 471–478. `doi:10.1109/ICCASIT55263.2022.9986912`.

[49] S. D. Glaser, A. Tolman, Sense of Sensing: From Data to Informed Decisions for the Built Environment, Journal of Infrastructure Systems 14 (1) (2008) 4–14. `doi:10.1061/(ASCE)1076-0342(2008)14:1(4)`.

[50] J. D. Achenbach, Structural health monitoring – What is the prescription?, Mechanics Research Communications 36 (2) (2009) 137–142. `doi:10.1016/j.mechrescom.2008.08.011`.

[51] M. Torzoni, A. Manzoni, S. Mariani, Structural health monitoring of civil structures: A diagnostic framework powered by deep metric learning, Computers & Structures 271 (2022) 106858. `doi:10.1016/j.compstruc.2022.106858`.

[52] E. García-Macías, F. Ubertini, Integrated SHM Systems: Damage Detection Through Unsupervised Learning and Data Fusion, in: A. Cury, D. Ribeiro, F. Ubertini, M. D. Todd (Eds.), Structural Health Monitoring Based on Data Science Techniques, Springer International Publishing, Cham, Switzerland, 2022, pp. 247–268. `doi:10.1007/978-3-030-81716-9\_12`.

[53] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B. D. Youn, M. D. Todd, S. Mahadevan, C. Hu, Z. Hu, A comprehensive review of digital twin – part 1: modeling and twinning enabling technologies, Structural and Multidisciplinary Optimization 65 (2022) 354. `doi:10.1007/s00158-022-03425-4`.

[54] M. Torzoni, A. Manzoni, S. Mariani, Enhancing Bayesian model updating in structural health monitoring via learnable mapping, arXiv preprint arXiv:2405.13648v1 (2024). `doi:10.48550/arXiv.2405.13648`.

[55] Matteo Torzoni, Model-based and data-driven methodologies toward predictive digital twins of structures, Ph.D. thesis, Politecnico di Milano (2024).

[56] B. Zakić, A. Ryzynski, C. Guo-Hong, J. Jokela, Classification of damage in concrete bridges, Materials and Structures 24 (1991) 268–275. `doi:10.1007/BF02472082`.

[57] K. Friston, S. Samothrakis, R. Montague, Active inference and agency: optimal control without cost functions, Biological Cybernetics 106 (8) (2012) 523–541. `doi:10.1007/s00422-012-0512-8`.

[58] A. Kamariotis, K. Vlachas, V. Ntertimanis, I. Koune, A. Cicirello, E. Chatzi, On the Consistent Classification and Treatment of Uncertainties in Structural Health Monitoring Applications, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering 11 (1) (2024) 011108. `doi:10.1115/1.4067140`.

[59] V. Narouie, H. Wessels, F. Cirak, U. Römer, Mechanical state estimation with a Polynomial-Chaos-Based Statistical Finite Element Method, Computer Methods in Applied Mechanics and Engineering 441 (2025) 117970. `doi:10.1016/j.cma.2025.117970`.

[60] K. P. Murphy, Probabilistic machine learning: Advanced topics, MIT press, Cambridge, MA, 2023.

[61] M. J. Wainwright, M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference, Foundations and Trends® in Machine Learning 1 (1–2) (2008) 1–305. `doi:10.1561/2200000001`.

[62] C. M. Bishop, Pattern Recognition and Machine Learning, Information Science and Statistics, Springer, New York, New York, 2006.

[63] C. Heins, B. Millidge, D. Demekas, B. Klein, K. Friston, I. D. Couzin, A. Tschantz, pymdp: A Python library for active inference in discrete state spaces, Journal of Open Source Software 7 (73) (2022) 4098. `doi:10.21105/joss.04098`.

[64] M. Ülker-Kaustell, Some aspects of the dynamic soil-structure interaction of a portal frame railway bridge, Ph.D. thesis, KTH Royal Institute of Technology (2009).

