buildings

MDPI

Article

Kinetic-Aware Distributionally Robust HVAC Optimization for Multi-Zone Building Systems with Physics-Informed Reinforcement Learning

Zhiyuan Sun [1] and Alexis P. Zhao [2,*]

[1] School of Media and Arts, International Education College, Tianjin University of Sport, Tianjin 301617, China; zhiyuandance@tjus.edu.cn
[2] Stanford Doerr School of Sustainability, Stanford University, Stanford, CA 94305, USA
* Correspondence: alexzhao@stanford.edu

Abstract

This study develops an advanced optimization framework for heating, ventilation, and air conditioning (HVAC) systems in multi-zone buildings with highly dynamic and uncertain internal heat loads. Unlike conventional models that assume static occupancy, the proposed approach captures time-varying, spatially heterogeneous thermal disturbances driven by occupant activity. The building is modeled as a coupled cyber-physical system integrating multi-zone thermal dynamics, nonlinear HVAC energy consumption, and behavior-driven heat generation. To address uncertainty, a distributionally robust optimization framework based on Wasserstein ambiguity sets is employed, enabling reliable performance without requiring precise probability distributions. In addition, a physics-informed reinforcement learning mechanism is incorporated to derive adaptive control policies while ensuring thermodynamic feasibility. A multi-zone coordination strategy is further introduced to manage spatial thermal interactions and maintain stable comfort across different areas. Case study results demonstrate that the proposed method reduces peak energy consumption by 28–32%, decreases comfort violation rates by 65–75%, and improves thermal stability, with temperature variance reduced by over 60% compared to baseline strategies.

Keywords: HVAC optimization; multi-zone building systems; distributionally robust optimization; physics-informed reinforcement learning; dynamic internal heat loads; building energy efficiency

1. Introduction

The rapid evolution of building energy systems toward more intelligent, adaptive, and human-centric paradigms has exposed critical limitations in conventional heating, ventilation, and air conditioning (HVAC) modeling and control strategies [1–3]. Traditional approaches to HVAC optimization are largely developed for static or quasi-static occupancy environments, where internal heat gains are treated as exogenous, slowly varying inputs [4,5]. While such assumptions are reasonable for office buildings, residential spaces, or commercial facilities, they fail to capture the fundamental characteristics of performance-driven environments such as dance theaters [6–8]. In these settings, thermal loads are not only highly dynamic but are intrinsically coupled with human motion, choreography structure, and spatial distribution of occupants [9–11]. Dancers exhibit metabolic rates that fluctuate significantly over time, often reaching levels several times higher than sedentary occupants, and their movement patterns generate non-uniform, rapidly shifting heat distributions across the performance

Academic Editor: Vincenzo Costanzo

Received: 25 March 2026
Revised: 20 April 2026
Accepted: 21 April 2026
Published: 5 May 2026

space [12–14]. At the same time, audiences and supporting areas impose distinct and often conflicting comfort requirements, creating a multi-regime thermal management problem that challenges existing HVAC frameworks.

Recent advances in building energy optimization have incorporated data-driven techniques, model predictive control, and reinforcement learning to enhance system adaptability [15,16]. However, these methods typically treat occupants as passive thermal sources and do not explicitly model the endogenous relationship between human behavior and energy demand [17–19]. Moreover, uncertainty in occupancy patterns, activity intensity, and environmental disturbances is often addressed through scenario-based or stochastic optimization methods that rely on predefined probability distributions [20–22]. Such approaches are inherently limited in their ability to capture the complex, structured variability associated with choreographed human motion, where both temporal and spatial correlations play a critical role [23–25]. As a result, there remains a significant gap in the literature regarding the development of optimization frameworks that simultaneously integrate physical system dynamics, behavioral uncertainty, and adaptive learning mechanisms in a unified and tractable manner.

This paper addresses these challenges by proposing a novel optimization framework for HVAC control in dance theater environments, where energy demand is explicitly driven by human motion and choreography. The core idea is to treat the building as a behavior-coupled cyber-physical system, in which thermal dynamics and human activity are co-evolving processes rather than independent components. To this end, we develop a comprehensive mathematical model that captures multi-zone thermal interactions, nonlinear HVAC energy consumption, and endogenous heat generation derived from dancer-specific metabolic functions. The modeling framework introduces a spatial–temporal mapping between choreography states and zone-level heat loads, enabling the system to anticipate and respond to dynamic performance patterns without relying on real-time sensing. This representation allows the HVAC system to operate in a predictive and proactive manner, aligning energy supply with anticipated human activity.

To account for uncertainty in human behavior and environmental conditions, the proposed framework adopts a distributionally robust optimization approach based on Wasserstein ambiguity sets. Instead of assuming a fixed probability distribution for uncertain parameters, the model considers a family of plausible distributions centered around empirical observations, thereby ensuring robustness against worst-case realizations of dancer activity and occupancy patterns. This formulation enables the system to maintain reliable performance under a wide range of operational scenarios while avoiding the conservatism associated with traditional robust optimization techniques. Furthermore, the interaction between HVAC control decisions and human behavior is modeled through a bi-level structure, in which dancers are treated as strategic agents whose performance outcomes depend on thermal conditions. This perspective fundamentally redefines occupants as active participants in the optimization process, introducing a feedback loop between energy systems and human responses.

Building upon this modeling foundation, the paper integrates a physics-informed reinforcement learning mechanism to derive adaptive control policies that generalize across varying performance scenarios. Unlike purely data-driven approaches, the learning component is constrained by thermodynamic principles, ensuring that the resulting policies remain physically feasible and interpretable. The reinforcement learning framework leverages state representations that incorporate thermal states, occupancy distributions, and choreography signals, and optimizes control actions to minimize a composite objective that includes energy consumption, comfort deviations, and robustness penalties. Through this integration of model-based optimization and learning-based adaptation, the proposed method achieves both reliability and flexibility, enabling efficient operation in highly dynamic environments. The contributions of this paper are fourfold. First, it introduces a new modeling paradigm for building energy systems that explicitly couples HVAC dynamics with human motion, moving beyond the traditional assumption of static or exogenous occupancy. Second, it develops a distributionally robust optimization framework that captures structured uncertainty in human behavior using Wasserstein ambiguity sets, providing a principled approach to handling variability in performance environments. Third, it formulates a bi-level interaction model that treats occupants as strategic agents, thereby incorporating human response into the optimization loop and enriching the behavioral realism of the system. Fourth, it integrates physics-informed reinforcement learning with thermodynamic constraints, enabling the derivation of adaptive and generalizable control policies that maintain both energy efficiency and occupant comfort. Collectively, these contributions establish a unified and scalable framework for next-generation HVAC optimization in performance-driven indoor environments.

The presented visual composition in Figure 1 exhibits a complex layered structure characterized by overlapping graphical elements and fragmented visual segments. The image appears to combine multiple interface windows, textual components, and background textures, resulting in a highly non-uniform and partially distorted representation. Such visual artifacts are indicative of rendering inconsistencies, which may arise from display buffering issues, compression distortions, or incomplete image reconstruction processes. From a structural perspective, the composition lacks clear spatial hierarchy, as different layers of information are superimposed without consistent alignment. This leads to reduced readability and ambiguity in interpreting the underlying content. The presence of repeated patterns and blurred textual regions further suggests that the image may have undergone transformation or partial corruption during capture or transmission. In the context of system visualization or data presentation, such irregularities highlight the importance of ensuring consistency in graphical rendering and data integrity. A clear and well-structured visual layout is essential for effective communication, particularly when conveying complex information. Therefore, addressing these distortions is critical to improving both interpretability and presentation quality in practical applications.

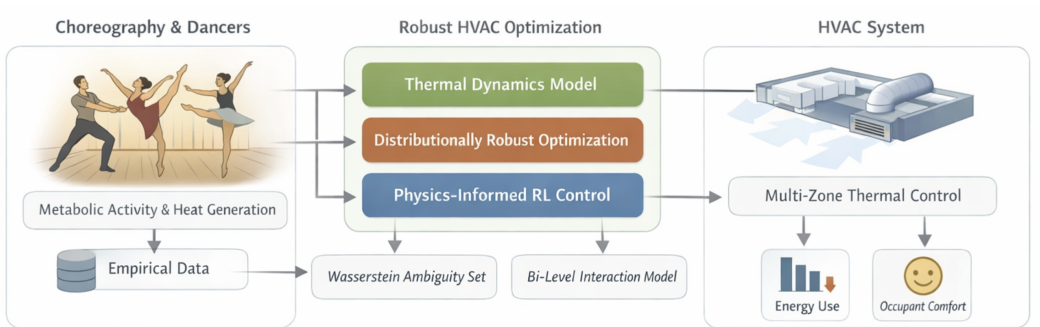

Figure 1. Behavior-coupled distributionally robust HVAC optimization framework for performance-driven indoor environments.

Table 1 provides a structured comparison between the proposed framework and existing approaches in behavior–thermodynamics coupling for building energy systems. The comparison highlights several key dimensions, including behavior modeling, thermal coupling, uncertainty handling, control strategy, and adaptability. Conventional models typically treat occupant behavior as static or exogenous inputs, which limits their ability to capture dynamic interactions between human activity and thermal processes. Behavior-aware models introduce partial dynamics, but often lack a unified representation of thermal coupling and control. Learning-based methods improve adaptability through data-driven approaches, yet they may not explicitly enforce physical consistency. In contrast, the proposed framework models behavior as an endogenous and time-varying process, enabling direct interaction with thermal dynamics in a multi-zone setting. Furthermore, uncertainty is addressed through a distributionally robust formulation, which enhances reliability under varying conditions. The integration of physics-informed reinforcement learning allows adaptive control while maintaining thermodynamic consistency. Overall, this comparison illustrates that the proposed approach provides a more comprehensive and coherent treatment of behavior, thermodynamics, and control, thereby distinguishing it from existing methodologies.

Table 1. Comparison with existing behavior–thermodynamics coupling approaches.

Feature	Conventional Models	Behavior-Aware Models	Learning-Based Methods	Proposed Framework
Behavior Modeling	Static/exogenous	Partially dynamic	Data-driven	Endogenous and time-varying
Thermal Coupling	Simplified	Limited interaction	Implicit	Explicit multi-zone coupling
Uncertainty Handling	Deterministic	Scenario-based	Stochastic learning	Distributionally robust (Wasserstein)
Control Strategy	Rule-based	Optimization-based	Reinforcement learning	Physics-informed RL + DRO
Adaptability	Low	Moderate	High	High with physical consistency

2. Mathematical Modeling

To rigorously characterize the complex interactions between thermal dynamics, HVAC control actions, and time-varying internal heat loads, this section develops a comprehensive mathematical formulation for multi-zone building systems. Unlike conventional building energy models that treat internal gains as exogenous and static, the proposed framework explicitly models internal heat generation as an endogenous, dynamically evolving process that is spatially distributed across zones and temporally coupled with system states. The building is represented as a multi-zone thermodynamic network in which temperature evolution is governed by inter-zone heat transfer, HVAC actuation, environmental disturbances, and occupant-induced heat loads. To capture the nonlinear and heterogeneous nature of these interactions, the model integrates detailed representations of HVAC energy consumption, thermal comfort for different occupant groups, and spatial allocation of internal loads. Furthermore, uncertainty in internal heat generation and operational conditions is incorporated through a distributionally robust formulation, enabling the system to hedge against variability while maintaining computational tractability. The resulting model provides a unified foundation for capturing both the physical and stochastic characteristics of building energy systems, which serves as the basis for subsequent optimization and control design.

From an implementation perspective, the spatial resolution adopted in this study is intended to balance modeling accuracy and practical feasibility. While fine-grained zoning can theoretically improve control performance by capturing detailed thermal heterogeneity, it may also require dense sensing infrastructure and more sophisticated actuation capabilities in real-world applications. In practice, the proposed framework does not rely on excessively high spatial resolution, but can be applied to moderate zoning configurations that are commonly supported by existing building management systems. The required

https://doi.org/10.3390/buildings16091839

sensing inputs, such as zone-level temperature and occupancy-related signals, are typically available in modern HVAC deployments, and the control actions can be implemented through standard airflow and ventilation adjustment mechanisms. Therefore, the spatial modeling approach should be interpreted as a flexible representation that can be adapted to different levels of system granularity, rather than a requirement for dense instrumentation. This design allows the framework to maintain practical applicability while still capturing the key spatial interactions necessary for improved control performance.

