Model-assisted optimal control framework for industrial system coupling problems

Wei, X.

Citation

Wei, X. (2026, July 7). *Model-assisted optimal control framework for industrial system coupling problems*. Retrieved from https://hdl.handle.net/1887/4307775

CHAPTER

2

Industrial System Coupling and Model-assisted Optimization Control

In this section, we provide a detailed analysis of coupling phenomena in industrial systems and the resulting challenges for optimal control. To address these issues, we introduce the model-assited optimization approach and, based on this methodology, propose a comprehensive solution framework.

2.1 Coupling in Industrial Systems and Their Impact on Optimal Control

Industrial coupling problems can be categorized into three types: coupling between industrial systems and the environment, coupling among components within industrial systems, and coupling between industrial systems and working media. Industrial system-environment coupling operates at the macro scale, focusing on the boundary interactions between the system and its external surroundings (Chen et al. 2023; Bhattacharya et al. 2023). It examines the mutual influences between the system and the environment and analyzes how environmental changes affect system operation during runtime. coupling among components within industrial systems functions at the meso-scale, emphasizing functional synergy among internal subsystems or units (Meng et al. 2022). Rooted in a holistic view of the system, it studies the mutual impacts between components and explores coordinated control strategies under tight-coupling conditions. Industrial system-working media coupling is addressed at the micro scale, concentrating on how the physical properties of the media influence system control (Huang et al. 2022; Chang et al. 2023; Wu et al. 2023). It reflects the constraints imposed by underlying physical mechanisms on system performance and investigates control methods to suppress interference from the media during the control process.

From the perspective of engineering hierarchical thinking, these three categories of coupling span from internal to external, macro to micro, and system to component levels, forming a progressive layered structure. It can be argued that virtually all coupling problems in industrial contexts can be classified into these three types, indicating the completeness of this classification. The following sections will elaborate on how coupling phenomena in these three dimensions impact optimal control.

2.1.1 Coupling Between Industrial Systems and the Environment

In many industrial systems, the external environment significantly impacts equipment performance (Xia et al. 2025). When environmental information is fully observable and interactions with the equipment are fully understood, this information can be leveraged to make rapid and precise control decisions (Arcieri et al. 2024). However, in most real-world applications, these environmental factors are often not fully observable (Kurniawati et al. 2022; Lauri et al. 2022; Morad et al. 2023). Environmental uncertainty extends beyond the initial state, encompassing continuous changes in external conditions (Zhang et al. 2024). For instance, in the control of autonomous vehicles, adaptive cruise control and collision avoidance require maintaining a safe distance based on road conditions and surrounding traffic (Shamsah et al. 2023; Guo et al. 2022). While onboard radar/cameras can detect the position and speed of leading vehicles, they cannot directly measure the friction coefficient between tires and the road surface, nor can they directly infer the acceleration or lane-changing intentions of nearby vehicles (Zheng et al. 2025; Xia et al. 2023). In UAV flight control, although the drone's position, altitude, and velocity can be measured via sensors, sidewinds, gusts during stable motion or hovering, and complex turbulence generated near structures are not directly observable (Ye et al. 2022; Xu et al. 2025). Similarly, in industrial robotic tasks such as grinding and assembly, the actual contour of a workpiece may deviate from the CAD model input into the computer (Xie et al. 2022; Jiang et al. 2022). Relying solely on the designed contour for machining may damage the workpiece, and such dimensional uncertainty is often random and unpredictable.

A control problem can be described as a Markov Decision Process (MDP) using the tuple: $M\left(S, A, T, R, \Omega, O, \gamma\right)$, where S denotes the state space, A represents the action space, T represents the state transition probability, $R: S \times A \rightarrow \mathbb{R}$ is the reward function, Ω is the set of observations, O is the conditional observation probability, and γ is the discount factor (Lavaei et al. 2022, Heinbach et al. 2023).

Fully Observable Markov Decision Process

The Fully Observable Markov Decision Process (MDP) is the most fundamental and core sequential decision-making model. It describes how an agent, in an ideal world where environmental states are fully observable and environmental dynamics are known, can maximize long-term cumulative rewards through a series of actions. A fully observable MDP is defined by a tuple: $M\left(S, A, T, R, \gamma\right)$, where S denotes the state space, A represents the action space, T represents the state transition probability, $R: S \times A \rightarrow \mathbb{R}$ is the reward function, and γ is the discount factor (Mahajan et al. 2024; Kurniawati 2022).

In a Fully Observable Markov Decision Process, the conditional probability distribution of the future state s_{t+1} depends solely on the current state s_t and the current action a_t, and is independent of any earlier history $(s_0, s_1, \ldots, s_{t-1})$ (Ororbia et al. 2022). The state transition can be expressed probabilistically as:

$$P(s_{t+1} \mid s_t, a_t, s_{t-1}, a_{t-1}, \ldots) = P(s_{t+1} \mid s_t, a_t) \tag{2.1}$$

The current state s_t is a sufficient statistic for decision-making. Eq. (2.1) indicates that in a Markov Decision Process, you do not need to remember all past interactions; knowing only what the environment looks like at present is sufficient to make optimal decisions about the future. This greatly simplifies the problem structure and serves as the foundation for almost all efficient algorithms currently in use.

In a fully observable environment, the agent can perceive the complete and true state s_t of the environment at each time step t without delay, noise, or ambiguity. Since the state is complete and satisfies the Markov property, the optimal policy π can be simply defined as a mapping from the state space S to the action space A. In this case, the optimal value function $V_*(s)$ and the optimal action-value function $Q^*(s,a)$ satisfy the self-consistent Bellman optimality equations as follows:

$$V_*(s) = \max[R(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a) V_*(s')] \tag{2.2}$$

As a fixed-point equation, the Bellman equation provides a clear iterative objective for algorithms. In the process of achieving an optimal policy, one only needs to interact with the environment to realize state transitions, compute the value function, and ultimately obtain the optimal policy through iterative training.

It is precisely these characteristics that make the fully observable MDP not only the theoretical core of reinforcement learning but also the most powerful and commonly used conceptual model for understanding, analyzing, and designing complex decision-making systems. It provides a clear ideal objective: if I could know everything, how should I act optimally? Real-world partially observable problems can then be viewed as this ideal objective overlaid with the additional challenge of “information acquisition.”

Partially Observable Markov Decision Process (POMDP)

The Partially Observable Markov Decision Process refers to practical systems in which the decision-maker cannot directly access the complete state of the environment. A Partially Observable Markov Decision Process can be represented as a tuple: $M(S, A, T, R, \Omega, O, \gamma)$, where S denotes the state space, A represents the action space, T represents the state transition probability, $R : S \times A \to \mathbb{R}$ is the reward function, Ω is the set of observations, O is the conditional observation probability, and γ is the discount factor. In this formulation, the state space S consists of two parts: the observable state s_t^o and the unobservable state s_t^h (Bennett et al. 2024, Shi et al. 2022). Unlike in a fully observable Markov Decision Process, the current observable state s_t^o does not contain all the information required for future predictions. Therefore, s_t^o does not satisfy the Markov property, i.e.:

$$P(s_{t+1}^o \mid s_t^o, a_t, s_{t-1}^o, a_{t-1}, \ldots) \neq P(s_{t+1}^o \mid s_t^o, a_t) \tag{2.3}$$

In such a scenario, relying solely on the observable state s_t^o cannot achieve accurate decision-making. In a Partially Observable Markov Decision Process, it is necessary to estimate a belief state $\hat{s}_t^h$ that represents the subjective assessment of the unobservable state

s_t^h , thereby compensating for the impact of missing information on decisions (Miao et al. 2022). The belief state is a subjective estimation of the unobservable state based on available knowledge (Huang et al. 2024). This estimated future belief depends only on the current belief and the action taken - in other words, it satisfies the Markov property.

A POMDP is not a simple extension of an MDP, but rather an honest acknowledgment and systematic treatment of the fundamental uncertainties present in the real world. It tells us that in an information-incomplete world, optimal decision-making depends not only on what you currently know, but also on how you act to reduce uncertainty. In summary, under conditions of partial observability, relying solely on directly measurable information cannot achieve precise control decisions. Optimal control relies on an accurate understanding of and feedback from the environment. When it is impossible to directly obtain all environmental state information, the accuracy and reliability of optimization results are inherently limited. Due to environmental uncertainty, the optimization scheme must possess adaptability - able to anticipate environmental changes, perform latent-state inference of unobservable states, and ultimately integrate observable states to make final decisions.

2.1.2 Coupling Among Components Within Industrial Systems

In complex systems, there are tightly coupled relationships among various components, where the optimization objectives of different parts often conflict with each other (Wang et al. 2024, Grieves 2024). Local optimization typically cannot guarantee the overall optimality of the system; improving the performance of one component may compromise the performance of others, ultimately reducing the overall efficiency of the system. Multi-variable coupling leads to an increase in the dimensionality of the problem space, making the solution space of the optimization problem highly intricate (Fang et al. 2022).

Before performing optimization, it is first necessary to consider how to characterize such complex coupling relationships. The optimization problem under this type of coupling can be described as:

$$
\begin{aligned}
&\min_{x_1,\dots,x_n} F(x_1,\dots,x_n) = \Phi(f_1,\dots,f_n) \\
&\text{s.t. } g_i(x_i, x_{-i}) \le 0, \quad i = 1,\dots,n
\end{aligned}
\tag{2.4}
$$

where $x_i \in \mathbb{R}^{d_i}$ represents the state of the *i*-th component, Φ is the aggregation function, f_i is the local objective function of the *i*-th component, and g denotes the constraint function. When the states x_i in F are highly coupled, local optimization will conflict with the effect of global optimization. Moreover, as the number of components increases, the dimensionality of the entire problem's search space grows exponentially, significantly raising computational costs.

Furthermore, in industrial coupling systems, due to the highly nonlinear interactions between local parameters and global performance, minor computational deviations can be progressively amplified through close component correlations, leading to fundamental errors in predicting the overall system behavior. Low-fidelity models cannot accurately capture the physical interactions and energy transfer mechanisms under such strong coupling (Feng et al. 2023). Their simplified assumptions may obscure critical modes and instabilities present in the real physical process, thereby rendering simulation results ineffective in guiding the actual system dynamics (Xu et al. 2024; Lyu et al. 2023).

The optimization of coupled systems inherently involves navigating a high-dimensional non-convex solution space, where objective functions and constraints interweave to form a complex topological structure. Low-fidelity simulations, which simplify the underlying physical mechanisms, tend to distort the geometric characteristics of the solution space, smoothing over critical local extrema and feasible-region boundaries (Zhang et al. 2022; Liu et al. 2023). This can cause optimization algorithms either to become trapped in spurious local optima or to entirely miss the true global optimum region. Insufficient computational accuracy fundamentally undermines both the convergence of the optimization process and the guarantee of solution optimality.

Control and decision-making in industrial systems rely on accurate prediction of dynamic evolution trajectories. Coupled dynamics exhibit sensitivity to initial conditions and parameters. Due to oversimplification of physical mechanisms, low-fidelity models introduce unquantifiable cumulative errors, severely compromising the reliability of both short-term predictions and long-term projections.

Only high-fidelity simulations can provide a reliable environment for the formulation and validation of coordinated control strategies, ensuring that decisions explored in virtual space can be executed safely and effectively in the physical world (Veers et al. 2022; Liu et al. 2022). However, high-precision models often incur substantial computational costs. First, the model must simultaneously resolve strongly coupled multi-physics equations and accurately capture the bidirectional feedback mechanisms among different fields. For example, in fluid–structure interaction problems, it is necessary to solve both the Navier–Stokes equations to describe the flow and the solid mechanics equations to describe deformation, requiring data exchange and iterative convergence at each time step and spatial node (Sheng 2022). This leads to a combinatorial increase in computational dimensionality, resulting in the "curse of dimensionality," whereby the required computational resources grow exponentially with problem scale.

Second, to accurately capture the nonlinear, non-stationary characteristics and critical details of the system, the model requires extreme discretization refinement in both time and space. The time step must be sufficiently small to capture transient processes and high-frequency dynamics, while the spatial grid must be highly refined to resolve local singular regions such as boundary layers, shock waves, and interfaces. Such fine-grained discretization leads to a vast number of unknowns, resulting in large-scale, ill-conditioned linear systems that must be solved at each iteration, imposing significant demands on memory, communication, and solution algorithms (Kudela et al. 2022; Fang et al. 2024).

Finally, high-precision solutions typically require the use of high-order numerical schemes and tight iterative tolerances (Jia et al. 2022). Although high-order schemes improve accuracy, they employ wider computational stencils and exhibit stronger data dependencies, which hinder parallel computing efficiency and can introduce numerical oscillations, necessitating complex stabilization treatments. Meanwhile, stringent convergence criteria force each iteration step to perform more sub-iterations, effectively doubling the overall computational effort. Furthermore, to assess model uncertainty and perform parameter optimization, it is often necessary to repeat simulations across multiple scenarios or parameter samples, further amplifying the total computational cost.

In scenarios characterized by high problem dimensionality, strong coupling, and expensive simulation models, employing heuristic iterative search methods — such as evolutionary algorithms—to obtain optimal control parameters is not appropriate. The reason lies in the irreconcilable conflict between computational cost and search efficiency.

Developing a method with lower computational overhead is therefore key to solving optimization problems in this class of coupled scenarios (Schwarz et al. 2022).

2.1.3 Coupling Between Industrial Systems and Working Media

The coupling between industrial systems and working media (such as gases and liquids) often exhibits marked nonlinearity and hysteresis, arising from the dependence of medium physical properties (e.g., density, viscosity) on the system state, as well as the finite time required for energy and mass transfer and accumulation within the equipment (Kundu et al. 2024; Peng et al. 2024). This nonlinearity manifests as a complex mapping between system inputs and outputs, including saturation, dead zones, and hysteresis rather than a simple proportional relationship (Nash et al. 1995; Bilal et al. 2024). The time-lag effect implies that the influence of a current control action appears in the controlled variable only after a certain delay (Ghasemzadeh et al. 2023, Nanmaran et al. 2023). Considering a controlled variable, its core dynamics can be abstracted into a delayed differential equation of the following form: the system dynamics under such effects can be represented as:

$$\tau_c \frac{dy(t)}{dt} + y(t) = f(u(t-\tau_d)) + d(t) \tag{2.5}$$

where $y(t)$ represents the output of the system, $u(t)$ is the input of the system, τ_c denotes the time constant, f is the nonlinear static mapping of the system, and $d(t)$ represents external disturbance. At any time *t*, the factor determining the rate of change of $y(t)$ (i.e., dy/dt) is not the current control action u(t), but rather its historical value $u(t-\tau_d)$. In other words, the “driving force” of the system at time t is determined by a decision made τ_d units of time earlier. Under this effect, control shifts from an instantaneous response problem to a sequential planning problem requiring forward prediction. This directly poses the following challenges to system safety:

First, time delay directly threatens the real-time safety-assurance capability of control (He et al. 2023). In safety-critical scenarios, the latency between detecting a hazard and executing a protective action may imply that the system has already crossed an irreversible safety boundary (Lu et al. 2023). For instance, in fluid machinery control, the delay between detecting compressor outlet pressure and adjusting valve opening may lead to loss of control, resulting in instability or surge (Shahriyari et al. 2024). Time delay keeps the control system in a state of “reactive lag” when responding to sudden threats, forcibly compressing the safety margin. The system must rely on higher-accuracy predictions and earlier warnings to compensate for this “uncontrolled period.”