[65] R. Smith, K. J. Friston, C. J. Whyte, A step-by-step tutorial on active inference and its application to empirical data, Journal of Mathematical Psychology 107 (2022) 102632. `doi:10.1016/j.jmp.2021.102632`.

[66] F. Gregoretti, G. Pezzulo, D. Maisto, cpp-AIF: A multi-core C++ implementation of Active Inference for Partially Observable Markov Decision Processes, Neurocomputing 568 (2024) 127065. `doi:10.1016/j.neucom.2023.127065`.

[67] L. Rosafalco, M. Torzoni, A. Manzoni, S. Mariani, A. Corigliano, Online structural health monitoring by model order reduction and deep learning algorithms, Computers & Structures 255 (2021) 106604. `doi:10.1016/j.compstruc.2021.106604`.

[68] T. J. Hughes, The Finite Element Method: Linear Static and Dynamic Finite Element Analysis, Dover Civil and Mechanical Engineering, Dover Publications, 2000.

[69] S. Eftekhar Azam, S. Mariani, Online damage detection in structural systems via dynamic inverse analysis: A recursive Bayesian approach, Engineering Structures 159 (2018) 28–45. `doi:10.1016/j.engstruct.2017.12.031`.

[70] A. Quarteroni, A. Manzoni, F. Negri, Reduced Basis Methods for Partial Differential Equations: An Introduction, Vol. 92, Springer, Cham, Switzerland, 2015. `doi:10.1007/978-3-319-15431-2`.

[71] F. Chinesta, A. Huerta, G. Rozza, K. Willcox, Model Reduction Methods, in: E. Stein, R. de Borst, T. J. R. Hughes (Eds.), Encyclopedia of Computational Mechanics, Second Edition, John Wiley & Sons, 2017, pp. 1–36.

[72] L. Sirovich, Turbulence and the dynamics of coherent structures. I. Coherent structures, Quarterly of Applied Mathematics 45 (3) (1987) 561–571. `doi:10.1090/qam/910462`.

[73] F. Negri, redbKIT, version 2.2, `http://redbkit.github.io/redbKIT` (2016).

[74] M. Torzoni, A. Manzoni, S. Mariani, A multi-fidelity surrogate model for structural health monitoring exploiting model order reduction and artificial neural networks, Mechanical Systems and Signal Processing 197 (2023) 110376. `doi:10.1016/j.ymssp.2023.110376`.

[75] M. Torzoni, L. Rosafalco, A. Manzoni, S. Mariani, A. Corigliano, SHM under varying environmental conditions: An approach based on model order reduction and deep learning, Computers & Structures 266 (2022) 106790. `doi:10.1016/j.compstruc.2022.106790`.

[76] M. Torzoni, A. Manzoni, S. Mariani, A Deep Neural Network, Multi-fidelity Surrogate Model Approach for Bayesian Model Updating in SHM, in: P. Rizzo, A. Milazzo (Eds.), European Workshop on Structural Health Monitoring, Springer International Publishing, Cham, Switzerland, 2023, pp. 1076–1086. `doi:10.1007/978-3-031-07258-1\_108`.

[77] C. R. Farrar, S. W. Doebling, D. A. Nix, Vibration–Based Structural Damage Identification, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 359 (1778) (2001) 131–149. `doi:10.1098/rsta.2000.0717`.

[78] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, D. Inman, A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications, Mechanical Systems and Signal Processing 147 (2021) 107077. `doi:10.1016/j.ymssp.2020.107077`.

[79] O. Fink, Q. Wang, M. Svensen, P. Dersin, W.-J. Lee, M. Ducoffe, Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications, Engineering Applications of Artificial Intelligence 92 (2020) 103678. `doi:10.1016/j.engappai.2020.103678`.

[80] F. Chollet, et al., Keras, `https://keras.io` (2015).

[81] K. Amunts, M. Axer, S. Banerjee, L. Bitsch, J. G. Bjaalie, P. Brauner, A. Brovelli, N. Calarco, M. Carrere, S. Caspers, et al., The coming decade of digital brain research: A vision for neuroscience at the intersection of technology and computing, Imaging Neuroscience 2 (2024) 1–35. `doi:10.1162/imag\_a\_00137`.

[82] K. J. Friston, M. Lin, C. D. Frith, G. Pezzulo, J. A. Hobson, S. Ondobaka, Active Inference, Curiosity and Insight, Neural Computation 29 (10) (2017) 2633–2683. `doi:10.1162/neco\_a\_00999`.

[83] M. Torzoni, Adt-code, github, `https://github.com/MatteoTorzy/DT-Active` (2025).

Appendix A. Expected free energy derivations

In this Appendix, we provide the complete derivation of the expected free energy expressions (10) and (35), as adapted to the adaptive digital twin framework from [63]:

Expected free energy.

$$
\begin{aligned}
G_t^\pi &= \mathbb{E}_{Q(O_t, D_t | \pi)}[\ln Q(D_t \mid \pi) - \ln \widetilde{p}(O_t, D_t \mid \pi)] \\
&= \mathbb{E}_{Q(O_t, D_t | \pi)}[\ln Q(D_t \mid \pi) - \ln \widetilde{p}(O_t, D_t \mid \pi) + \underbrace{\ln Q(D_t \mid O_t, \pi) - \ln Q(D_t \mid O_t, \pi)}_{=0}] \\
&= \mathbb{E}_{Q(O_t, D_t | \pi)}[\ln Q(D_t \mid \pi) - \ln Q(D_t \mid O_t, \pi) - \ln \widetilde{p}(O_t) \\
&\qquad - \ln p(D_t \mid O_t, \pi) + \ln Q(D_t \mid O_t, \pi)] \\
&= -\mathbb{E}_{Q(O_t, D_t | \pi)}[\ln Q(D_t \mid O_t, \pi) - \ln Q(D_t \mid \pi)] - \mathbb{E}_{Q(O_t, D_t | \pi)}[\ln \widetilde{p}(O_t)] \\
&\qquad + \mathbb{E}_{Q(O_t, D_t | \pi)}[\ln Q(D_t \mid O_t, \pi) - \ln p(D_t \mid O_t, \pi)] \\
&= -\underbrace{\mathbb{E}_{Q(O_t | \pi)}[D_{\mathrm{KL}}[Q(D_t \mid O_t, \pi) \mid\mid Q(D_t \mid \pi)]]}_{\text{Epistemic value (information gain)}} - \underbrace{\mathbb{E}_{Q(O_t | \pi)}[\ln \widetilde{p}(O_t)]}_{\text{Pragmatic value (utility)}} \\
&\qquad + \underbrace{\mathbb{E}_{Q(O_t | \pi)}[D_{\mathrm{KL}}[Q(D_t \mid O_t, \pi) \mid\mid p(D_t \mid O_t, \pi)]]}_{\text{Expected variational approximation error } (\geq 0)}.
\end{aligned}
\tag{A.1}
$$

Expected free energy with model parameters.

$$
\begin{aligned}
G_t^\pi &= \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(D_t, \phi \mid \pi) - \ln \widetilde{p}(O_t, D_t, \phi \mid \pi)] \\
&= \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(D_t, \phi \mid \pi) - \ln \widetilde{p}(O_t, D_t, \phi \mid \pi) \\
&\qquad + \underbrace{\ln Q(D_t, \phi \mid O_t, \pi) - \ln Q(D_t, \phi \mid O_t, \pi)}_{=0}] \\
&= \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(D_t \mid \pi) + \ln Q(\phi \mid \pi) - \ln \widetilde{p}(O_t) - \ln p(D_t, \phi \mid O_t, \pi) \\
&\qquad + \ln Q(D_t, \phi \mid O_t, \pi) - \ln Q(D_t, \phi \mid O_t, \pi)] \\
&= \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(D_t \mid \pi) - \ln Q(D_t \mid O_t, \pi) + \ln Q(\phi \mid \pi) - \ln Q(\phi \mid O_t, \pi) \\
&\qquad - \ln \widetilde{p}(O_t) - \ln p(D_t, \phi \mid O_t, \pi) + \ln Q(D_t, \phi \mid O_t, \pi)] \\
&= -\mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(D_t \mid O_t, \pi) - \ln Q(D_t \mid \pi)] \\
&\qquad - \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(\phi \mid O_t, \pi) - \ln Q(\phi \mid \pi)] - \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln \widetilde{p}(O_t)] \\
&\qquad + \mathbb{E}_{Q(O_t, D_t, \phi | \pi)}[\ln Q(D_t, \phi \mid O_t, \pi) - \ln p(D_t, \phi \mid O_t, \pi)] \\
&= -\underbrace{\mathbb{E}_{Q(O_t | \pi)}[D_{\mathrm{KL}}[Q(D_t \mid O_t, \pi) \mid\mid Q(D_t \mid \pi)]]}_{\text{Epistemic value (digital state information gain)}} - \underbrace{\mathbb{E}_{Q(O_t | \pi)}[D_{\mathrm{KL}}[Q(\phi \mid O_t, \pi) \mid\mid Q(\phi \mid \pi)]]}_{\text{Epistemic value (model parameters information gain)}} \\
&\qquad - \underbrace{\mathbb{E}_{Q(O_t | \pi)}[\ln \widetilde{p}(O_t)]}_{\text{Pragmatic value (utility)}} + \underbrace{\mathbb{E}_{Q(O_t | \pi)}[D_{\mathrm{KL}}[Q(D_t, \phi \mid O_t, \pi) \mid\mid p(D_t, \phi \mid O_t, \pi)]]}_{\text{Expected variational approximation error } (\geq 0)}.
\end{aligned}
\tag{A.2}
$$

MOX Technical Reports, last issues

Dipartimento di Matematica

Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

40/2025 Tentori, C.A.; Gregorio, C.; ...; Ieva, F.; Della Porta, M.G.
Clinical and Genomic-Based Decision Support System to Define the Optimal Timing of Allogeneic Hematopoietic Stem-Cell Transplantation in Patients With Myelodysplastic Syndromes

37/2025 Spreafico, M.; Ieva, F.; Fiocco, M.
Causal effect of chemotherapy received dose intensity on survival outcome: a retrospective study in osteosarcoma

Gimenez Zapiola, A.; Boselli, A.; Menafoglio, A.; Vantini, S.
Hyper-spectral Unmixing algorithms for remote compositional surface mapping: a review of the state of the art

35/2025 Perotto, S.; Ferro, N.; Speroni, G.; Temellini, E.
Anisotropic recovery-based error estimators and mesh adaptation for real-life engineering innovation

34/2025 Bucelli, M.; Dede', L.
Coupling models of resistive valves to muscle mechanics in cardiac fluid-structure interaction simulations

32/2025 De Sanctis, M.F.; Di Battista, I.; Arnone, E.; Castiglione, C.; Palummo, A.; Bernardi, M.; Ieva, F.; Sangalli, L.M.
Exploring nitrogen dioxide spatial concentration via physics-informed multiple quantile regression

33/2025 Di Battista, I.; De Sanctis, M.F.; Arnone, E.; Castiglione, C.; Palummo, A.; Sangalli, L.M.
A semiparametric space-time quantile regression model

31/2025 Botteghi, N.; Fresca, S.; Guo, M.; Manzoni, A.
HypeRL: Parameter-Informed Reinforcement Learning for Parametric PDEs

30/2025 Rosafalco, L.; Conti, P.; Manzoni, A.; Mariani, S.; Frangi, A.
Online learning in bifurcating dynamic systems via SINDy and Kalman filtering

29/2025 Centofanti, E.; Ziarelli, G.; Parolini, N.; Scacchi, S.; Verani, M. ; Pavarino, L. F.
Learning cardiac activation and repolarization times with operator learning