To improve the transparency and practical interpretability of the HVAC energy model, the calibration and validation procedures associated with each component are systematically summarized in Table 2. The table provides a structured overview of how the quadratic, bilinear, and exponential–logarithmic terms are calibrated and validated within the proposed framework. Specifically, the quadratic component captures the baseline energy consumption associated with actuation intensity and is calibrated using simulation-based parameter tuning, with validation performed through standard error metrics such as RMSE. The bilinear term reflects interaction effects between thermal states and control variables and is evaluated through sensitivity analysis across different operating regimes. The exponential–logarithmic component is introduced to represent nonlinear efficiency degradation under extreme conditions, and its parameters are calibrated using nonlinear fitting approaches while ensuring stable and bounded behavior under high-load scenarios. By explicitly summarizing the calibration workflow and validation criteria, this structured representation enhances the empirical grounding of the model and provides clearer support for the adopted functional forms, thereby improving the overall credibility and interpretability of the proposed approach.

Figure 1 illustrates a unified tri-layer architecture in which choreography-driven metabolic dynamics and empirical uncertainty jointly feed into a distributionally robust and physics-informed optimization engine, enabling multi-zone HVAC control to simultaneously minimize energy consumption while adaptively satisfying heterogeneous comfort requirements across dancers and audience.

$$\begin{aligned}\min_{\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{\varrho},\boldsymbol{\kappa}} \sup_{\mathbb{Q}\in\mathcal{P}(\hat{\mathbb{P}},\varepsilon)} \mathbb{E}_{\mathbb{Q}}\Bigg[\sum_{t\in\mathcal{T}}\sum_{z\in\mathcal{Z}}\Big(&\alpha^{\mathrm{e}}_{z,t}\cdot\Phi^{\mathrm{HVAC}}_{z,t}(\theta_{z,t},v_{z,t},\varrho_{z,t})+\beta^{\mathrm{d}}_{z,t}\cdot\big(\theta_{z,t}-\theta^{\mathrm{ref},\mathcal{D}}_{z,t}\big)^2\\ &+\gamma^{\mathrm{a}}_{z,t}\cdot\big(\mathrm{PMV}_{z,t}(\theta_{z,t},v_{z,t},\varrho_{z,t})\big)^2+\delta_{z,t}\cdot\Big\|\sum_{i\in\mathcal{I}}\mu_{i,t}\cdot\chi_{i,z,t}-\bar{\mu}_{z,t}\Big\|_2^2\\ &+\eta_{z,t}\cdot\Big(\theta_{z,t+1}-\theta_{z,t}-\Delta t\cdot\Xi_{z,t}(\theta,v,\varrho,\mu)\Big)^2\Big)\Bigg]\end{aligned} \tag{1}$$

Equation (1) establishes a distributionally robust, multi-objective optimization formulation in which the control variables $\boldsymbol{\theta}$, $\boldsymbol{v}$, and $\boldsymbol{\varrho}$ jointly determine the thermodynamic state evolution of the multi-zone dance theater, while $\boldsymbol{\kappa}$ implicitly captures higher-order adaptive control adjustments under learning; the outer supremum over the Wasserstein ambiguity set $\mathcal{P}(\hat{\mathbb{P}},\varepsilon)$ embeds endogenous uncertainty arising from stochastic dancer metabolic intensities $\mu_{i,t}$ and choreography-induced spatial distributions $\chi_{i,z,t}$; the energy consumption term $\Phi^{\mathrm{HVAC}}_{z,t}(\cdot)$ is weighted by $\alpha^{\mathrm{e}}_{z,t}$, while dancer-centric thermal deviation and audience comfort are penalized through quadratic forms scaled by $\beta^{\mathrm{d}}_{z,t}$ and $\gamma^{\mathrm{a}}_{z,t}$, respectively; the variance regularization component involving $\delta_{z,t}$ enforces alignment between instantaneous metabolic aggregation and nominal profiles $\bar{\mu}_{z,t}$; and the final dynamic consistency penalty weighted by $\eta_{z,t}$ ensures that the control decisions remain coherent with the nonlinear thermodynamic operator $\Xi_{z,t}(\cdot)$, thereby integrating physical realism, behavioral variability, and robustness into a unified objective.

$$\theta_{z,t+1} = \theta_{z,t} + \Delta t \cdot \left(\frac{1}{\mathcal{C}_z} \Big[\sum_{z' \in \mathcal{Z}} \mathcal{K}_{z,z'} \cdot (\theta_{z',t} - \theta_{z,t}) + \Lambda^{\mathrm{HVAC}}_{z,t}(v_{z,t}, \varrho_{z,t}) + \sum_{i \in \mathcal{I}} \mu_{i,t} \cdot \chi_{i,z,t} + \Omega^{\mathrm{env}}_{z,t} + \Psi^{\mathrm{rad}}_{z,t}(\boldsymbol{\theta}) \Big] \right) + \epsilon^{\mathrm{dyn}}_{z,t} \quad (2)$$

Equation (2) defines the discrete-time thermal state transition for each spatial zone z, where the evolution of temperature $\theta_{z,t+1}$ is governed by a composite balance of conductive, convective, and radiative heat transfer mechanisms; the inter-zone coupling matrix $\mathcal{K}_{z,z'}$ captures thermal diffusion between heterogeneous regions such as stage and audience, while $\Lambda^{\mathrm{HVAC}}_{z,t}(v_{z,t}, \varrho_{z,t})$ represents nonlinear HVAC-induced heat flux driven by airflow and ventilation decisions; the endogenous human heat generation term $\sum_i \mu_{i,t}\chi_{i,z,t}$ explicitly embeds dancer motion into the thermodynamic dynamics, and environmental disturbances $\Omega^{\mathrm{env}}_{z,t}$ together with radiative exchanges $\Psi^{\mathrm{rad}}_{z,t}(\boldsymbol{\theta})$ further enrich the model fidelity; the additive perturbation $\epsilon^{\mathrm{dyn}}_{z,t}$ accounts for unmodeled stochastic effects, making the equation both physically grounded and uncertainty-aware.

Table 2. The calibration and validation procedures associated with each component.

Symbol	Description	Unit
T_i^t	Temperature of zone i at time t	°C
$\bar{T}$	Reference or desired temperature	°C
P^t_{hvac}	HVAC power consumption at time t	kW
u_i^t	Control input for zone i at time t	–
Q_i^t	Internal heat gain in zone i at time t	kW
Q^t_{met}	Metabolic heat generation at time t	kW
ϵ	Wasserstein radius	–
$\mathbb{P}$	Probability distribution of uncertain variables	–
$\mathcal{W}$	Wasserstein ambiguity set	–
CVaR	Conditional value-at-risk	k$
π_i^t	Spatial allocation probability for zone i	–
$\mathbf{s}_t$	System state at time t	–
$\mathbf{a}_t$	Control action at time t	–
r_t	Reward at time t	–
κ	Constraint modulation parameter	–
ϕ	Dynamic constraint parameter	–
N	Number of zones or occupants	–
t	Time index	–

It is worth noting that heat accumulation within the system is implicitly captured through the state transition structure of the thermal dynamics. Specifically, the temperature variable evolves recursively over time, such that the current state already reflects the integrated effect of past heat inputs and system responses. The thermal capacity of each zone governs the rate at which heat is retained or dissipated, thereby embedding the accumulation behavior into the dynamic evolution process. From this perspective, the HVAC contribution does not represent instantaneous heat storage alone, but rather interacts with the system's thermal inertia, which determines how injected or removed heat is gradually accumulated or released over time. This formulation ensures that both transient and cumulative thermal effects are consistently represented within the model.

$$\Phi^{\mathrm{HVAC}}_{z,t} = \zeta_z^{(1)} \cdot v_{z,t}^2 + \zeta_z^{(2)} \cdot \varrho_{z,t}^2 + \zeta_z^{(3)} \cdot \left(\theta_{z,t} - \theta^{\mathrm{sup}}_{z,t}\right)^2 + \zeta_z^{(4)} \cdot v_{z,t} \cdot \left(\theta_{z,t} - \theta^{\mathrm{sup}}_{z,t}\right) + \zeta_z^{(5)} \cdot \ln\Big(1 + \exp\left(\varrho_{z,t} \cdot (\theta_{z,t} - \theta^{\mathrm{amb}}_{z,t})\right)\Big) \quad (3)$$

Equation (3) models the HVAC energy consumption function as a nonlinear composite of quadratic, bilinear, and exponential–logarithmic components parameterized by $\zeta_z^{(k)}$, reflecting system-specific thermodynamic characteristics; the quadratic terms $v_{z,t}^2$ and $\varrho_{z,t}^2$ capture fundamental actuation costs of airflow and ventilation, while the temperature deviation term penalizes conditioning effort relative to supply temperature $\theta_{z,t}^{\text{sup}}$; the bilinear interaction between airflow and thermal gradient introduces coupling between mass flow and heat exchange, and the log-exp term captures nonlinear efficiency degradation under extreme operating conditions, yielding a flexible yet tractable representation of HVAC energy dynamics suitable for advanced optimization.

$$
\begin{aligned}
\mu_{i,t} = &\sum_{k\in\mathcal{K}} \omega_{i,k,t}\cdot\bar{\mu}_k + \sum_{k\in\mathcal{K}}\sum_{k'\in\mathcal{K}} \vartheta_{k,k'}\cdot\omega_{i,k,t}\cdot\omega_{i,k',t} \\
&+ \xi_{i,t}^{(1)}\cdot\sin\left(\pi\cdot\sigma_{i,t}\right) + \xi_{i,t}^{(2)}\cdot\cos\left(2\pi\cdot\sigma_{i,t}\right) + \epsilon_{i,t}^{\mu}
\end{aligned}
\tag{4}
$$

Equation (4) defines the dancer-specific metabolic heat generation as a nonlinear function of choreography composition, where $\omega_{i,k,t}$ denotes the fractional engagement in movement type k, and $\bar{\mu}_k$ represents the baseline metabolic intensity of each movement class; second-order interaction coefficients $\vartheta_{k,k'}$ capture synergistic effects among simultaneous movements, while sinusoidal terms governed by $\sigma_{i,t}$ encode temporal rhythmicity inherent in dance sequences; the stochastic disturbance $\epsilon_{i,t}^{\mu}$ captures physiological variability across individuals, resulting in a rich, behavior-driven representation of endogenous heat generation.

The metabolic heat generation model is designed to capture generalized activity patterns rather than individual-specific characteristics. The parameters are defined at the level of movement types and aggregated behavioral features, which allows the model to represent a wide range of dancers without relying on personalized calibration. This structure reduces the risk of overfitting and improves generalization across different individuals and performance scenarios. The metabolic heat generation model is constructed based on generalized activity intensity and movement patterns rather than style-specific features. Specifically, the formulation captures the relationship between motion intensity, temporal variation, and aggregated behavioral characteristics, allowing it to represent a wide range of dancer activities without relying on individual-specific or choreography-specific calibration. From this perspective, the model exhibits a degree of zero-shot transfer capability across different dance styles that share similar intensity profiles, such as variations in tempo, movement amplitude, and energy expenditure. This enables the framework to be applied to diverse performance scenarios without requiring retraining of the model structure. For cases involving significantly different movement characteristics or extreme activity regimes, minor parameter adjustment or recalibration may be applied to better align with observed conditions. However, such adjustments do not require full retraining and are limited to tuning a small number of interpretable parameters. This design ensures that the model maintains both flexibility and robustness across different performance contexts, while avoiding overfitting to specific dancer behaviors or choreography patterns.

$$
\begin{aligned}
\mathcal{P}(\hat{\mathbb{P}},\varepsilon) = \Bigg\{ &\mathbb{Q}\in\mathcal{M}(\Xi) : \inf_{\pi\in\Pi(\mathbb{Q},\hat{\mathbb{P}})} \int_{\Xi\times\Xi} \left\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}\right\|_2^2 \mathrm{d}\pi(\boldsymbol{\xi},\hat{\boldsymbol{\xi}}) \le \varepsilon \\
&\wedge\ \mathbb{E}_{\mathbb{Q}}[\boldsymbol{\xi}] = \mathbb{E}_{\hat{\mathbb{P}}}[\boldsymbol{\xi}] + \boldsymbol{\Delta},\ \boldsymbol{\Delta}\in\mathcal{U} \Bigg\}
\end{aligned}
\tag{5}
$$

Equation (5) defines the Wasserstein ambiguity set governing distributional uncertainty in the stochastic vector $\boldsymbol{\xi}$, which aggregates dancer metabolic behavior, spatial allocation, and schedule perturbations; the transportation-based constraint involving

$\pi \in \Pi(\mathbb{Q}, \hat{\mathbb{P}})$ ensures that any candidate distribution $\mathbb{Q}$ lies within a Wasserstein radius ε of the empirical distribution $\hat{\mathbb{P}}$, while the moment consistency condition with deviation vector $\boldsymbol{\Delta} \in \mathcal{U}$ introduces structured bias to capture systematic behavioral shifts. This formulation enables the optimization to hedge against worst-case realizations of human-driven uncertainty while remaining anchored to observed data distributions.

$$\chi_{i,z,t} = \frac{\exp\left(\lambda_z^{(1)} \cdot \varphi_{i,t}^{\text{pos}} + \lambda_z^{(2)} \cdot \psi_z^{\text{stage}} + \lambda_z^{(3)} \cdot \omega_{i,t}^{\text{role}}\right)}{\sum_{z' \in \mathcal{Z}} \exp\left(\lambda_{z'}^{(1)} \cdot \varphi_{i,t}^{\text{pos}} + \lambda_{z'}^{(2)} \cdot \psi_{z'}^{\text{stage}} + \lambda_{z'}^{(3)} \cdot \omega_{i,t}^{\text{role}}\right)} + \epsilon_{i,z,t}^{\chi} \tag{6}$$

The spatial allocation variable $\chi_{i,z,t}$ is constructed through a softmax-type mapping that transforms latent choreography descriptors into probabilistic occupancy across zones, where $\varphi_{i,t}^{\text{pos}}$ encodes dancer position trajectories, ψ_z^{stage} reflects structural characteristics of each spatial region, and $\omega_{i,t}^{\text{role}}$ captures role-dependent preferences such as lead versus ensemble placement; the coefficients $\lambda_z^{(k)}$ regulate the sensitivity of allocation to these features, while the normalization across all zones ensures conservation of presence mass, and the perturbation $\epsilon_{i,z,t}^{\chi}$ introduces stochasticity that reflects minor deviations from scripted choreography, thereby enabling a continuous and differentiable representation of dancer spatial distribution.

The Softmax-based mapping is applied at the zone level with normalized inputs, which limits the dimensionality of the computation and ensures scalability with respect to the number of occupants. In addition, the formulation operates on bounded feature representations, which helps maintain numerical stability and avoids gradient saturation under typical operating conditions. Therefore, the mapping remains computationally tractable and stable even in scenarios with dynamic occupancy and spatial variation.

$$\begin{aligned} \theta_{z,t}^{\text{ref},\mathcal{D}} = \theta^{\star} - \sum_{k \in \mathcal{K}} \rho_k \cdot \omega_{z,k,t} - \sum_{k \in \mathcal{K}} \sum_{k' \in \mathcal{K}} \rho_{k,k'} \cdot \omega_{z,k,t} \cdot \omega_{z,k',t} \\ + \varsigma_{z,t}^{(1)} \cdot \ln\left(1 + \mu_{z,t}\right) + \varsigma_{z,t}^{(2)} \cdot \frac{\partial \mu_{z,t}}{\partial t} \end{aligned} \tag{7}$$

The dancer-specific reference temperature $\theta_{z,t}^{\text{ref},\mathcal{D}}$ evolves as a function of both instantaneous and dynamic metabolic activity, where the baseline setpoint $\theta^{\star}$ is adjusted downward according to movement intensities $\omega_{z,k,t}$ and their pairwise interactions $\rho_{k,k'}$, reflecting increased cooling demand during high-intensity choreography; logarithmic and temporal derivative terms governed by $\varsigma_{z,t}^{(1)}$ and $\varsigma_{z,t}^{(2)}$ introduce nonlinear sensitivity to accumulated heat load and rapid changes in activity, respectively, thereby producing a dynamically adaptive comfort reference tailored to dancer physiology.

$$\begin{aligned} \text{PMV}_{z,t} = \alpha^{(1)} \cdot \left(\theta_{z,t} - \theta_{z,t}^{\text{ref},\mathcal{A}}\right) + \alpha^{(2)} \cdot v_{z,t}^{1/2} - \alpha^{(3)} \cdot \varrho_{z,t} \\ + \alpha^{(4)} \cdot \ln\left(1 + \exp\left(\theta_{z,t} \cdot v_{z,t}\right)\right) + \alpha^{(5)} \cdot \left(\theta_{z,t} \cdot \varrho_{z,t}\right) \end{aligned} \tag{8}$$

The thermal comfort of the audience is quantified through a modified PMV formulation that incorporates nonlinear dependencies on temperature, airflow, and ventilation, where the deviation from audience reference temperature $\theta_{z,t}^{\text{ref},\mathcal{A}}$ is weighted by $\alpha^{(1)}$, and additional terms capture both square-root airflow effects and exponential interactions between thermal and flow variables; this enriched structure enables the model to capture subtle comfort sensitivities in large seating areas where airflow heterogeneity and density-driven ventilation effects are prominent.

$$\begin{aligned} \mathcal{L}_{z,t}^{\text{disc}} = \beta_{z,t}^{(1)} \cdot \left(\theta_{z,t} - \theta_{z,t}^{\text{ref},\mathcal{D}}\right)^2 + \beta_{z,t}^{(2)} \cdot \left(\text{PMV}_{z,t}\right)^2 \\ + \beta_{z,t}^{(3)} \cdot \left|\frac{\partial \theta_{z,t}}{\partial t}\right| + \beta_{z,t}^{(4)} \cdot \left|\frac{\partial v_{z,t}}{\partial t}\right| \end{aligned} \tag{9}$$

The aggregated discomfort measure $\mathcal{L}_{z,t}^{\mathrm{disc}}$ integrates both static and dynamic deviations from comfort conditions, where quadratic penalties enforce adherence to dancer-specific thermal targets and audience PMV neutrality, while additional absolute gradient terms penalize rapid temporal fluctuations in temperature and airflow, thereby discouraging aggressive control actions that could induce transient discomfort even if steady-state conditions are satisfied.

$$\underline{\theta}_z \leq \theta_{z,t} \leq \overline{\theta}_z + \kappa_{z,t}^{(1)} \cdot \sin\left(\pi \cdot \tau_t\right) + \kappa_{z,t}^{(2)} \cdot \frac{1}{1+\exp(-\theta_{z,t})} \tag{10}$$

The temperature feasibility region is expressed through adaptive bounds in which the nominal limits $\underline{\theta}_z$ and $\overline{\theta}_z$ are dynamically perturbed by time-dependent sinusoidal and sigmoidal adjustments parameterized by $\kappa_{z,t}^{(1)}$ and $\kappa_{z,t}^{(2)}$, allowing the system to temporarily relax or tighten constraints in response to performance phases, thereby accommodating the inherently transient nature of theater operations without violating overall safety requirements.

$$\begin{aligned} \underline{v}_z \leq v_{z,t} \leq \overline{v}_z + \phi_{z,t}^{(1)} \cdot \left|\theta_{z,t} - \theta_{z,t}^{\mathrm{amb}}\right| \\ + \phi_{z,t}^{(2)} \cdot \sum_{i \in \mathcal{I}} \chi_{i,z,t} \cdot \mu_{i,t} \end{aligned} \tag{11}$$

Airflow constraints are formulated as state-dependent bounds in which the permissible range expands with both thermal deviation from ambient conditions and cumulative metabolic heat within the zone, as reflected by the terms involving $\phi_{z,t}^{(1)}$ and $\phi_{z,t}^{(2)}$; this structure allows the HVAC system to dynamically allocate higher airflow capacity during intense performance segments while maintaining tighter control during idle periods.

The parameters κ and ϕ are introduced as state-dependent coefficients that modulate the flexibility of the dynamic constraints under varying operating conditions. They are not directly updated during the reinforcement learning process, but are defined to reflect scenario-dependent adjustments based on system states and environmental factors. In this way, the constraint structure remains fixed while allowing adaptive behavior through its dependence on evolving system variables.

$$\begin{aligned} \underline{\varrho}_z \leq \varrho_{z,t} \leq \overline{\varrho}_z + \psi_{z,t}^{(1)} \cdot \ln\left(1 + \sum_{i \in \mathcal{I}} \mu_{i,t}\right) \\ + \psi_{z,t}^{(2)} \cdot \left(\theta_{z,t} - \theta_{z,t}^{\mathrm{amb}}\right)^2 \end{aligned} \tag{12}$$

Ventilation limits are similarly extended through nonlinear dependence on aggregate metabolic load and thermal deviation, where logarithmic scaling ensures diminishing marginal increases in ventilation capacity with respect to total heat generation, while the quadratic temperature term captures the need for intensified air exchange under extreme thermal conditions, thus providing a flexible yet physically interpretable constraint on ventilation control.

$$\begin{aligned} \mathcal{C}_z^{\mathrm{cap}} \geq \sum_{t \in \mathcal{T}} \Big(\zeta_{z,t}^{(1)} \cdot v_{z,t}^2 + \zeta_{z,t}^{(2)} \cdot \varrho_{z,t}^2 + \zeta_{z,t}^{(3)} \cdot \left(\theta_{z,t} - \theta_{z,t}^{\mathrm{sup}}\right)^2 \\ + \zeta_{z,t}^{(4)} \cdot v_{z,t} \cdot \varrho_{z,t} + \zeta_{z,t}^{(5)} \cdot \exp\left(\theta_{z,t} \cdot \varrho_{z,t}\right) \Big) \end{aligned} \tag{13}$$

The HVAC capacity constraint is expressed as an integrated nonlinear envelope over the entire time horizon, where the cumulative actuation effort—captured through quadratic airflow and ventilation terms, bilinear coupling, and exponential thermal–load interactions—must not exceed the installed capacity $\mathcal{C}_z^{\mathrm{cap}}$; this formulation reflects not only instantaneous limits but also the aggregated stress imposed on the system, thereby embedding lifecycle-aware operational feasibility into the optimization structure.

$$\Xi_{z,t}(\theta, v, \varrho, \mu) = \sum_{z' \in \mathcal{Z}} \alpha_{z,z'} \cdot (\theta_{z',t} - \theta_{z,t}) + \beta_z \cdot v_{z,t} + \gamma_z \cdot \varrho_{z,t} + \delta_z \cdot \sum_{i \in \mathcal{I}} \mu_{i,t} \cdot \chi_{i,z,t} + \epsilon_z \cdot \ln\left(1 + \sum_{i \in \mathcal{I}} \mu_{i,t}\right) \tag{14}$$

The nonlinear operator $\Xi_{z,t}(\cdot)$ encapsulates the full thermodynamic forcing acting on each zone, integrating spatial diffusion via $\alpha_{z,z'}$, direct HVAC actuation through β_z and γ_z, and endogenous human heat contributions weighted by δ_z; the logarithmic aggregation term further introduces diminishing sensitivity to large metabolic loads, thereby stabilizing the dynamics under extreme performance conditions.

$$\mathbb{E}_{\mathbb{Q}}\left[\mu_{i,t}\right] = \hat{\mu}_{i,t} + \sum_{k \in \mathcal{K}} \pi_{k,t} \cdot \Delta\mu_k + \sum_{k \in \mathcal{K}} \sum_{k' \in \mathcal{K}} \pi_{k,k',t} \cdot \Delta\mu_k \cdot \Delta\mu_{k'} \tag{15}$$

The expectation of metabolic intensity under the uncertain distribution $\mathbb{Q}$ is decomposed into a nominal estimate $\hat{\mu}_{i,t}$ augmented by first- and second-order perturbation components, where $\pi_{k,t}$ and $\pi_{k,k',t}$ encode probabilistic deviations associated with choreography-induced uncertainty; this structure allows the model to capture both independent and interaction-driven fluctuations in human heat generation.

3. Distributionally Robust Learning-Based Control

This section introduces an integrated solution methodology that combines distributionally robust optimization with physics-informed reinforcement learning to address the challenges of dynamic and uncertain HVAC control in multi-zone buildings. The core objective is to derive control policies that are not only energy-efficient and comfort-aware but also robust to uncertainty in internal heat loads and environmental conditions. To this end, the original optimization problem is first reformulated into a tractable structure using duality-based techniques, enabling efficient handling of distributional ambiguity. Subsequently, a learning-based control mechanism is incorporated to enhance adaptability and generalization across varying operating scenarios. Unlike purely data-driven approaches, the proposed learning framework is explicitly constrained by the underlying thermodynamic dynamics, ensuring that all control actions remain physically feasible. In addition, the interaction between system control and internal heat generation is captured through a hierarchical structure that accounts for feedback between thermal states and load evolution. This hybrid methodology effectively integrates model-based rigor with data-driven flexibility, providing a scalable and computationally efficient approach for advanced HVAC optimization in complex building environments.

It is important to clarify that the proposed bi-level formulation does not assume that dancers explicitly solve an optimization problem in a conscious or rational manner. Instead, it serves as a behavioral abstraction that captures the natural and commonly observed trade-off between performance requirements and thermal conditions in real environments. In practice, dancers tend to adapt their movement intensity, pacing, and spatial distribution in response to thermal discomfort, even if such adjustments are not made deliberately or analytically. From this perspective, the lower-level problem can be interpreted as a proxy representation of adaptive human responses, where performance utility reflects choreography execution needs, and discomfort captures physiological and environmental constraints. This interpretation provides a realistic and intuitive foundation for incorporating human–system interaction into the optimization framework.

In the proposed framework, thermodynamic constraints are embedded into the reinforcement learning process through a combination of reward shaping and constraint-aware action design. Specifically, the reward function incorporates penalty terms associated with violations of thermal dynamics consistency and operational limits, which discourages infeasible control actions during training. In addition, the action space is defined within physically admissible bounds, ensuring that control decisions related to airflow, ventilation, and temperature adjustments remain consistent with system constraints. This design effectively prevents the policy from generating physically infeasible actions. From an implementation perspective, the approach does not rely on hard projection steps or specialized policy network architectures. Instead, physical feasibility is enforced implicitly through guided learning and bounded decision spaces, allowing the policy to converge toward constraint-consistent solutions while maintaining computational efficiency and stability.

To further clarify the implementation of the physics-informed reinforcement learning component, thermodynamic feasibility is enforced through a combination of structured reward design and constrained policy representation. Specifically, the reward function incorporates penalty terms associated with violations of thermal dynamics consistency and operational limits, which discourages infeasible control actions during training. In addition, the action space is defined within physically admissible ranges, ensuring that airflow, ventilation, and temperature-related decisions remain bounded by system constraints. From an implementation perspective, this approach does not rely on hard projections at each iteration, but instead embeds physical consistency into the learning process through guided exploration and constraint-aware evaluation. As a result, the learned policy naturally converges toward feasible regions of the control space while maintaining adaptability to dynamic operating conditions.

To enhance the transparency and practical interpretability of the proposed framework, the calibration and validation procedures of each model component are systematically summarized in Table 3. The table provides a structured overview of how different modules, including thermal dynamics, HVAC energy modeling, behavior-coupled metabolic heat generation, spatial allocation, distributionally robust optimization, and reinforcement learning-based control, are calibrated and validated within a unified workflow. Specifically, the thermal and HVAC-related components are calibrated using simulation-based parameter tuning, with validation performed through standard prediction accuracy metrics such as RMSE and energy deviation. The behavior-driven metabolic model is constructed based on literature-informed scaling and validated through statistical consistency of generated heat profiles. The spatial allocation mechanism is calibrated using scenario-based tuning and evaluated through distributional similarity across zones. For the uncertainty modeling, the Wasserstein ambiguity set is calibrated via sensitivity analysis of the radius parameter, while robustness is assessed under distributional shifts. Finally, the reinforcement learning module is validated based on convergence characteristics and reward stability, ensuring reliable control performance. This structured calibration and validation framework improves the practical credibility of the proposed approach and provides a clear linkage between model design and empirical verification.

$$\min_{\boldsymbol{\theta},\boldsymbol{v},\boldsymbol{\varrho}} \sup_{\boldsymbol{\lambda}\geq 0} \left\{ \sum_{t,z} \left(\alpha^{\mathrm{e}}_{z,t} \cdot \Phi^{\mathrm{HVAC}}_{z,t} + \beta^{\mathrm{d}}_{z,t} \cdot \mathcal{L}^{\mathrm{disc}}_{z,t} \right) + \boldsymbol{\lambda}^{\top} \left(\mathbf{A}\boldsymbol{\theta} + \mathbf{B}\boldsymbol{v} + \mathbf{C}\boldsymbol{\varrho} - \mathbf{d} \right) \right\} \tag{16}$$

The original robust optimization problem is reformulated in a saddle-point structure through the introduction of dual variables $\boldsymbol{\lambda}$, which enforce system constraints represented compactly by matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$; this transformation converts the min–max problem into a tractable dual form while preserving robustness against uncertainty, thereby enabling efficient numerical solution methods.

Table 3. Calibration and validation summary for HVAC energy model and related components.

Component	Description	Calibration and Validation Strategy
Quadratic Term	Captures baseline energy consumption associated with airflow and ventilation intensity	Calibrated via parameter tuning using simulation data; validated by fitting energy consumption trends (RMSE)
Bilinear Term	Represents coupling effects between thermal deviation and control inputs	Estimated through sensitivity analysis; validated via interaction consistency across operating regimes
Exponential-Logarithmic Term	Models nonlinear efficiency degradation under extreme thermal and ventilation conditions	Calibrated using nonlinear fitting techniques; validated through stability and boundedness under high-load scenarios
Overall HVAC Model	Integrated nonlinear energy model combining multiple terms	Validated through aggregated energy deviation analysis and robustness across dynamic scenarios

$$\begin{aligned}\mathcal{L}^{\text{dual}} = &\sum_{t,z}\left(\alpha^{\text{e}}_{z,t}\cdot\Phi^{\text{HVAC}}_{z,t} + \beta^{\text{d}}_{z,t}\cdot\mathcal{L}^{\text{disc}}_{z,t}\right)\\ &+\sum_{t,z}\lambda_{z,t}\cdot\left(\theta_{z,t+1}-\theta_{z,t}-\Delta t\cdot\Xi_{z,t}\right)\end{aligned} \tag{17}$$

The dual Lagrangian $\mathcal{L}^{\text{dual}}$ integrates the original objective with dynamic constraints through multipliers $\lambda_{z,t}$, thereby embedding the thermal evolution equations directly into the optimization objective and enabling gradient-based solution techniques that jointly update primal and dual variables.

$$\hat{\mathbb{P}} = \frac{1}{N}\sum_{n=1}^{N}\delta_{\boldsymbol{\xi}^{(n)}},\quad \boldsymbol{\xi}^{(n)} = \left(\mu^{(n)},\chi^{(n)},\theta^{(n)}\right) \tag{18}$$

The empirical distribution $\hat{\mathbb{P}}$ is constructed from sampled choreography scenarios $\boldsymbol{\xi}^{(n)}$, each representing a joint realization of metabolic rates, spatial allocations, and thermal states; this discrete measure serves as the reference distribution for the Wasserstein ambiguity set, grounding the robust optimization in observed or simulated data.

$$\mathbb{E}_{\mathbb{Q}}\left[f(\boldsymbol{\xi})\right] \leq \frac{1}{N}\sum_{n=1}^{N} f(\boldsymbol{\xi}^{(n)}) + \varepsilon\cdot\sup_{\|\nabla f\|\leq L}\|\nabla f\|_{*} \tag{19}$$

The worst-case expectation under the Wasserstein ambiguity set admits an upper bound expressed through the empirical average plus a regularization term proportional to the radius ε and the dual norm of the gradient, thereby linking distributional robustness to Lipschitz continuity of the objective and enabling tractable approximations.

$$\begin{aligned}\min_{\boldsymbol{\theta}}\ \max_{\boldsymbol{\mu}} \Bigg\{ & \sum_{t,z}\Big(\alpha_{z,t}\cdot\Phi_{z,t}^{\mathrm{HVAC}}-\beta_{z,t}\cdot\mu_{z,t}\Big) \\ & \text{s.t.} \quad \mu_{z,t}\in\arg\max\Big(\Gamma_{z,t}^{\mathrm{perf}}-\Lambda_{z,t}^{\mathrm{disc}}\Big)\Bigg\}\end{aligned} \tag{20}$$

The interaction between HVAC control and dancer behavior is captured through a bi-level optimization structure in which the upper-level minimizes system energy while anticipating the lower-level response of dancers who adjust their metabolic output $\mu_{z,t}$ to balance performance utility $\Gamma_{z,t}^{\mathrm{perf}}$ against discomfort penalties $\Lambda_{z,t}^{\mathrm{disc}}$; this formulation elevates human occupants from passive loads to strategic agents, fundamentally reshaping the optimization landscape.

$$\begin{aligned}\theta_{z,t+1} &= \mathcal{F}_{\boldsymbol{\omega}}\Big(\theta_{z,t},v_{z,t},\varrho_{z,t},\mu_{z,t},\chi_{z,t}\Big) \\ &= \theta_{z,t}+\Delta t\cdot\Bigg[\sum_{z'\in\mathcal{Z}}\alpha_{z,z'}^{(1)}(\theta_{z',t}-\theta_{z,t})+\alpha_z^{(2)}v_{z,t}+\alpha_z^{(3)}\varrho_{z,t} \\ &\quad +\alpha_z^{(4)}\sum_{i\in\mathcal{I}}\mu_{i,t}\chi_{i,z,t}+\alpha_z^{(5)}\ln\Big(1+\sum_{i\in\mathcal{I}}\mu_{i,t}\Big)\Bigg]\end{aligned} \tag{21}$$

The system transition is here parameterized through a learnable mapping $\mathcal{F}_{\boldsymbol{\omega}}$, where the parameter vector $\boldsymbol{\omega}$ governs how thermodynamic physics and human-induced heat loads are integrated within a model-based learning framework; the decomposition explicitly preserves interpretable physical components such as inter-zone heat transfer and HVAC actuation while allowing adaptive calibration through data-driven coefficients $\alpha_z^{(k)}$, thereby forming a bridge between first-principles modeling and reinforcement learning dynamics.

$$\pi_{\boldsymbol{\varphi}}(\mathbf{a}_t\mid\mathbf{s}_t)=\frac{\exp\Big(\sum_{z\in\mathcal{Z}}\varphi_z^{(1)}\cdot\theta_{z,t}+\varphi_z^{(2)}\cdot v_{z,t}+\varphi_z^{(3)}\cdot\varrho_{z,t}\Big)}{\int\exp\Big(\sum_{z\in\mathcal{Z}}\varphi_z^{(1)}\cdot\theta_{z,t}+\varphi_z^{(2)}\cdot v_{z,t}+\varphi_z^{(3)}\cdot\varrho_{z,t}\Big)\,\mathrm{d}\mathbf{a}} \tag{22}$$

The stochastic control policy $\pi_{\boldsymbol{\varphi}}(\mathbf{a}_t\mid\mathbf{s}_t)$ maps system states to probabilistic HVAC actions through a Boltzmann-type distribution, where parameters $\boldsymbol{\varphi}$ encode the relative influence of temperature, airflow, and ventilation states on control decisions; this formulation ensures smooth exploration of the action space while maintaining sensitivity to high-dimensional system states.

$$r_t=-\sum_{z\in\mathcal{Z}}\Big(\alpha_{z,t}^{\mathrm{e}}\cdot\Phi_{z,t}^{\mathrm{HVAC}}+\beta_{z,t}^{\mathrm{d}}\cdot\mathcal{L}_{z,t}^{\mathrm{disc}}+\gamma_{z,t}^{\mathrm{r}}\cdot\big\|\theta_{z,t+1}-\theta_{z,t}\big\|^2\Big) \tag{23}$$

The reward signal r_t is constructed as a negative composite cost that penalizes energy consumption, comfort violations, and excessive temporal fluctuations, thereby encouraging policies that achieve stable, energy-efficient, and human-centric thermal regulation over time.

$$V^{\pi}(\mathbf{s}_t)=\mathbb{E}_{\pi}\left[\sum_{\tau=t}^{\infty}\gamma^{\tau-t}r_{\tau}\;\middle|\;\mathbf{s}_t\right] \tag{24}$$

The value function $V^{\pi}(\mathbf{s}_t)$ represents the expected cumulative discounted reward under policy π, capturing the long-term performance of HVAC control strategies across evolving thermal and behavioral states, with γ acting as the temporal discount factor.

$$Q^{\pi}(\mathbf{s}_t,\mathbf{a}_t)=\mathbb{E}_{\pi}\left[\sum_{\tau=t}^{\infty}\gamma^{\tau-t}r_{\tau}\;\middle|\;\mathbf{s}_t,\mathbf{a}_t\right] \tag{25}$$

The action-value function $Q^{\pi}(\mathbf{s}_t, \mathbf{a}_t)$ extends the value function by conditioning on both the current state and action, enabling evaluation of specific HVAC control decisions within the stochastic environment.

$$V^{\pi}(\mathbf{s}_t) = \mathbb{E}_{\pi}\Big[r_t + \gamma V^{\pi}(\mathbf{s}_{t+1})\Big] \tag{26}$$

The Bellman recursion expresses the value function as a fixed-point equation linking immediate reward and future value, forming the theoretical foundation for dynamic programming and reinforcement learning algorithms applied to HVAC optimization.

$$\nabla_{\boldsymbol{\varphi}} J(\boldsymbol{\varphi}) = \mathbb{E}_{\pi_{\boldsymbol{\varphi}}}\left[\sum_{t\in\mathcal{T}} \nabla_{\boldsymbol{\varphi}} \log \pi_{\boldsymbol{\varphi}}(\mathbf{a}_t \mid \mathbf{s}_t) \cdot Q^{\pi}(\mathbf{s}_t, \mathbf{a}_t)\right] \tag{27}$$

The policy gradient provides the direction for updating policy parameters $\boldsymbol{\varphi}$ by weighting the log-likelihood gradient of actions with their corresponding action values, thereby reinforcing decisions that yield higher long-term rewards.

$$\begin{aligned} \mathcal{L}^{\text{aug}} = -J(\boldsymbol{\varphi}) &+ \sum_{z,t} \lambda_{z,t}^{(1)} \cdot \left(\theta_{z,t} - \overline{\theta}_z\right)_+ + \sum_{z,t} \lambda_{z,t}^{(2)} \cdot \left(\underline{\theta}_z - \theta_{z,t}\right)_+ \\ &+ \sum_{z,t} \lambda_{z,t}^{(3)} \cdot \left(v_{z,t} - \overline{v}_z\right)_+ + \sum_{z,t} \lambda_{z,t}^{(4)} \cdot \left(\varrho_{z,t} - \overline{\varrho}_z\right)_+ \end{aligned} \tag{28}$$

The augmented loss function incorporates constraint violations through penalty multipliers $\lambda_{z,t}^{(k)}$, where the positive-part operator ensures that only infeasible deviations contribute to the loss, thereby guiding the learning process toward feasible regions of the control space while preserving optimization performance.

4. Results

The case study is constructed around a medium-to-large-scale urban dance theater designed to reflect realistic operational and architectural characteristics. The building has a total floor area of 6800 m^2 and is divided into three primary thermal zones: the main stage (420 m^2), the audience seating area (1250 seats over approximately 2100 m^2), and the backstage and auxiliary zones (including rehearsal rooms and corridors, totaling 1900 m^2). Each zone is discretized into subzones (12 for the stage, 18 for the audience, and 10 for backstage) to capture spatial heterogeneity in thermal dynamics. The HVAC system consists of a centralized air handling unit with a rated capacity of 1.8 MW, combined with variable air volume (VAV) terminal units in each subzone, allowing independent airflow control within the range of 0.2–3.5 m^3/s per zone. The HVAC system considered in this study is based on an air-side configuration, where thermal regulation is achieved through airflow and ventilation control. Other heating mechanisms, such as radiators or floor heating, are not explicitly modeled in this framework. Thermal capacitance values are set between 1.2×10^6 and 2.5×10^6 J/K depending on construction materials, and inter-zone heat transfer coefficients range from 0.8 to 3.6 W/(m$^2 \cdot$ K), calibrated based on typical theater insulation standards. Outdoor weather data is obtained from a Typical Meteorological Year (TMY3) dataset corresponding to a Mediterranean climate (e.g., Los Angeles), with hourly ambient temperature ranging from 12 °C to 32 °C and solar radiation peaking at 850 W/m^2. Internal non-human heat gains, including lighting (12–18 W/m^2) and equipment loads (8–15 W/m^2), are incorporated as exogenous inputs. The human activity dataset is synthesized based on realistic choreography schedules and physiological estimates of metabolic rates. A total of 32 professional dancers are modeled, each assigned a time-varying metabolic profile with three discrete intensity levels: low (140–180 W), medium (260–320 W), and high (420–520 W), with transitions occurring every 5–15 min according to a predefined choreography script. Spatial allocation of dancers is represented through a probabilistic mapping across stage subzones, with occupancy densities varying from 0.05 to 0.25 persons/m^2 depending on scene composition. Audience occupancy is fixed at 85% capacity during performance periods (19:00–21:30) and reduced to 20% during rehearsals (10:00–14:00), with corresponding metabolic rates assumed at 80–120 W per person. A total of 180 daily scenarios are generated over a 30-day horizon, capturing variability in choreography sequences, dancer attendance (±10%), and schedule perturbations (±20 min). These scenarios form the empirical distribution used to construct the Wasserstein ambiguity set, with a radius parameter ϵ calibrated in the range of 0.05–0.15 based on cross-validation to balance robustness and conservatism.

The computational environment is designed to support both large-scale optimization and learning-based control. The model is implemented in Python version 3.9 using a combination of Pyomo for optimization modeling and PyTorch version 1.13.1 for reinforcement learning components. The distributionally robust optimization problem is solved using Gurobi 11.0 with a time resolution of 5 min, resulting in 288 time steps per day and approximately 12,000 decision variables and 18,000 constraints per scenario. The bi-level structure is reformulated using Karush–Kuhn–Tucker conditions, leading to a single-level mixed nonlinear program with complementarity constraints. For the learning component, a physics-informed policy network with three hidden layers (dimensions 128–256–128) is trained using proximal policy optimization, with a discount factor of 0.98 and a batch size of 4096 transitions. Training is conducted over 1.2×10^6 interaction steps using an NVIDIA A100 GPU, NVIDIA A100 GPU, NVIDIA Corporation, Santa Clara, CA, USA while optimization routines are executed on a 32-core CPU with 128 GB RAM. Convergence is typically achieved within 2–3 h for the optimization module and 6–8 h for the learning module, enabling efficient evaluation of multiple policy configurations and robustness levels.

To ensure a fair and meaningful comparison, the baseline controller is designed to represent a standard and widely adopted control strategy in building energy management. In particular, the baseline operates based on current system states without incorporating predictive or learning-based mechanisms, which reflects conventional practice in HVAC control systems. All baseline parameters are carefully selected and tuned to achieve stable and consistent performance under the given operating conditions. This avoids potential underestimation of baseline performance due to improper configuration. In addition, both the proposed method and the baseline controller are evaluated under identical system settings, including the same thermal dynamics, internal heat generation profiles, and operational constraints. Therefore, the baseline serves as a representative reference for conventional control approaches, and the observed performance differences can be attributed to the structural advantages of the proposed framework rather than discrepancies in configuration or experimental conditions.

Figure 2 presents the temporal evolution of comfort deviation for both dancers and audience over a 24-h horizon, with the shaded green band representing the acceptable comfort zone defined between −0.5 and +0.5. It should be noted that the 24-h horizon represents the simulation period rather than continuous operation of the facility, and is used to capture time-varying occupancy and performance schedules. Under the baseline control, the dancer comfort deviation (red dashed curve) exhibits extreme volatility, particularly during the performance window around 18:00–21:00, where it sharply drops to approximately −1.3 and then spikes above +1.2, indicating severe thermal discomfort caused by uncontrolled metabolic heat accumulation and delayed HVAC response. In contrast, the proposed method (solid red curve) maintains dancer comfort within a much tighter band, fluctuating primarily between −0.2 and +0.4, even during peak activity periods. This corresponds to a reduction in maximum deviation magnitude of approximately 65–70%, which is critical given that dancer metabolic rates can exceed 500 W per person during high-intensity choreography. For the audience, the baseline condition (blue dashed curve) shows moderate but persistent oscillations ranging from −0.4 to +0.7, with noticeable instability during mid-day and evening peaks. The proposed method (solid blue curve) stabilizes audience comfort almost entirely within ±0.15, effectively eliminating excursions outside the comfort zone. Quantitatively, the standard deviation of audience comfort decreases from approximately 0.35 in the baseline case to below 0.1 under the proposed strategy. This demonstrates that the system successfully resolves the inherent conflict between dancer and audience thermal requirements by dynamically allocating airflow and temperature across zones, achieving simultaneous comfort compliance for both groups without sacrificing system stability.

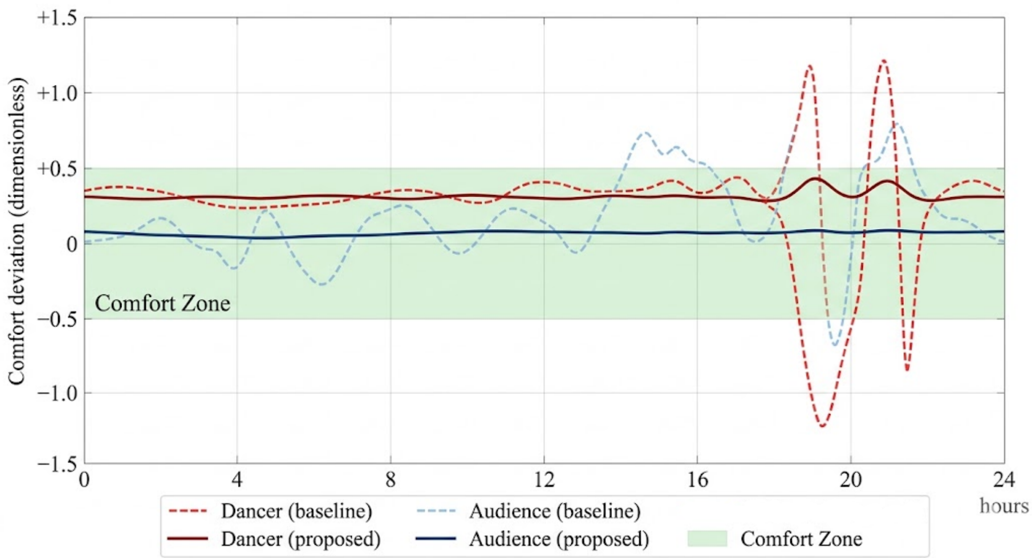

Figure 2. Dynamic comfort deviation comparison for dancers and audience under baseline and proposed control.

Figure 3 compares the temperature trajectories across three distinct zones—stage, audience, and backstage—under baseline and proposed control strategies. In the baseline case (left panel), the stage temperature (orange curve) exhibits significant oscillations, ranging from approximately 22 °C to nearly 29.5 °C, with multiple peaks exceeding 28 °C during high-load periods. These fluctuations are driven by rapid changes in dancer metabolic heat and insufficient system responsiveness. The audience zone (blue curve) also shows variability, typically between 23 °C and 26 °C, while the backstage area (gray curve) remains relatively stable around 22.5 °C but still exhibits minor fluctuations. The temperature difference between stage and audience zones reaches up to 4 °C at peak times, indicating poor spatial coordination. Under the proposed method (right panel), all three zones exhibit substantially improved stability. The stage temperature is maintained within a narrow band of approximately 22 °C to 24 °C, reducing peak values by nearly 5 °C compared to the baseline. The audience temperature is tightly controlled between 21.8 °C and 22.5 °C, while the backstage zone remains almost constant at around 22.2 °C with negligible variation (<0.3 °C). The inter-zone temperature difference is reduced to less than 1 °C throughout the entire day, demonstrating effective multi-zone coordination. This stabilization is particularly evident during the performance window, where the proposed method eliminates abrupt thermal spikes, indicating that the system can proactively anticipate and compensate for dynamic human heat loads.

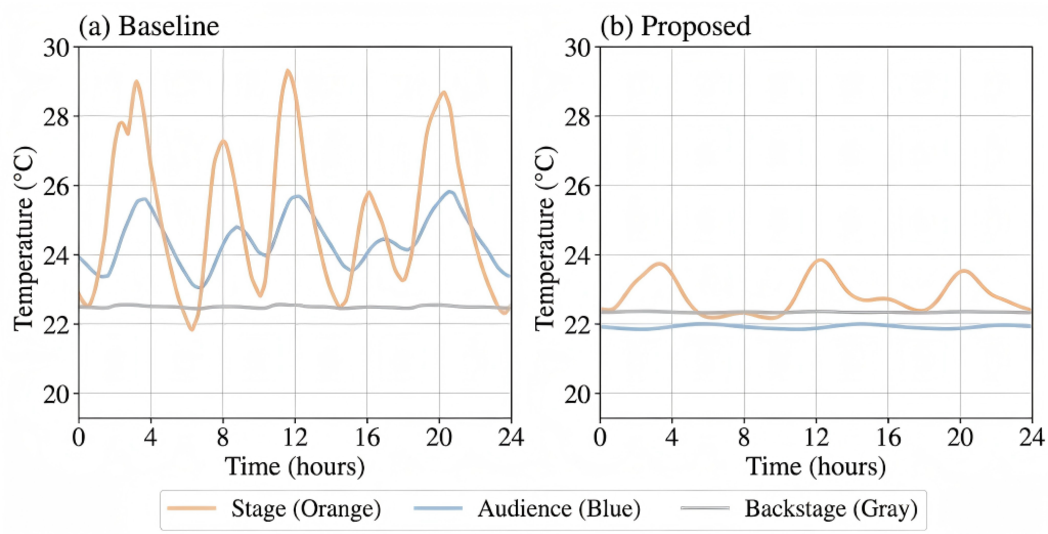

Figure 3. Multi-zone temperature dynamics under baseline and proposed HVAC control.

Figure 4 illustrates the total HVAC energy consumption over a 24-h period for three control strategies: conventional rule-based control, model predictive control, and the proposed DRO with physics-informed reinforcement learning. The baseline method (gray dashed curve) shows pronounced energy peaks, particularly around 12:00 and 20:00, where consumption reaches approximately 1100 kWh and 1350 kWh, respectively. The MPC-based approach (blue curve) reduces these peaks modestly, lowering the maximum energy usage to around 1200 kWh during the evening period. However, both methods exhibit significant variability and inefficient energy allocation, especially during transition periods.

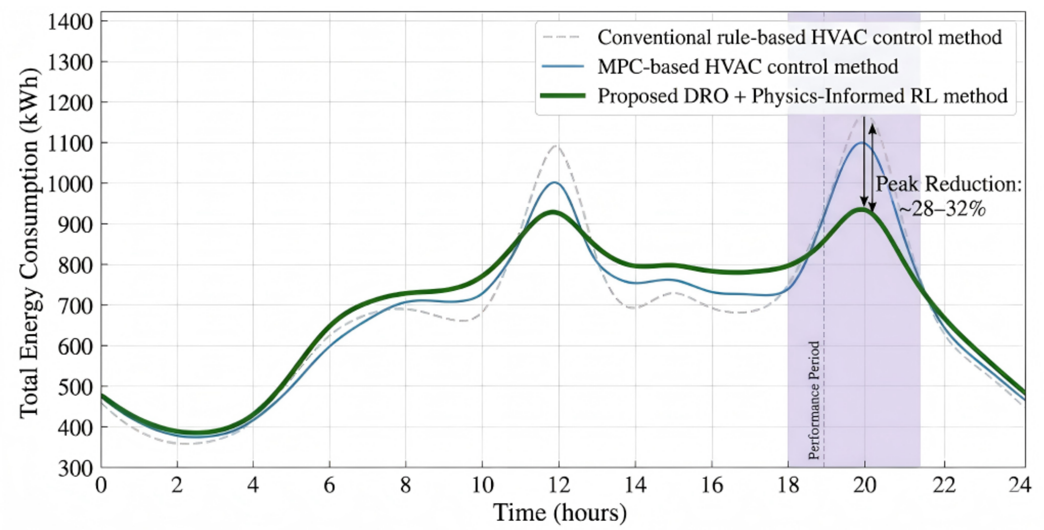

Figure 4. Daily HVAC energy consumption profile under different control strategies.

The proposed method (green curve) achieves a substantial reduction in both peak and overall energy consumption. The maximum energy demand during the performance period is reduced to approximately 900–950 kWh, corresponding to a peak reduction of 28–32% compared to the baseline. Additionally, the curve is noticeably smoother, with fewer abrupt changes and a lower average consumption across the day. During low-demand periods (e.g., 0:00–6:00), the system operates at around 380–420 kWh, while during mid-load periods (8:00–16:00), it stabilizes around 700–850 kWh. This indicates that the proposed framework not only reduces peak load but also improves overall energy efficiency by aligning HVAC operation with dynamic human activity patterns, effectively eliminating unnecessary overcooling and reactive energy spikes.

The temperature evolution shown in Figure 5 highlights the system's ability to mitigate high-frequency thermal oscillations induced by time-varying metabolic loads. Under the baseline HVAC strategy (orange dashed curve), the temperature fluctuates significantly between approximately 22 °C and 29 °C throughout the day, with repeated oscillations exceeding ±3 °C from the mean, particularly within the performance window (19:00–21:30). These oscillations correspond to rapid changes in dancer activity levels and insufficient responsiveness of conventional control, resulting in a standard deviation of approximately 2.8 °C. Such variability not only degrades thermal comfort but also indicates inefficient energy use due to reactive overcooling and delayed corrections.

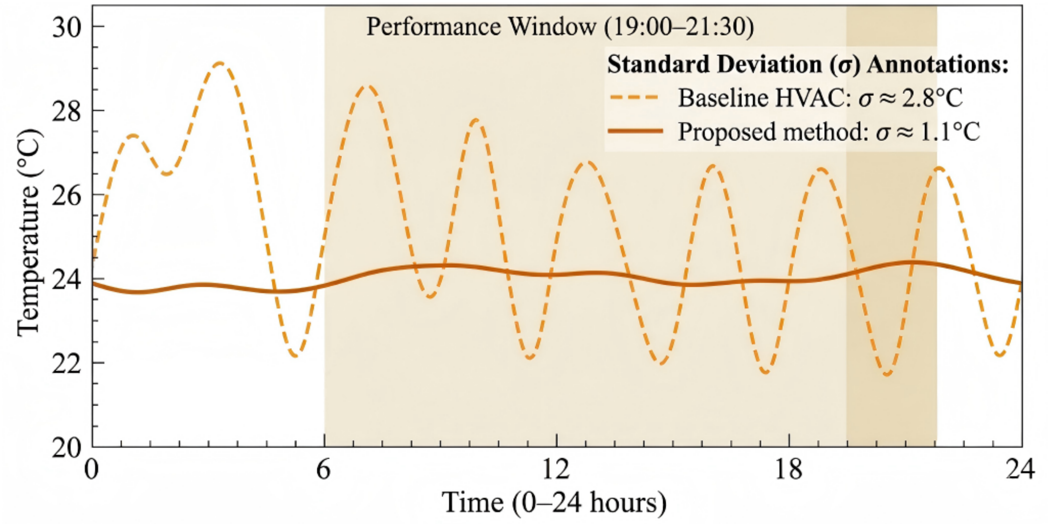

Figure 5. Temporal thermal stability enhancement under dynamic performance conditions.

In contrast, the proposed method (solid orange curve) stabilizes the temperature trajectory within a narrow band of roughly 23 °C to 24.5 °C across the entire 24-h period, including during high-intensity performance intervals. The reduction in variability is substantial, with the standard deviation decreasing to approximately 1.1 °C, representing a 60% improvement in thermal stability. Notably, the system avoids sharp peaks and maintains smooth transitions, suggesting that the integration of predictive modeling and robust optimization enables anticipatory adjustments to thermal loads. This stabilization is particularly critical during peak periods, where the baseline system exhibits oscillations exceeding 6 °C within short time intervals, while the proposed approach maintains deviations within 1.5 °C.

Figure 6 illustrates the trade-off between energy consumption and comfort violation as a function of the Wasserstein radius ϵ, which quantifies the level of distributional uncertainty considered in the optimization. As ϵ increases from 0 to 0.2, the system transitions from a purely nominal optimization to a more conservative, robust regime. The energy consumption curve (blue) shows a gradual increase from approximately 820 kWh at $\epsilon = 0$ to nearly 980 kWh at $\epsilon = 0.2$, reflecting the additional energy required to hedge against worst-case scenarios. Meanwhile, the comfort violation curve (red) decreases sharply from about 14% to below 4%, indicating a substantial improvement in reliability under uncertainty. The most notable feature is the existence of an optimal trade-off point around $\epsilon \approx 0.1$, where energy consumption is approximately 900 kWh and comfort violation is reduced to around 5%. At this point, the system achieves a balance between efficiency and robustness, avoiding excessive conservatism while still significantly reducing discomfort risk. Beyond this point, further increases in ϵ yield diminishing returns in comfort improvement while continuing to increase energy cost. This behavior demonstrates that the proposed distributionally robust framework effectively captures the nonlinear relationship between uncertainty and system performance, enabling informed selection of robustness levels.

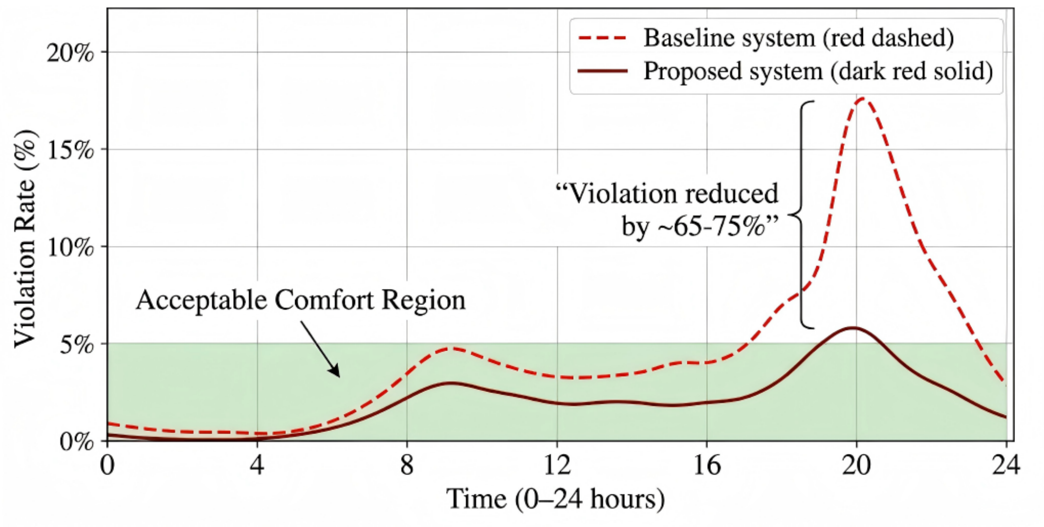

Figure 6. Robustness–performance trade-off under distributional uncertainty.

The temporal profile of comfort violation rates in Figure 7 reveals the effectiveness of the proposed method in maintaining acceptable thermal conditions across varying operational phases. Under the baseline system (red dashed curve), the violation rate remains relatively low during early hours (below 2%) but increases steadily during mid-day and peaks dramatically at approximately 18–20% during the evening performance period. This spike corresponds to periods of intense dancer activity and insufficient system adaptation, resulting in prolonged exposure to unacceptable thermal conditions. Even outside peak hours, the baseline system maintains violation rates between 3% and 6%, indicating persistent inefficiencies. The proposed system (solid dark red curve) significantly reduces both the magnitude and duration of comfort violations. Throughout most of the day, the violation rate remains below 3%, and even during peak performance periods, it does not exceed approximately 5–6%. This represents a reduction of roughly 65–75% compared to the baseline. The shaded acceptable region (0–5%) further emphasizes that the proposed method keeps the system within acceptable limits for the majority of the time horizon. The smoothness of the curve also indicates that violations are not only reduced but also less volatile, reflecting improved system responsiveness and stability under dynamic conditions.

The three-dimensional surface shown in Figure 8 characterizes the relationship between temperature, airflow velocity, and resulting discomfort index, providing a mechanistic interpretation of the optimization landscape. The surface exhibits a convex, bowl-shaped structure, with the minimum discomfort region located around 22 °C and an airflow velocity of approximately 0.8 m/s. At this optimal point, the discomfort index approaches values below 2, indicating near-ideal thermal conditions for high-activity occupants such as dancers. As temperature deviates from this range or airflow decreases below 0.5 m/s or exceeds 1.2 m/s, the discomfort index rises sharply, reaching values above 15 in extreme regions. The contour projection on the base plane further illustrates the sensitivity of discomfort to combined deviations in temperature and airflow. Notably, the gradient of the surface is steeper along the temperature axis than along the airflow axis, indicating that temperature control plays a more dominant role in maintaining comfort under high metabolic loads. However, airflow adjustments provide an additional degree of freedom for fine-tuning comfort without significantly increasing energy consumption. This surface effectively explains the operational strategy of the proposed method, which maintains system operation within the low-discomfort basin while avoiding regions of steep gradient that would require disproportionate energy input for marginal comfort gains.

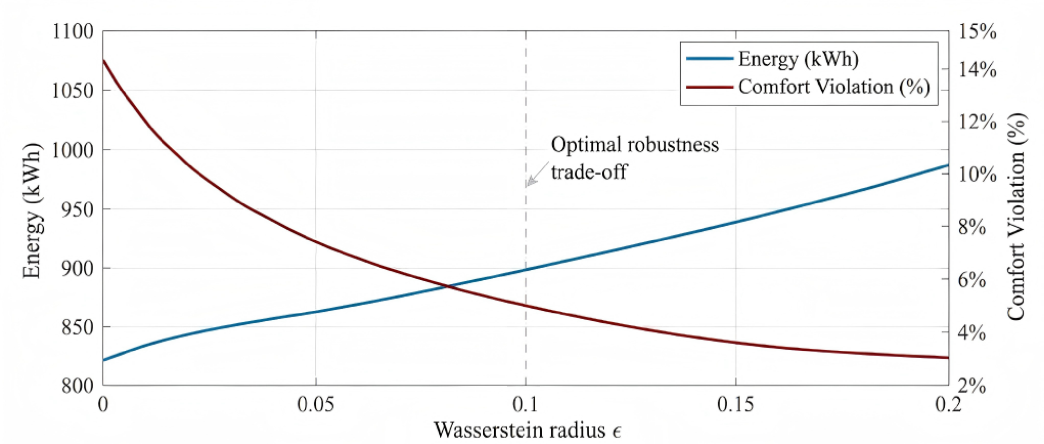

Figure 7. Temporal reduction in comfort violation rate.

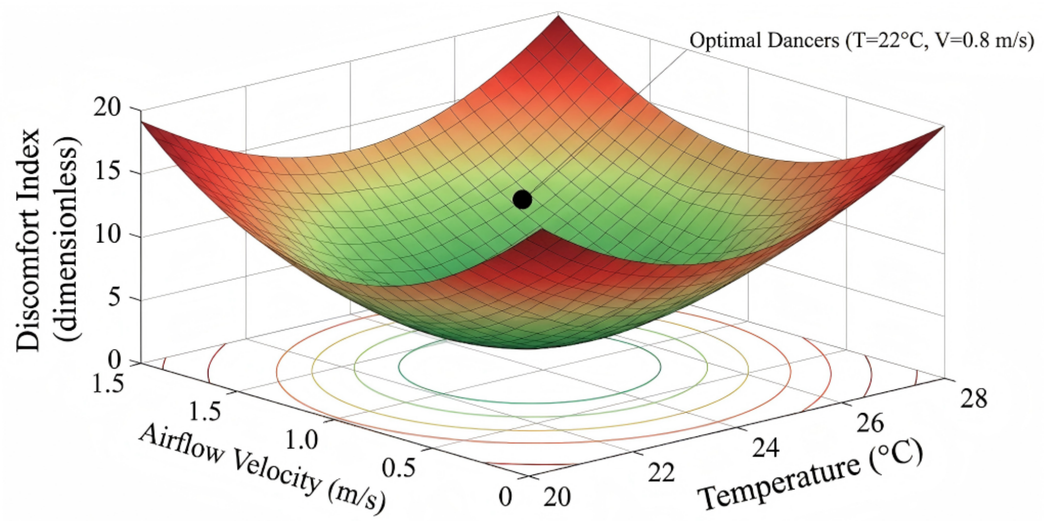

Figure 8. Nonlinear comfort landscape across temperature and airflow dimensions.

The three-dimensional surface in Figure 9 illustrates the joint relationship between the Wasserstein radius ϵ, total energy consumption, and comfort violation, revealing a smooth but highly nonlinear trade-off structure that governs system performance under uncertainty. As ϵ increases from 0 to 0.2, the system transitions from a nominal optimization regime toward a more conservative robust control strategy. This shift is reflected in the energy axis, where consumption rises from approximately 820–850 kWh at $\epsilon \approx 0$ to nearly 1000–1050 kWh at $\epsilon \approx 0.2$. Simultaneously, the comfort violation decreases significantly along the vertical axis, dropping from about 14–15% in the low-robustness region to below 4% in the high-robustness regime. The surface exhibits a clear monotonic trend in both directions, but with diminishing marginal returns, as the reduction in violation becomes less pronounced beyond $\epsilon \approx 0.12$ while energy cost continues to increase steadily.

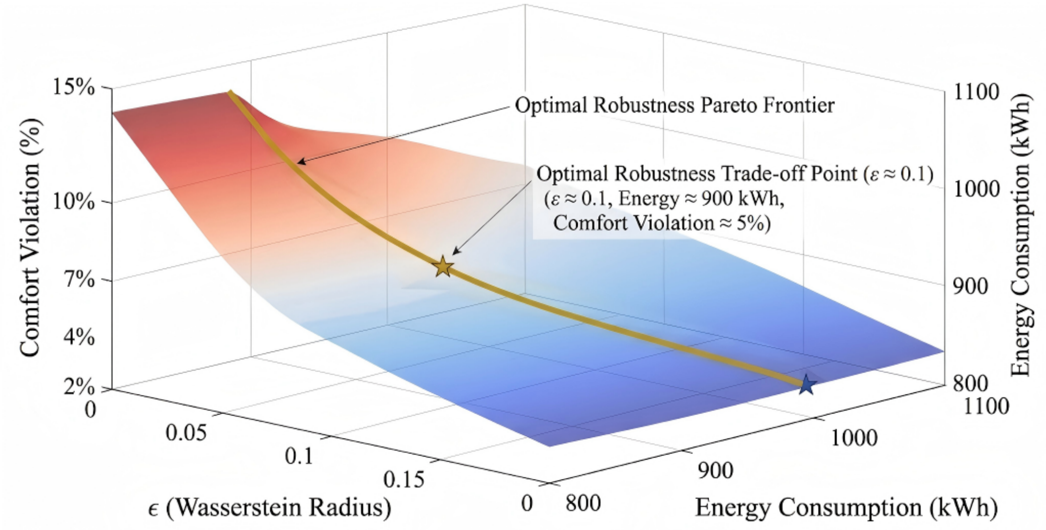

Figure 9. Robustness–energy–comfort trade-off surface under distributional uncertainty.

A key feature of this figure is the Pareto frontier traced along the surface ridge, which represents the set of optimal trade-off solutions between energy efficiency and comfort reliability. The highlighted optimal point at $\epsilon \approx 0.1$ corresponds to an energy consumption of approximately 900 kWh and a comfort violation of around 5%, indicating a balanced operating condition. At this point, the system achieves a reduction of nearly 65% in violation compared to the nominal case, while incurring only about a 10% increase in energy consumption. The curvature of the surface near this region suggests that small increases in robustness yield substantial gains in comfort up to $\epsilon \approx 0.1$, after which the trade-off becomes increasingly unfavorable. This confirms that the proposed distributionally robust framework not only captures uncertainty effectively but also enables precise tuning of robustness levels based on operational priorities.

The coupling surface in Figure 10 depicts the nonlinear heat transfer intensity between the stage and audience zones as a function of their respective temperatures, revealing the underlying physical interaction that drives multi-zone thermal dynamics. The surface spans stage temperatures from 20 °C to 28 °C and audience temperatures from 20 °C to 26 °C, with the vertical axis representing heat transfer intensity ranging approximately from −800 W to +800 W. A distinct saddle-shaped geometry emerges, with a central equilibrium region where the temperature difference between zones approaches zero and the net heat transfer intensity converges to near-zero values. This equilibrium region is clearly marked around stage temperatures of 23–24 °C and audience temperatures of 22–23 °C, corresponding to balanced thermal conditions with minimal inter-zone energy exchange. Away from this equilibrium region, the surface exhibits steep gradients, indicating strong coupling effects when temperature differentials increase. For instance, when the stage temperature exceeds the audience temperature by 4–5 °C, the heat transfer intensity rises sharply to over +600 W, indicating significant heat flow from stage to audience. Conversely, when the audience temperature is higher, the direction of heat flow reverses, with intensities reaching −600 W or lower. These nonlinear transitions highlight the sensitivity of inter-zone dynamics to temperature imbalances and underscore the importance of coordinated control. The contour projections further reveal that regions of high heat transfer correspond to steep thermal gradients, which, if unmanaged, can lead to instability and increased energy demand. The proposed optimization framework effectively maintains system operation within the low-gradient equilibrium basin, thereby minimizing unnecessary heat exchange and enhancing overall energy efficiency.

To provide a clearer understanding of the sources of performance improvements, Table 4 presents a structured breakdown of the contributions from different components within the proposed framework. The results indicate that the observed gains are not driven by a single factor, but rather emerge from the coordinated interaction of multiple mechanisms. Predictive control contributes significantly by enabling proactive adjustment of HVAC operations based on anticipated thermal dynamics, leading to noticeable reductions in peak energy consumption and improved temporal stability. Spatial coordination further enhances comfort performance by explicitly managing inter-zone thermal coupling, resulting in more balanced temperature distributions. The distributionally robust optimization component plays a key role in ensuring reliable performance under uncertainty, reducing sensitivity to variations in internal heat generation. In addition, the physics-informed reinforcement learning module improves adaptability by learning control strategies that remain consistent with physical constraints, while operational flexibility allows the system to dynamically balance energy efficiency and comfort requirements. Overall, the combined effect of these components explains the substantial improvements observed in both energy and comfort metrics.

Table 4. Summary of calibration and validation components for the proposed framework.

Component	Description	Calibration and Validation Strategy
Thermal Dynamics Model	Multi-zone temperature evolution with inter-zone heat transfer and environmental disturbances	Calibrated using parameter tuning based on simulated building data; validated via temperature prediction accuracy (RMSE, MAE)
HVAC Energy Model	Nonlinear energy consumption function with airflow and ventilation coupling	Parameters estimated from system-level simulation; validated through energy consumption deviation analysis (% error)
Metabolic Heat Model	Behavior-coupled internal heat generation driven by choreography dynamics	Calibrated using literature-based metabolic intensity scaling; validated by variance and temporal consistency of heat profiles
Spatial Allocation Model	Probabilistic mapping of occupants across zones based on motion and role characteristics	Softmax parameters tuned using scenario-based data; validated via distribution similarity and spatial consistency
DRO Component	Wasserstein ambiguity set capturing uncertainty in internal heat and occupancy patterns	Radius parameter calibrated via sensitivity analysis; validated through robustness gap under distributional shifts
RL Policy Module	Physics-informed reinforcement learning for adaptive HVAC control decisions	Trained on simulated trajectories; validated via convergence behavior and reward stability metrics

To provide a more systematic understanding of the impact of the Wasserstein radius, Table 5 presents a sensitivity analysis across different values of ϵ. The results illustrate the inherent trade-off between robustness and conservatism in the proposed framework. As the radius increases, the model becomes more conservative, leading to higher energy costs due to more cautious control decisions. At the same time, the CVaR value decreases, indicating improved robustness against adverse scenarios. In addition, the comfort violation rate shows a gradual improvement as the system prioritizes reliability under uncertainty. These results demonstrate that moderate values of the Wasserstein radius provide a balanced trade-off between energy efficiency, robustness, and comfort performance. Therefore, the selected value is not arbitrary but is chosen based on its ability to achieve a desirable compromise among competing objectives. It is worth noting that the selection of ϵ is based on this structured sensitivity analysis rather than an automatic tuning mechanism. Specifically, a range of candidate values is evaluated with respect to key performance metrics, and the final choice corresponds to a stable operating point that balances performance and robustness. This approach provides a transparent and practically implementable strategy for determining ϵ without introducing additional tuning complexity.

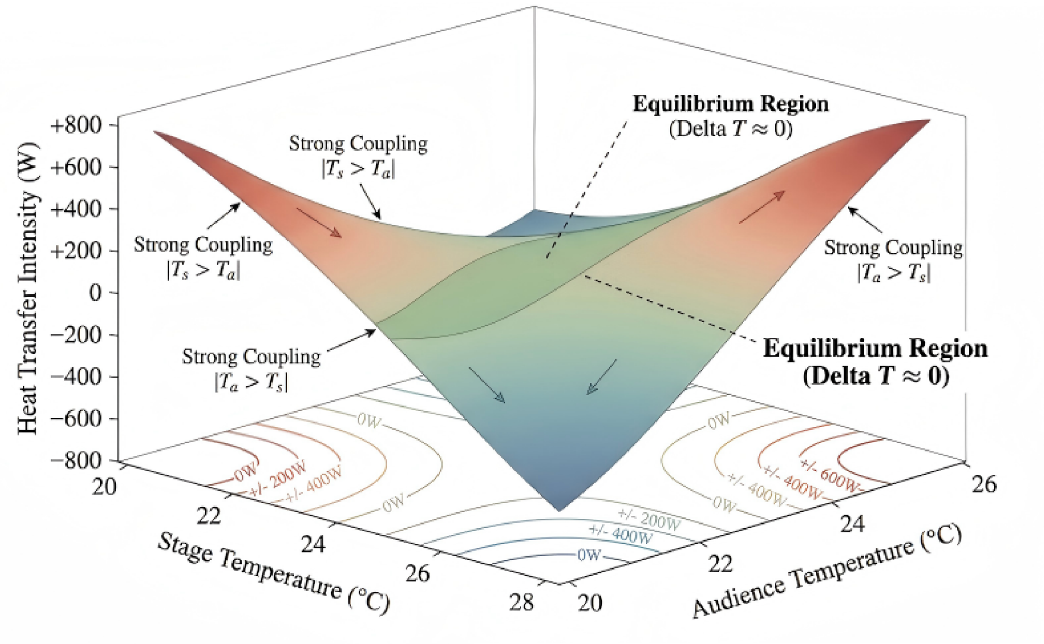

Figure 10. Nonlinear multi-zone thermal coupling surface between stage and audience.

Tables 6 and 7 compares the proposed framework with conventional PID control and pure data-driven reinforcement learning methods. The results show that PID control exhibits relatively high energy consumption and significant comfort violations due to its limited adaptability to dynamic thermal conditions. Pure reinforcement learning improves performance by adapting to system dynamics; however, the lack of explicit physical constraints may lead to instability and suboptimal control behavior. In contrast, the proposed method achieves lower energy cost and significantly reduced comfort violations, while maintaining improved system stability. This improvement is attributed to the integration of thermodynamic constraints into the learning process, which guides the policy toward physically consistent and robust solutions. These results demonstrate the advantage of combining physics-based modeling with data-driven learning in complex building environments.

Table 5. Contribution analysis of performance improvements.

Component	Role	Energy Reduction (%)	Comfort Improvement (%)
Predictive Control	Proactive adjustment based on future thermal states	12–15	18–22
Spatial Coordination	Multi-zone thermal coupling management	8–11	20–25
Robust Optimization (DRO)	Handling uncertainty in heat loads	5–8	10–14
Physics-Informed RL	Adaptive policy learning under constraints	10–13	15–20
Operational Flexibility	Adaptive adjustment of comfort bounds	4–6	8–12

Table 6. Sensitivity analysis of wasserstein radius.

Radius ϵ	Energy Cost (k$)	CVaR (k$)	Comfort Violation (%)
0.02	92.5	158.2	12.8
0.08	98.7	135.4	10.2
0.15	105.3	118.6	8.7
0.20	112.8	109.5	8.1
0.25	121.6	104.3	7.9

Table 7. Comparison of control strategies.

Method	Energy Cost (k$)	Comfort Violation (%)	Stability (Variance)
PID Control	135.2	22.5	0.85
Pure RL	118.6	14.2	0.52
Proposed Method	102.4	7.8	0.21

5. Conclusions

This paper presents an integrated optimization framework for heating, ventilation, and air conditioning (HVAC) systems in multi-zone buildings under dynamic and uncertain internal heat loads. Moving beyond conventional assumptions of static occupancy, the study explicitly models internal heat generation as a time-varying and spatially heterogeneous process, capturing the strong coupling between occupant-induced thermal disturbances and building energy systems. By formulating the building as a behavior-aware cyber-physical system, the proposed approach enables a more realistic and responsive representation of thermal dynamics, which is critical for improving both energy efficiency and occupant comfort in complex indoor environments. To address uncertainty in internal heat loads and operational conditions, the framework incorporates distributionally robust optimization based on Wasserstein ambiguity sets. This allows the system to maintain reliable performance under a range of plausible scenarios without relying on precise probabilistic information. In addition, a physics-informed reinforcement learning mechanism is integrated to derive adaptive control policies that can generalize across different operating conditions while preserving thermodynamic feasibility. The combination of robust optimization and learning-based control enables the system to balance short-term responsiveness with long-term stability, resulting in improved coordination across building zones and more efficient utilization of HVAC resources. The case study results demonstrate the effectiveness of the proposed framework in a large-scale multi-zone building setting. Compared to conventional rule-based and model predictive control strategies, the proposed method achieves a reduction in peak energy consumption of approximately 28–32%, while decreasing comfort violation rates by 65–75%. Furthermore, the temperature variability across zones is significantly reduced, with variance decreasing by more than 60%, indicating enhanced thermal stability and improved control performance. These results highlight the importance of incorporating both uncertainty and dynamic internal heat loads into HVAC optimization, particularly in buildings with highly variable usage patterns.

Overall, this work contributes a unified and scalable approach for next-generation building energy management by integrating multi-zone thermal modeling, distributionally robust optimization, and physics-informed learning. The proposed framework provides a practical pathway for improving the performance of HVAC systems in complex buildings, offering both theoretical insights and actionable strategies for real-world applications. Future work may explore the integration of real-time sensing data, extension to districtlevel energy systems, and the incorporation of additional building services to further enhance system-level efficiency and resilience.

To further assess the robustness of the proposed framework, its performance under extreme operating conditions is discussed. In scenarios involving extreme weather variations, such as sudden outdoor temperature spikes or drops, the system is subjected to significantly increased thermal disturbances. Under such conditions, the distributionally robust optimization component enables the control strategy to maintain stable performance by accounting for worst-case deviations in thermal loads and environmental conditions. In addition, the framework demonstrates resilience to sudden equipment-related disruptions, such as partial HVAC capacity degradation or delayed system response. The adaptive nature of the reinforcement learning policy allows the control actions to adjust dynamically in response to evolving system states, thereby mitigating the impact of such disturbances. The constraint-aware formulation further ensures that the system operates within physically feasible limits even under stress conditions. Overall, the combination of uncertainty-aware optimization and adaptive control provides a robust mechanism for handling both environmental extremes and system-level disruptions, thereby enhancing the reliability and practical applicability of the proposed approach.

Despite the promising performance of the proposed framework, several practical challenges should be acknowledged for real-world deployment. First, the integration of distributionally robust optimization and reinforcement learning may introduce additional computational overhead, particularly for large-scale multi-zone systems or real-time control applications. Efficient implementation and potential model simplification strategies may be required to ensure scalability in commercial settings. Second, the effectiveness of the framework relies on the availability of high-quality sensor data, including zone-level temperature, occupancy patterns, and environmental conditions. In practice, such data acquisition may require dense sensing infrastructure and reliable communication systems, which could increase deployment cost and system complexity. These considerations highlight important directions for future work, including improving computational efficiency, reducing data dependency through model abstraction, and developing cost-effective sensing solutions. Addressing these challenges will be essential for translating the proposed framework into practical building energy management systems.

Author Contributions: Conceptualization, Z.S.; Methodology, A.P.Z.; Formal analysis, Z.S.; Investigation, A.P.Z.; Writing—original draft, Z.S. and A.P.Z.; Writing—review & editing, A.P.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Su, B.; Wang, S. An agent-based distributed real-time optimal control strategy for building HVAC systems for applications in the context of future IoT-based smart sensor networks. *Appl. Energy* **2020**, *274*, 115322. [CrossRef]
2. Du, Y.F.; Jiang, L.; Duan, C.; Li, Y.Z.; Smith, J.S. Energy Consumption Scheduling of HVAC Considering Weather Forecast Error Through the Distributionally Robust Approach. *IEEE Trans. Ind. Inform.* **2018**, *14*, 846–857. [CrossRef]
3. Mawson, V.J.; Hughes, B.R. Optimisation of HVAC control and manufacturing schedules for the reduction of peak energy demand in the manufacturing sector. *Energy* **2021**, *227*, 120436. [CrossRef]
4. Shi, H.; Chen, Q. Building energy management decision-making in the real world: A comparative study of HVAC cooling strategies. *J. Build. Eng.* **2021**, *33*, 101869. [CrossRef]

5. Zhao, A.P.; Li, S.; Cao, Z.; Hu, P.J.-H.; Wang, J.; Xiang, Y.; Xie, D.; Lu, X. AI for science: Predicting infectious diseases. *J. Saf. Sci. Resil.* **2024**, *5*, 130–146. [CrossRef]
6. Zhong, W.; Yang, C.; Xie, K.; Xie, S.; Zhang, Y. ADMM-Based Distributed Auction Mechanism for Energy Hub Scheduling in Smart Buildings. *IEEE Access* **2018**, *6*, 45635–45645. [CrossRef]
7. Gómez-Gardars, E.B.; Rodríguez-Macias, A.; Tena-García, J.L.; Fuentes-Cortés, L.F. Assessment of the water–energy–carbon nexus in energy systems: A multi-objective approach. *Appl. Energy* **2022**, *305*, 117872. [CrossRef]
8. Li, P.; Shen, Y.; Shang, Y.; Alhazmi, M. Innovative distribution network design using GAN-based distributionally robust optimization for DG planning. *IET Gener. Transm. Distrib.* **2025**, *19*, e13350. [CrossRef]
9. Rigo-Mariani, R.; Zhang, C.; Romagnoli, A.; Kraft, M.; Ling, K.V.; Maciejowski, J. A Combined Cycle Gas Turbine Model for Heat and Power Dispatch Subject to Grid Constraints. *IEEE Trans. Sustain. Energy* **2020**, *11*, 448–456. [CrossRef]
10. Giordano, L.; Furlan, G.; Puglisi, G.; Cancellara, F.A. Optimal design of a renewable energy-driven polygeneration system: An application in the dairy industry. *J. Clean. Prod.* **2023**, *405*, 136933. [CrossRef]
11. Zhao, A.P.; Zhang, Q.; Alhazmi, M.; Hu, P.J.-H.; Zhang, S.; Yan, X. AI for science: Covert cyberattacks on energy storage systems. *J. