Physics-Informed Neural Networks for Enterprise-Scale Cyber-Physical Systems

A Practical Framework for State Estimation and Control Under Partial Observability

Author: Ashwin Gupta
Affiliation: Coforge Limited
Date: December, 2025
Industries: Aviation & Transportation Systems, Energy & Infrastructure Systems

Table of Contents

Executive Summary

Modern enterprise systems in aviation, energy, and large-scale infrastructure operate under known physical laws, yet their internal states are only partially observable in practice. Critical variables such as temperature distributions, structural stress, degradation rates, and latent control states are rarely measured directly due to cost, safety, or physical constraints. Operational decisions are therefore made under uncertainty, often relying on simplified models or heuristics that fail outside nominal conditions.

Existing approaches struggle to close this gap. First-principles models degrade when boundary conditions and parameters are uncertain or time-varying. Purely data-driven machine learning models interpolate observed data but extrapolate poorly, violate physical constraints, and fail silently under distribution shift. Hybrid rule-based systems encode operational experience but do not scale across assets or operating regimes.

This paper addresses a narrow but recurring class of enterprise problems characterized by **continuous-time dynamics, partial observability, and high operational risk**. In these systems, the core challenge is **inverse inference**: estimating hidden states and parameters that are consistent with both sparse measurements and governing physical laws.

Physics-Informed Neural Networks (PINNs) are evaluated as a targeted solution to this problem. They are not presented as a general AI technique, nor as a replacement for classical solvers or control theory. Instead, PINNs are treated as a mechanism for enforcing physical consistency while learning from operational data in settings where traditional methods and purely data-driven models are insufficient.

The paper introduces a canonical PINNs solution pattern suitable for enterprise environments and demonstrates its application across two verticals—**Aviation & Transportation Systems** and **Energy & Infrastructure Systems**—through six focused workflows, including thermal state estimation, structural deformation inference, and physics-constrained control. Across these cases, PINNs enable earlier detection of abnormal conditions, reduced dependence on dense instrumentation, improved control stability, and more informed maintenance and operational decisions.

The paper is explicit about limitations. PINNs introduce training complexity, computational cost, and convergence risk, and are inappropriate for discrete, event-driven systems or problems lacking well-defined governing dynamics.

The central claim is deliberately narrow: **for partially observable, physics-governed enterprise systems, PINNs provide a viable and repeatable approach to inverse inference that is otherwise difficult to achieve in production environments**.

1. The Enterprise Problem: Operating Physical Systems with Incomplete State Information

Enterprise IT systems increasingly sit in direct control or supervision of physical processes. In aviation, software platforms monitor and regulate thermal behavior, structural loads, and environmental control systems. In energy and infrastructure, IT systems manage cooling, power delivery, and long-lived assets whose internal condition evolves continuously over time. These systems are governed by well-established physical laws, but are operated with limited direct visibility into their true internal state.

The constraint is not a lack of theory. It is a lack of observability.

In production environments, internal system states are rarely measured in full. Temperature sensors are placed sparsely due to cost and reliability concerns. Strain gauges and inspection points are limited by accessibility and certification requirements. Degradation processes unfold over long time horizons, while measurements remain indirect and intermittent. As a result, the quantities that matter most for safety, efficiency, and asset longevity are inferred indirectly, often using simplified assumptions.

This gap between **system behavior** and **system visibility** creates persistent operational risk. Overheating is detected after performance degrades. Structural fatigue is identified during scheduled inspections rather than inferred continuously. Control systems react to surface measurements without accounting for latent dynamics, leading to oscillation, inefficiency, or conservative operating margins.

From an IT and systems engineering perspective, this is not primarily a data volume problem. In many of these environments, telemetry is limited by design and cannot be made dense without prohibitive cost or disruption. Nor is it a purely computational problem; high-fidelity simulation exists but requires parameters and boundary conditions that are not known precisely during operation.

The core issue is that enterprise systems must make decisions based on **partial information about continuous physical processes**. The internal state evolves according to known dynamics, but only fragments of that state are observable at any given time. This mismatch is structural, not incidental.

Traditional enterprise architectures address this gap using a combination of static models, safety margins, and operational heuristics. These approaches are serviceable under nominal conditions but degrade as systems age, environments change, or operating envelopes expand. The result is either increased risk or increased conservatism, both of which carry measurable cost.

Any solution that claims to improve outcomes in these settings must therefore address a specific requirement:
the ability to infer hidden physical states and parameters from sparse operational data while remaining consistent with known system dynamics.

The remainder of this paper examines why existing approaches struggle to meet this requirement and under what conditions physics-informed learning methods provide a defensible alternative.

2. Why Existing Approaches Break Down

The difficulty of operating partially observable physical systems is not new. What is new is the expectation that software platforms should manage these systems continuously, at scale, and under changing conditions. Existing approaches fail not because they are poorly implemented, but because they are mismatched to the structure of the problem.

2.1 Limits of First-Principles Models in Production

Physics-based models encode governing equations explicitly and provide strong guarantees under controlled assumptions. In enterprise environments, those assumptions rarely hold for long.