Second, time delay exacerbates system uncertainty and fragility, conflicting with the inherent robustness requirements of safety-critical control (Deng et al. 2022). Time delay amplifies the impact of model errors and disturbances—a predictive safety decision based on an erroneous model may become entirely ineffective when executed after the delay due to accumulated deviations (Wu et al. 2024). Furthermore, time delay dramatically complicates the dynamic management of safety boundaries (Lu et al 2022; Makhbouche et al. 2023). In static or delay-free systems, safety boundaries are typically expressed as static constraints in the state space. However, again taking fluid machinery as an example, when delay is present, current safety depends not only on the immediate state but also on historical control commands that are still “in transit” and have not yet taken effect. This transforms the safety boundary into a time-varying, history-dependent complex set. The system may appear entirely normal based on current measurements, yet an issued but not-yet-effective aggressive command may be pushing it toward danger.

Therefore, it is imperative to develop a safe controller that addresses the impact of system time delay. First, such a safety controller must possess “foresight” capability, enabling feedforward control based on system states to enhance response speed. Second, the controller must accommodate wide-range state variations in the system, i.e., exhibit high adaptability to control tasks under different operating conditions (Huang et al. 2022).

2.2 Model-assisted Optimal Control

In industrial coupling problems, models often play a decisive role (Fang et al. 2022). The essence of industrial coupling lies in interaction and mutual constraint. These models provide a means to understand and quantify such interactions (Tsaramirsis et al. 2022). The primary challenge of industrial coupling is to achieve a structured understanding of the interactions among different parts of a system (Kenett et al. 2022). Precise mathematical models, based on system operational principles, can translate vague “mutual influence” into explicit mathematical relationships (Buede et al. 2024). This formalization not only makes coupling quantifiable but also reveals the underlying mechanisms of coupling. Secondly, in high-cost, high-risk industrial scenarios where frequent physical experiments are impractical, models serve as indispensable virtual testbeds. They allow for the exploration of extreme operating conditions (e.g., investigating compressor surge boundaries) without physical risks and enable large-scale parameter scanning at speeds far exceeding real-time rates. More importantly, models are tools capable of providing future state predictions—for time-delay systems, current measurements only reflect the outcomes of past decisions; only models can project future system behavior based on current inputs and historical trajectories. This predictive capability forms the cornerstone of advanced algorithms such as model predictive control, shifting control from passive response to proactive guidance.

Model-assisted optimization (MAO) refers to a class of solution strategies that closely integrate mathematical models with optimization algorithms. Here, the term “model” does not denote the mathematical model of the optimization problem itself, but rather an additional simulation model or surrogate model capable of simulating the relationship between system inputs and outputs. Its fundamental purpose is to address complex optimization problems where the “evaluation” of the objective function or constraints is extremely costly or lacks an explicit mathematical expression.

In MAO, commonly used modeling methods mainly fall into three categories: simulation modeling based on computer-assisted engineering, data-driven modeling, and mathematical modeling based on physical mechanisms.

Computer-assisted engineering simulation modeling mainly includes numerical modeling methods such as the finite element method, finite volume method, and computational fluid dynamics (Zhang et al. 2023). It approximates real physical fields through highly refined discretization (e.g., mesh generation), offering high fidelity and multi-physics coupling capabilities to handle complex physical partial differential equations (Zhao et al. 2024). However, high precision also implies high computational costs. Consequently, integrating this type of model with heuristic or data-driven optimization algorithms poses significant challenges (Wang et al. 2022).

Data-driven modeling refers to models that learn the mapping between inputs and outputs from data without explicit reliance on physical principles (Wang et al. 2022; Wang et al. 2022, Zubair et al. 2024). Such models are characterized by "train-once, fast-evaluate" behavior. Among them, surrogate modeling represents one of the most important and typical applications of data-driven models in engineering optimization (He et al. 2023). A surrogate model uses data generated by a computationally expensive high-fidelity model to build a low-cost mathematical approximation (Ma et al. 2025; Ghafariasl et al. 2024). By running a limited number of high-cost simulations in advance to collect data, and then training a fast approximate model with this data, subsequent operations such as optimization, sensitivity analysis, and visualization are performed on the fast surrogate, thereby dramatically reducing the overall computational time (Cheng et al. 2024; Ling et al. 2022). Once trained, a surrogate model typically requires only milliseconds or microseconds for a single prediction, which enables the use of computationally intensive optimization algorithms (e.g., genetic algorithms) to tackle complex engineering problems (Wang et al. 2022). It should be noted that the accuracy of data-driven surrogate models is often lower than that of high-fidelity data sources. This is primarily because such models are trained on limited data. Establishing an accurate surrogate model generally requires the combined contribution of high-fidelity data sources, precise experimental design, and well-structured training procedures.

Physics-based modeling is the most classical and fundamental modeling approach in engineering and scientific fields. Also known as white-box models or first-principles models, they are constructed based on fundamental physical laws and conservation principles (such as Newton’s laws, energy conservation, mass conservation, momentum conservation, Maxwell’s equations, etc.) (Yu et al. 2023; Rajulapati et al. 2022). In such models, the behavior of the system is fully described and determined by its inherent physical mechanisms (Sheng et al. 2024). This description typically takes the form of ordinary differential equations (ODEs), partial differential equations (PDEs), or differential-algebraic equations (DAEs) (Gao et al. 2025, Brunton et al. 2024, Huang et al. 2024). These models offer high interpretability, transparency, and strong extrapolation capability. However, when dealing with complex phenomena such as multiphase flow, a single mechanistic model often cannot achieve accurate simulation and prediction. Therefore, physics-based models are frequently employed in localized studies of specific characteristics of real-world application scenarios.

It is worth mentioning that, building upon these three types of models, digital twin models have garnered widespread attention in the field of industrial optimization (Xia et al. 2023; Liu et al. 2022). A digital twin is a dynamic, continuously updated virtual representation of a physical system. It fully utilizes data from physical models, sensor updates, operational history, etc., to create a mapping in virtual space that reflects the full lifecycle process of the corresponding physical equipment (Goodwin et al. 2024). Through bidirectional interaction with the physical system, digital twin models are characterized by real-time responsiveness and synchronization (Liu et al. 2021). By integrating multi-physics modeling, they enable predictive maintenance and optimization of the physical system throughout its lifecycle. Based on these features, optimization driven by digital twins often exhibits online, dynamic, and adaptive characteristics.

At its core, optimization is the process of finding the best possible solution for a given objective function under specified constraints (Bäck et al. 1993). When this theoretical framework is applied to dynamic systems with the aim of designing control strategies that achieve desired performance metrics, it gives rise to the field of optimal control. In industrial coupling problems, appropriate models for different coupling scenarios can provide rigorous mathematical formulations, which form the foundation for the successful application of optimization algorithms in industry and for achieving accurate and safe optimal control.

2.3 Optimal Control Framework Design for Industrial Coupling Problems

For different coupling scenarios, this thesis proposes an optimization framework tailored to industrial coupling problems. Specifically, based on the characteristics of optimization challenges in each coupling context, the framework introduces distinct modeling approaches and improvements to optimization algorithms:

Coupling between industrial systems and the environment

Taking TBM trajectory navigation as an example, the key issue lies in achieving accurate optimal control under partially observable environmental states. In this framework, we propose a method that leverages historical time-series information from the control process to predict and compensate for missing environmental data required for optimal control. Moreover, to accurately simulate equipment-environment interaction, fragmented modeling techniques such as finite element methods are typically needed to build multi-field models for both the equipment and the environment. However, due to the prohibitively high computational cost of traditional finite-element models, we developed a simplified Patch-based Contact Dynamics Model (PCDM) that efficiently simulates TBM motion and equipment-environment interaction.

Coupling among components within industrial systems

Precise real-time simulation of dynamic interactions between components requires a high-fidelity digital twin. The difficulty arises because high-fidelity dynamic simulation models are computationally expensive, and coordinating multiple simulation parameters across system components drastically expands the search space, leading to an exponential increase in the number of joint simulations. Hence, the main challenge in this scenario is how to avoid extensive iteration while completing the optimization task. In this framework, we propose a one-shot optimization method for industrial control parameters based on surrogate modeling. Using the reflow soldering process as a case study, a Transformer-CNN architecture is employed to capture global temporal and local spatial features, constructing a dual inverse mapping from requirements to targets and from targets to parameters, thereby enabling inverse design of temperature-control parameters.

Coupling between industrial systems and working media

Illustrated by compressor coordinated control, the capacitive nature of the working medium introduces strong nonlinearity into the system, often manifested as significant response lag and severe overshoot. Such overshoot increases the risk of the compressor's operating point entering the surge region. To address this, we propose a safety-oriented coordinated control strategy for compressors based on the CMA-ES algorithm. By employing evolutionary computation to optimize PID controller parameters, the strategy maintains system response speed while effectively reducing overshoot and ensuring operational safety.

Although the three coupling scenarios differ in their physical nature and mathematical formulation, the proposed framework is built upon a set of shared design principles that transcend individual applications:

Model Role Allocation: In each scenario, a model is not simply a passive simulator but is assigned a specific role determined by the coupling characteristics. The role dictates what the model provides to the optimization process — training data (Chapter 3), a high-fidelity surrogate (Chapter 4), or a safety boundary quantifier (Chapter 5).

Problem Restructuring: Rather than directly applying optimization algorithms to the original problem, the framework restructures the problem into a form that aligns with the assigned model role. This includes decomposing a POMDP into hierarchical layers (Chapter 3), converting iterative search into inverse mapping learning (Chapter 4), and reframing the control objective from minimizing response time to planning feasible trajectories within safety constraints (Chapter 5).

Algorithm-Model Synergy: The optimization algorithm is selected or designed not in isolation, but based on the model's role and the restructured problem. The algorithm compensates for what the model lacks—temporal inference for partial observability (GRU in Chapter 3), shape-to-parameter mapping for high-dimensional coupling (CNN in Chapter 4), and derivative-free multi-variable optimization for nonlinear hysteresis (CMA-ES in Chapter 5).

While the three coupling scenarios share a common design philosophy, their fundamental differences justify the use of distinct modeling and optimization strategies. Understanding these differences is essential for readers to appreciate why a single method cannot address all industrial coupling problems.

System-Environment coupling (Chapter 3) is fundamentally an information problem. The system dynamics are relatively well-understood, but the environment is only partially observable. The key challenge is not the complexity of the system itself, but the missing information required for accurate decision-making. This is why the framework for this scenario focuses on inference—extracting latent environmental states from observable temporal patterns—rather than on system simplification or speedup.

Inter-Component coupling (Chapter 4) is fundamentally a complexity problem. The physics are fully observable and well-understood in principle, but the high-dimensional parameter space and expensive simulations make traditional iterative optimization computationally prohibitive. The key challenge is not missing information, but computational tractability. This is why the framework for this scenario centers on learning an inverse mapping that bypasses the need for iterative search entirely.

System-Working Medium coupling (Chapter 5) is fundamentally a safety problem. The system is nonlinear and time-delayed, which means that aggressive control actions can inadvertently push the system into unsafe regions (e.g., surge in compressors). The key challenge is not information or computation, but managing the trade-off between responsiveness and safety. This is why the framework for this scenario explicitly incorporates a physics-based safety boundary into the optimization objective.

This contrast—information, complexity, safety—can serve as a guide for practitioners facing new industrial coupling problems. By diagnosing which of these three challenges is dominant in their own system, they can select the appropriate model role and optimization strategy from the framework presented in this thesis.

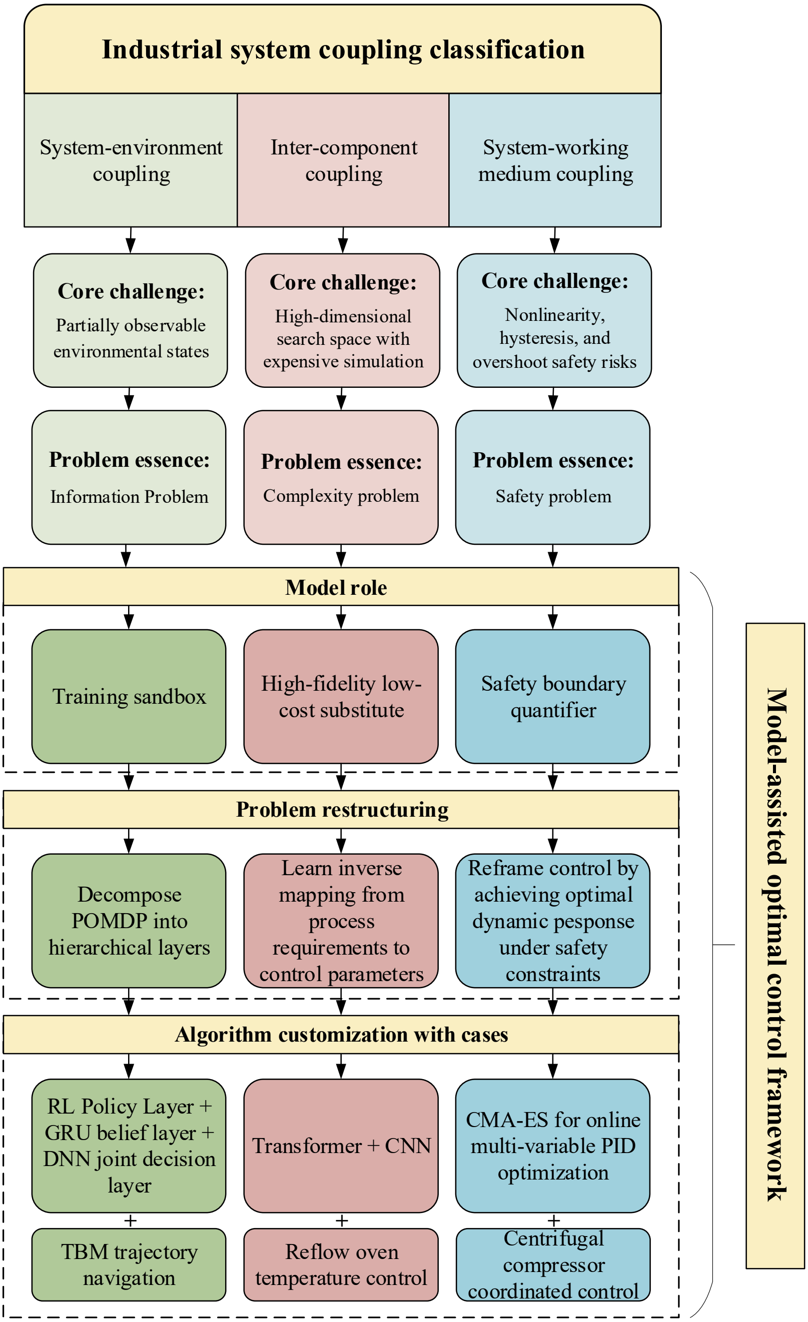

Figure 2.1: The structure of model-assisted optimal control framework.