Accurate simulation requires knowledge of boundary conditions, material properties, loads, and environmental factors that are either unmeasured or time-varying in real operations. As assets age, parameters drift. As operating regimes expand, nominal models lose fidelity. Recalibration is typically manual, episodic, and reactive, relying on inspections or offline analysis rather than continuous inference.

As a result, first-principles models are often reduced to advisory tools rather than operational decision engines. They are trusted in design and validation phases, but treated cautiously in live systems where uncertainty accumulates faster than models can be updated.

The limitation is not physics. It is observability.

2.2 Failure Modes of Purely Data-Driven Machine Learning

Data-driven models approach the problem from the opposite direction: they infer system behavior directly from observed data. When telemetry is dense and operating conditions are stable, this approach can be effective.

In the systems considered in this paper, neither condition holds. Sensors are sparse by necessity, not by oversight. Operating regimes shift due to load changes, environmental variation, and control interventions. Under these conditions, purely data-driven models learn correlations that are insufficiently constrained.

Two failure modes dominate in practice. First, extrapolation beyond observed regimes produces physically invalid states without explicit warning. Second, distribution shift degrades performance gradually, making failures difficult to detect until consequences are visible at the system level.

The absence of physical constraints means that correctness is evaluated statistically rather than structurally. In safety- and cost-critical systems, this is not an acceptable trade.

2.3 Why Hybrid Heuristics and Control Logic Do Not Scale

Most enterprise systems rely on a combination of simplified models, thresholds, and control heuristics derived from operational experience. These approaches are pragmatic and often effective locally.

Their weakness is scalability. As systems grow in complexity, heuristics accumulate rather than generalize. Edge cases multiply. Control logic becomes brittle as interactions between subsystems increase. The resulting systems are difficult to reason about, difficult to extend, and costly to maintain.

Crucially, these approaches do not improve observability. They react to surface measurements rather than inferring internal state. As a result, they either operate conservatively to avoid risk or accept risk implicitly without quantifying it.

2.4 The Common Failure Pattern

Across physics-based models, data-driven learning, and heuristic control, the same limitation appears repeatedly:
None of these approaches reliably infer hidden system states under partial observability while remaining consistent with known dynamics.

Physics-based models assume too much knowledge.
Data-driven models assume too much data.
Heuristics assume stability that does not persist.

This gap is the technical motivation for exploring physics-informed learning methods. Any alternative must address inverse inference directly, without requiring dense instrumentation or abandoning physical consistency.

The next section defines the conditions under which such methods are justified—and where they are not.

3. When Physics-Informed Neural Networks Are Justified — and When They Are Not

Physics-Informed Neural Networks are expensive to train, sensitive to formulation, and unforgiving of poor assumptions. They should not be treated as a general modeling technique. Their use is justified only when a specific set of conditions is met. Outside those conditions, they add complexity without commensurate value.

3.1 Conditions That Justify Their Use

PINNs are appropriate when **all** of the following hold.

Known governing dynamics exist.
The system is governed by differential equations that are well understood at the relevant scale. These equations may be approximate, but they must be meaningful enough to constrain behavior. PINNs do not compensate for unknown or poorly defined physics.

Internal states are not directly observable.
The quantities of interest—temperature fields, stress distributions, degradation parameters, latent control states—cannot be fully measured in operation. If dense, reliable sensing is already available, simpler methods dominate.

Data is sparse, noisy, or irregular by design.
Telemetry limitations arise from cost, safety, accessibility, or operational constraints, not from immature instrumentation. PINNs are not a substitute for missing data pipelines.

The primary task is inverse inference, not forward prediction.
The goal is to infer hidden states or parameters consistent with both data and physics. If the task is classification, ranking, or short-horizon forecasting, PINNs are usually unnecessary.

Physical consistency matters operationally.
Incorrect but plausible-looking outputs are unacceptable. Violations of conservation laws or stability constraints carry real cost or risk.

When these conditions are met, PINNs provide a mechanism to combine limited data with structural knowledge in a way that neither classical solvers nor unconstrained learning can achieve alone.

3.2 Conditions That Disqualify Their Use

PINNs should be rejected outright under the following circumstances.

The system is primarily discrete or event-driven.
Workflow orchestration, transaction processing, user behavior modeling, and most enterprise application logic do not benefit from physics-based constraints.

Governing dynamics are weak, unknown, or highly stochastic.
If the equations themselves are speculative or dominated by unmodeled effects, enforcing them adds bias rather than robustness.

High-quality labeled data is abundant.
When direct supervision is available at scale, simpler models are easier to train, validate, and maintain.

Real-time constraints dominate accuracy requirements.
PINNs are not lightweight. If inference latency must be minimal and approximate solutions suffice, classical control or reduced-order models are preferable.

Interpretability requirements exceed model fidelity needs.
In some regulatory or operational contexts, explicit analytical models are preferred even if they are less accurate.

3.3 What PINNs Are — and Are Not — in This Context

Within the scope of this paper, PINNs are treated as **constrained inference engines**, not as universal predictors or AI platforms. Their role is to recover internal system state and parameters that cannot be measured directly, while respecting known dynamics.

They do not replace simulation, control theory, or domain expertise. They sit between these components, filling a gap created by partial observability.

The remainder of this paper focuses exclusively on problems that meet these criteria. No claims are made beyond that boundary.

4. The Canonical PINNs Pattern for Enterprise Systems

When Physics-Informed Neural Networks are justified, their effectiveness depends less on architectural novelty and more on disciplined formulation. Across aviation and energy systems, successful deployments follow a common structure. This section defines that structure at a level suitable for enterprise implementation, not academic experimentation.

4.1 Problem Formulation

Each workflow addressed in this paper can be expressed as a continuous system governed by differential equations over space and time. The objective is not to solve these equations forward under known conditions, but to infer unobserved internal states or parameters that are consistent with both the equations and available measurements.

The formulation therefore consists of three elements:

- a set of governing dynamics representing conservation laws or constitutive behavior,
- a limited set of observational constraints derived from telemetry,
- and boundary or initial conditions that may be uncertain or time-varying.

The role of the PINN is to represent candidate system states while enforcing these constraints during training.

4.2 Model Structure and Constraint Enforcement

The neural network acts as a continuous function approximator over the system domain. Physical consistency is enforced by embedding the governing equations directly into the training objective. Rather than fitting outputs to labels alone, the model is penalized for violating known dynamics.

This constraint-based formulation distinguishes PINNs from surrogate models trained on simulation data. The network is not learning a mapping from inputs to outputs in isolation; it is learning a representation of system behavior constrained by physics.

In enterprise settings, this constraint enforcement serves two purposes:

- it regularizes learning under sparse data,
- and it prevents physically invalid solutions during inference.

4.3 Data Assimilation Under Partial Observability

Operational telemetry enters the model as sparse observational constraints rather than dense supervision. Measurements may be unevenly distributed in space and time and may carry nontrivial noise.

The training objective balances fidelity to observed data against adherence to governing dynamics. This balance is not static. In practice, it is adjusted to reflect confidence in sensors, uncertainty in boundary conditions, and the evolving operational context.

This approach allows the model to interpolate and extrapolate internal states while remaining anchored to known system behavior.

4.4 Training Stability and Convergence Considerations

PINN training is sensitive to scaling, initialization, and loss weighting. In enterprise environments, this sensitivity cannot be treated as a tuning inconvenience; it must be managed explicitly.

Effective implementations rely on:

- nondimensionalization or normalization of governing equations,
- staged or curriculum-based training strategies,
- and convergence diagnostics tied to physical consistency rather than loss reduction alone.

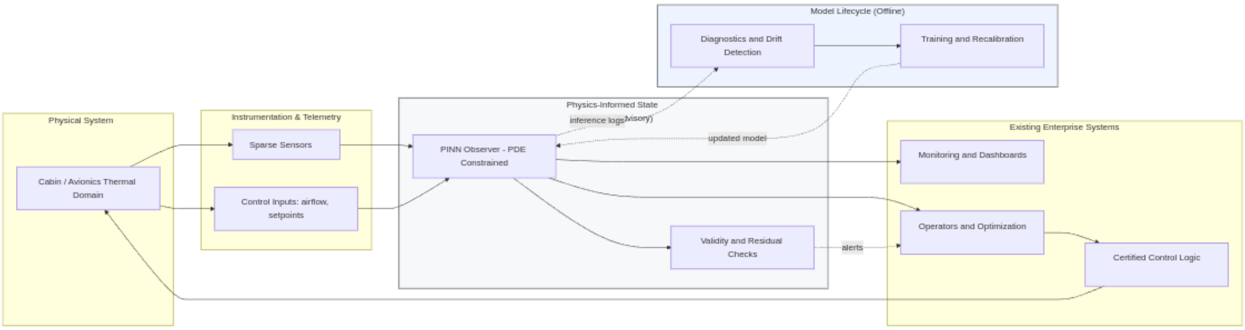
Training is treated as an engineering process, not an automated pipeline.

4.5 Deployment and Runtime Use

Once trained, the PINN operates as a state inference component within a larger system. It does not replace existing control logic or monitoring infrastructure. Instead, it augments them by providing estimates of internal states or parameters that are otherwise unavailable.

Runtime monitoring focuses on detecting violations of modeling assumptions rather than maximizing prediction accuracy. When the system moves outside the domain for which the model is valid, inference is flagged rather than silently extrapolated.

This pattern—physics-constrained inference, sparse data assimilation, and conservative deployment—is reused across all workflows described in the following sections.



5. Aviation & Transportation Systems

Aviation systems combine continuous physical dynamics with strict operational constraints. Thermal behavior, structural response, and environmental control evolve according to known laws, yet direct measurement of internal states is limited by weight, certification, accessibility, and reliability constraints. As a result, flight operations and maintenance decisions are routinely made with incomplete information about system condition.

The workflows in this section address this gap by applying physics-informed inference to recover internal states that are not directly observable during operation. Each workflow instantiates the same inference pattern described in Section 4, adapted to aviation-specific constraints and failure modes.

5.1 Aircraft Subsystem Thermal State Estimation:

Aircraft subsystems such as avionics bays, power distribution units, and localized cooling assemblies dissipate heat continuously under variable load and environmental conditions. Thermal margins are narrow, and localized overheating accelerates component degradation and increases failure risk. Despite this, full thermal instrumentation is infeasible in operational aircraft.

The engineering problem is therefore not predicting temperatures at known sensor locations, but **inferring the evolving internal temperature field of the subsystem** using sparse measurements and incomplete boundary information.

System Dynamics

At the subsystem level, thermal behavior can be represented by transient heat transfer dynamics:

$$\frac{\partial T(x,t)}{\partial t} - \nabla \cdot (k \nabla T(x,t)) = q(x,t)$$

where T(x,t) is the temperature field, k represents effective thermal conductivity, and q(x,t) captures internal heat generation. In operation, both q(x,t) and boundary heat fluxes vary with mission profile, airflow, and component loading, and are not directly measured.

Why Conventional Methods Are Insufficient

Physics-based thermal models require accurate boundary conditions and material parameters to remain reliable. In flight, these inputs drift continuously and are only partially observable, forcing models to be calibrated conservatively or used offline.

Data-driven models trained on historical temperature readings interpolate known sensor behavior but provide no structural constraints on unmeasured regions. Under new operating regimes or abnormal heat loads, they can produce internally inconsistent thermal states without clear indication of failure.

Neither approach provides reliable, full-field thermal inference during live operation.

PINNs-Based Inference

In this workflow, the temperature field T(x,t)T(x,t)T(x,t) is represented by a neural network constrained to satisfy the governing heat equation. Sparse temperature measurements act as observational constraints, while uncertain heat sources and boundary conditions are inferred implicitly during training.

The PINN is trained to produce thermal states that:

- satisfy known heat transfer dynamics,
- remain consistent with available telemetry,
- and adapt to changing operating conditions without manual recalibration.

This formulation allows the model to infer internal gradients and localized hotspots that are not instrumented directly.

Operational Use

Once trained, the model operates as a thermal state estimation component integrated into existing monitoring systems. It provides continuous estimates of internal temperatures and flags deviations from expected physical behavior rather than reacting only to sensor threshold violations.

Operationally, this enables earlier detection of abnormal thermal behavior, improved maintenance scheduling based on inferred thermal stress, and reduced reliance on conservative safety margins imposed due to uncertainty rather than evidence.

The contribution is not increased sensing resolution, but **improved observability through physics-constrained inference**.

5.2 Structural Deformation and Fatigue Inference

Aircraft structures experience continuous mechanical loading driven by aerodynamic forces, maneuvers, pressurization cycles, and environmental conditions. Over time, these loads induce deformation and fatigue that accumulate below visible thresholds. Structural failure is rarely sudden; it is the result of internal stress evolution that remains unobserved until inspections or limit violations occur.

In operational aircraft, **direct measurement of full stress or deformation fields is infeasible**. Strain gauges are sparse, placed selectively due to weight, certification, and maintenance constraints. Inspections are periodic and offline. The internal structural state between these checkpoints is inferred indirectly, often conservatively.

The problem is therefore not detecting known damage, but **inferring internal deformation and fatigue progression under partial observability**.

System Dynamics

At a simplified structural level, deformation can be expressed through beam or structural mechanics relationships:

$$EI\frac{d^4w(x)}{dx^4} = F(x)$$

where w(x) denotes structural deflection, E the effective modulus, I the moment of inertia, and F(x) the applied load distribution. In practice, load profiles, boundary conditions, and material properties vary with flight conditions and aging, and are not fully known during operation.

Why Conventional Methods Are Insufficient

Structural analysis models are reliable under known loads and boundary conditions, typically during design or controlled testing. In flight, loads are inferred indirectly and boundary conditions evolve due to wear, retrofits, and environmental exposure. As a result, these models are not used continuously for operational inference.

Data-driven approaches trained on limited strain measurements face a complementary limitation. They interpolate observed sensor responses but provide no guarantee that inferred internal stresses satisfy mechanical consistency. Under rare or evolving load regimes, they may underestimate fatigue accumulation without explicit indication of error.

Scheduled inspections mitigate risk but do not eliminate uncertainty between inspection intervals. Conservative safety margins compensate for this gap at the cost of reduced utilization.

PINNs-Based Inference

In this workflow, the structural deformation field is represented as a continuous function constrained by known mechanical relationships. Sparse strain or displacement measurements provide observational anchors, while load distributions and boundary effects are treated as latent variables inferred implicitly.

The PINN is trained to infer deformation states that:

- satisfy governing structural dynamics,
- align with sparse sensor data,
- and remain consistent across varying load conditions.

This enables continuous estimation of internal deformation and stress proxies without requiring dense instrumentation.

Operational Use

Deployed within structural health monitoring systems, the model provides ongoing estimates of deformation and inferred fatigue indicators. Rather than waiting for threshold exceedance or inspection findings, operators gain visibility into trends that indicate elevated risk.

Operational benefits include earlier identification of abnormal load paths, more targeted inspections, and improved maintenance planning based on inferred structural stress rather than conservative assumptions.

The value lies in **continuous structural inference between inspections**, not in replacing certification-grade analysis or inspection regimes.

5.3 Cabin and Avionics Climate Control Optimization

Aircraft cabin and avionics cooling systems operate in regimes where thermal behavior is dominated not only by conduction, but also by **localized heat exchange with a circulating medium**—airflow driven by environmental control systems or forced convection within enclosures. Near surfaces and occupied zones, simple boundary-condition models fail to capture this interaction accurately.

The control challenge is therefore not regulating a single measured temperature, but **stabilizing a distributed thermal field whose internal exchange mechanisms are only partially observable**.

System Dynamics

In this workflow, thermal evolution is modeled using a formulation analogous to Pennes' bioheat equation, which extends classical heat transfer by explicitly accounting for distributed exchange with a transport medium:

$$\rho c \frac{\partial T(x,t)}{\partial t} = \nabla \cdot (k \nabla T(x,t)) + q(x,t) + \omega(x)\, c_f\, (T_f(t) - T(x,t))$$

where T(x,t) represents the internal temperature field, q(x,t) denotes internal heat generation, and the exchange term models localized heat transfer between the solid domain and a circulating fluid at reference temperature Tf(t). The exchange coefficient ω(x) captures effective flow-mediated heat transport, which varies spatially and is not directly measured during operation.

This formulation is particularly relevant in cabin and avionics environments where airflow-driven heat exchange dominates near boundaries and occupied zones.

Why Conventional Control Approaches Are Insufficient

Traditional climate control systems rely on lumped models and surface temperature feedback. These approaches assume uniform mixing or fixed exchange coefficients, assumptions that break down under changing airflow patterns, load distributions, and occupancy conditions.

Pure feedback control reacts to measured temperature deviations without accounting for internal thermal inertia or spatial gradients. This leads to delayed response, oscillatory behavior, or conservative tuning that sacrifices efficiency to preserve stability.

The limitation is not actuator capability, but **inadequate state representation** of the thermal system being controlled.

PINNs-Based State Inference for Control

In this workflow, a Physics-Informed Neural Network is used to infer the internal thermal state governed by the bioheat formulation. The network represents the temperature field T(x,t)T(x,t)T(x,t) and is constrained to satisfy the governing equation while assimilating sparse temperature measurements and known control inputs.

The exchange coefficient ω(x)\omega(x)ω(x) and effective fluid temperature Tf(t)T_f(t)Tf(t) are treated as latent variables inferred implicitly, allowing the model to adapt to changing airflow and load conditions without explicit reparameterization.

The PINN acts as a **physics-constrained state observer**, recovering internal thermal dynamics that are not captured by sensor feedback alone.

Operational Use

The inferred thermal state is integrated into the existing control loop to provide physically consistent state feedback rather than replacing the controller itself. This improves anticipation of delayed thermal response, reduces oscillation, and enables tighter control without aggressive overcooling.

Operationally, this results in improved energy efficiency, smoother control behavior, and more stable thermal conditions for both occupants and avionics systems.

The contribution is not a new control strategy, but **a better-informed control system grounded in distributed thermal inference**.

6. Energy & Infrastructure Systems

Energy and infrastructure platforms operate physical assets at scale over long lifetimes. Thermal behavior in data centers, power electronics, and cooling infrastructure directly impacts reliability, efficiency, and capital utilization. As in aviation systems, internal thermal states evolve continuously while direct measurement remains sparse by design.

The workflows in this section apply the same physics-informed inference pattern to infrastructure-scale systems, demonstrating that the approach is not domain-specific, but **problem-structure–specific**.

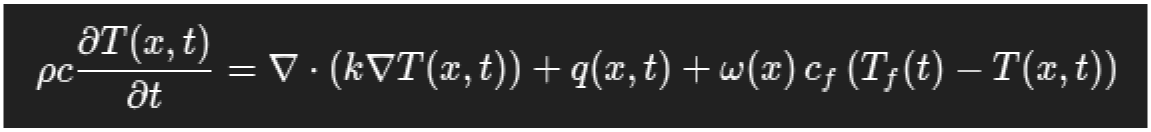
6.1 Data Center Thermal State Estimation

Modern data centers host densely packed compute hardware with highly non-uniform heat generation. Localized hotspots emerge due to workload variation, airflow imbalance, and cooling inefficiencies. While surface and inlet temperatures are monitored, **the internal thermal state across racks, boards, and components is largely unobserved**.

The operational objective is not retrospective hotspot detection, but **continuous inference of internal thermal conditions** to support proactive control and capacity planning.

System Dynamics

Thermal behavior within a data center zone can be represented as a transient heat transfer system with distributed exchange between solid components and circulating coolant or air:



$$\rho c \frac{\partial T(x,t)}{\partial t} = \nabla \cdot \left(k \nabla T(x,t)\right) + q(x,t) + \omega(x)\, c_f \left(T_f(t) - T(x,t)\right)$$

Here, T(x,t) denotes the internal temperature field, q(x,t) captures time-varying compute-driven heat generation, and the exchange term represents heat transfer with airflow or liquid cooling systems. The effective exchange coefficient ω(x) depends on local flow conditions and is not directly observable in operation.

This formulation captures the dominant thermal behavior near boards, heat sinks, and enclosure boundaries where lumped models are insufficient.

Why Conventional Methods Are Insufficient

Computational fluid dynamics and thermal simulation tools provide high-fidelity insight during design but are too computationally expensive and sensitive to boundary assumptions for continuous operational use. Simplified thermal models used in production rely on fixed parameters that fail under dynamic workloads and changing airflow patterns.

Purely data-driven models trained on sensor data interpolate known measurements but do not enforce thermal consistency across uninstrumented regions. Under load shifts or cooling anomalies, these models can misrepresent internal conditions without clear indication of error.

Neither approach provides reliable, full-field thermal inference in real time.

PINNs-Based Inference

In this workflow, a Physics-Informed Neural Network represents the internal temperature field and enforces the governing heat transfer dynamics. Sparse temperature measurements from sensors act as observational constraints, while internal heat generation and exchange coefficients are inferred implicitly.

The model is trained to recover thermal states that:

- satisfy known physical relationships,
- adapt to changing workloads and cooling conditions,
- and remain stable under sparse instrumentation.

This allows inference of hotspots and gradients that are not directly sensed.

Operational Use

Deployed as part of the data center monitoring stack, the inferred thermal state supports earlier detection of cooling inefficiencies, improved workload placement, and tighter integration with cooling control systems. Rather than reacting to threshold violations, operators gain visibility into emerging thermal stress.

The value is not higher-resolution monitoring, but **improved thermal observability using existing telemetry**.

6.2 Direct-to-Chip Cooling Control Systems

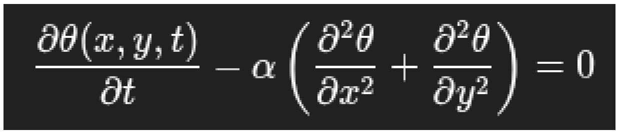
Direct-to-chip cooling systems are designed to remove heat at the point of generation, where thermal gradients are steep and time constants are short. Unlike room- or rack-level cooling, these systems operate in regimes where small delays or modeling errors translate directly into throttling, performance loss, or hardware damage. Control decisions must therefore be made with limited tolerance for uncertainty.

Despite this, **the internal thermal state of the chip and substrate remains largely unobserved**. Temperature sensors are sparse and positioned for protection rather than full-field observability. Coolant behavior near the chip surface—where most heat exchange occurs—is not measured directly. As a result, control systems act on delayed, surface-level signals while the most critical thermal dynamics unfold internally.

The core challenge is not cooling capacity, but **state-aware control under fast, partially observable thermal dynamics**.

System Dynamics

At the chip and substrate level, thermal evolution is dominated by transient conduction with strong boundary effects. A representative governing equation is the two-dimensional transient heat equation:



$$\frac{\partial\theta(x,y,t)}{\partial t} - \alpha\left(\frac{\partial^2\theta}{\partial x^2} + \frac{\partial^2\theta}{\partial y^2}\right) = 0$$

where θ represents the internal temperature field and $\alpha$the effective thermal diffusivity. In operational systems, this equation is influenced by boundary interactions with coolant flow and heat spreaders that vary dynamically with load and control input.

Critically, these boundary effects dominate thermal response near the chip surface but are not directly observable during runtime.

Why Conventional Control Breaks Down

Classical control approaches rely on lumped thermal models and surface temperature feedback. These models assume uniform internal behavior and fixed response delays, assumptions that fail under high power density and rapid workload variation.

When internal gradients are steep, surface measurements lag true internal temperature peaks. Controllers react late, compensate aggressively, and induce oscillations or conservative operating limits. Increasing controller gain improves responsiveness but destabilizes the system when boundary conditions shift.

Offline thermal simulations capture these effects accurately but are computationally impractical for real-time control and require boundary conditions that are not known precisely in operation.

The failure mode is consistent: **control decisions are made without reliable knowledge of the internal thermal state**.

PINNs-Based State Inference for Control Support

In this workflow, a Physics-Informed Neural Network is used to infer the internal temperature field governed by the transient heat equation. The network represents θ(x,y,t)\theta(x,y,t)θ(x,y,t) as a continuous function constrained by the governing dynamics, while assimilating sparse temperature measurements and known control inputs.

Rather than attempting to model coolant behavior explicitly, the PINN infers the effective impact of boundary heat exchange implicitly through its constraint structure. This allows the model to recover internal thermal gradients and transient peaks that are not visible to surface sensors.

The PINN functions as a **physics-constrained state observer**, producing physically consistent internal state estimates in regimes where direct measurement is unavailable and purely data-driven observers are unstable.

Operational Use

The inferred thermal state is integrated into the existing cooling control loop as an augmentation, not a replacement. Control logic continues to operate within certified and validated boundaries, but with access to state estimates that anticipate internal thermal response rather than reacting to delayed measurements.

This enables:

- earlier intervention under rapid load changes,
- reduced oscillatory behavior in cooling actuation,

- tighter thermal margins without increased risk,
- and improved utilization of cooling capacity.

The outcome is improved performance stability and hardware protection under high-density compute workloads.

The contribution of this approach is not a new cooling mechanism or control law, but **state-aware control enabled by physics-constrained inference** in a regime where traditional observability assumptions no longer hold.

6.3 Thermal Degradation Modeling of Grid and Power Electronics Assets

Power electronics and grid-connected assets operate under sustained thermal stress over long service lifetimes. Transformers, inverters, converters, and switchgear experience temperature-driven degradation mechanisms that accumulate gradually, often remaining undetected until performance degrades or failures occur. Unlike fast transient cooling problems, the dominant challenge here is **long-term internal condition estimation under sparse monitoring**.

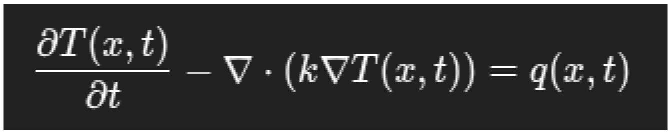
Direct measurement of internal temperatures and material condition is limited by accessibility, safety, and cost. Sensors are typically placed at external or surrogate locations, while degradation processes evolve internally over months or years. As a result, asset health is inferred indirectly, often using conservative lifetime models or periodic inspections.

The operational objective is not predicting instantaneous temperature, but **estimating hidden thermal stress and degradation state to inform maintenance and replacement decisions**.

System Dynamics

At the asset level, thermal behavior can be represented as a diffusion-driven process with internal heat generation and boundary exchange. Over longer time scales, this evolution governs degradation rates of insulation, semiconductors, and structural materials.

A representative formulation is:



$$\frac{\partial T(x,t)}{\partial t} - \nabla \cdot (k \nabla T(x,t)) = q(x,t)$$

where T(x,t) represents internal temperature and q(x,t) captures load-dependent heat generation. While the governing physics is well understood, internal temperature fields are not directly observable during normal operation.

Degradation models depend on cumulative thermal exposure rather than instantaneous values, making accurate inference of internal temperature histories critical.

Why Conventional Asset Health Models Are Insufficient

Traditional asset management relies on simplified thermal indices, empirical aging curves, or static design assumptions. These approaches approximate degradation using external measurements or nameplate ratings, often failing to reflect actual operating conditions.

Physics-based simulations provide detailed insight but require parameter calibration and boundary conditions that drift over time. Data-driven health models correlate sensor readings with failures but struggle to extrapolate beyond observed degradation patterns and offer limited interpretability.

Both approaches share a common limitation: **they do not continuously infer internal thermal state**, which is the primary driver of degradation.

PINNs-Based Parameter and State Inference

In this workflow, a Physics-Informed Neural Network is used to infer internal temperature fields consistent with observed telemetry and known thermal dynamics. Sparse external measurements anchor the inference, while the network recovers internal temperature evolution over long horizons.

Crucially, the PINN enables inference not only of temperature fields but also of latent parameters associated with thermal transport and heat generation that evolve as assets age. This allows degradation indicators to be computed from inferred internal conditions rather than estimated indirectly.

The approach treats degradation as a consequence of inferred thermal history, not as a separate predictive task.

Operational Use

Integrated into asset management platforms, the inferred thermal and degradation states support condition-based maintenance and more accurate remaining useful life estimation. Maintenance actions can be prioritized based on inferred internal stress rather than conservative schedules or threshold breaches.

Operationally, this reduces unplanned outages, avoids premature replacement, and improves confidence in long-term planning decisions.

The value lies in **continuous inference of internal condition from sparse operational data**, not in replacing existing protection or inspection mechanisms.

7. Operational Reality and Trade-offs

Physics-Informed Neural Networks introduce capabilities that are difficult to achieve with conventional approaches, but they do so at a cost. In enterprise settings, these costs are not theoretical; they directly affect delivery timelines, operational risk, and long-term ownership. Any serious deployment must account for these trade-offs explicitly.

7.1 Training Cost and Engineering Overhead

PINNs are optimization-heavy. Training requires repeated evaluation of differential operators, careful scaling of variables, and deliberate loss balancing between physical constraints and observational data. This makes training more computationally expensive and more sensitive to formulation choices than standard machine learning models.

In practice, successful implementations treat training as an engineering task rather than an automated pipeline. This includes problem-specific normalization, staged training strategies, and manual inspection of convergence behavior. Teams without strong numerical and domain expertise will struggle to reach stable solutions.

The implication is clear: PINNs are unsuitable for rapid prototyping or low-effort deployment models.

7.2 Convergence Risk and Model Validity

Unlike supervised learning, PINNs do not fail loudly when assumptions break. A model may converge numerically while violating physical intuition due to poor constraint weighting, insufficient coverage of the domain, or incorrect boundary representations.

For this reason, convergence must be evaluated using **physical diagnostics**, not loss values alone. Residual behavior, boundary consistency, and stability under perturbation are more meaningful indicators of validity than aggregate error metrics.

Enterprise deployments must include explicit checks for when the system operates outside the domain for which the model is valid. Silent extrapolation is unacceptable in safety- or cost-critical systems.

7.3 Deployment and Runtime Constraints

PINNs are typically trained offline and used online for inference. While inference is significantly cheaper than training, it is still heavier than simple regression models. Latency constraints therefore limit where and how PINNs can be embedded.

In the workflows described in this paper, PINNs are deployed as **state inference components**, not as real-time control decision engines. This separation reduces operational risk and allows existing certified control logic to remain unchanged.

This architectural choice is not optional; it is a prerequisite for adoption in regulated or safety-critical environments.

7.4 Maintenance, Drift, and Recalibration

Physical systems change over time. Materials age, operating envelopes shift, and boundary conditions evolve. PINNs do not eliminate drift; they make it visible.

Recalibration strategies must be defined upfront. In some cases, retraining is required. In others, updating boundary or latent parameters is sufficient. The cost of retraining must be weighed against the operational value of improved inference.

Ownership does not end at deployment. PINNs demand lifecycle management comparable to that of high-fidelity simulation tools, not lightweight analytics models.

7.5 Summary of Trade-offs

PINNs trade simplicity for observability. They reduce uncertainty about hidden internal states but increase system complexity and engineering effort. They are justified only when the cost of operating blindly exceeds the cost of building and maintaining physics-informed inference.

This trade is acceptable in the domains examined in this paper. It is not acceptable everywhere.

8. When Physics-Informed Neural Networks Should Not Be Used

Physics-Informed Neural Networks are not a default choice. In many enterprise scenarios, they are the wrong tool. This section defines clear non-use conditions to prevent misapplication and false confidence.

8.1 Discrete, Event-Driven, or Transactional Systems

PINNs rely on continuous governing dynamics. Systems dominated by discrete events—transaction processing, workflow orchestration, user behavior modeling, fraud detection, or scheduling—do not satisfy this requirement.

In such systems, enforcing differential constraints adds no value and introduces unnecessary complexity. Classical machine learning, rules-based logic, or optimization methods are more appropriate.

8.2 Problems Without Well-Defined Governing Dynamics

If the underlying physics is unknown, weakly understood, or dominated by unmodeled stochastic effects, PINNs should not be used. Enforcing incorrect or oversimplified equations biases the solution and reduces robustness.

PINNs do not discover physics. They assume it.

When the governing equations cannot be written down with confidence, unconstrained or weakly constrained data-driven methods are preferable.

8.3 Scenarios With Dense, High-Quality Measurements

When internal states are already well observed through dense, reliable sensing, the primary advantage of PINNs disappears. In these cases, numerical solvers or supervised learning models can achieve higher accuracy with lower computational and engineering overhead.

Using PINNs in such settings increases system complexity without improving decision quality.

8.4 Applications Where Latency Dominates Fidelity

PINNs are not lightweight models. Although inference is cheaper than training, it remains more expensive than simple regressors or reduced-order models.

In applications where strict real-time constraints dominate and approximate state estimates are sufficient, classical control observers or empirical models are a better fit.

8.5 Use Cases Requiring Deterministic Guarantees

In some regulated or safety-critical contexts, deterministic analytical models are required regardless of accuracy limitations. PINNs, like other neural network–based methods, do not provide formal guarantees of correctness or bounded error under all conditions.

Where certification requires explicit analytical traceability, PINNs should be used only as advisory tools, if at all.

8.6 Summary of Non-Applicability

PINNs should not be deployed to:

- replace numerical solvers in well-specified problems,
- outperform machine learning on densely labeled data,
- compensate for missing domain understanding,
- or simplify system engineering effort.

They are justified only when the cost of operating with partial observability exceeds the cost of building and maintaining physics-informed inference.

9. Conclusion

This paper examined a specific and recurring class of enterprise problems: operating physical systems governed by known dynamics with incomplete visibility into their internal state. In aviation and energy infrastructure, this mismatch between system behavior and observability introduces persistent risk, inefficiency, and conservatism that cannot be resolved through incremental improvements to sensing, simulation, or data-driven modeling alone.

Physics-Informed Neural Networks were evaluated not as a general-purpose AI technique, but as a constrained inference mechanism for these settings. Their value does not lie in superior numerical accuracy or broad applicability. It lies in their ability to recover physically consistent internal states and parameters from sparse operational data when classical solvers require unavailable inputs and unconstrained machine learning fails outside observed regimes.

Across six workflows spanning aviation and energy systems, a single canonical pattern was shown to apply: physics-constrained state inference under partial observability, deployed as an augmentation to existing monitoring and control architectures. The approach enables earlier detection of abnormal conditions, improved control stability, and more informed operational decisions without replacing certified systems or requiring dense instrumentation.

The paper was explicit about limits. PINNs introduce computational cost, formulation risk, and ongoing maintenance burden. They are inappropriate for discrete systems, poorly understood dynamics, or environments with abundant high-quality measurements. Their use is justified only when the cost of operating blindly exceeds the cost of building and sustaining physics-informed inference.

The central conclusion is deliberately narrow:

For partially observable, physics-governed enterprise systems, Physics-Informed Neural Networks provide a viable and repeatable approach to state and parameter inference that is otherwise difficult to achieve in production environments.

Nothing more is claimed. Nothing less is required.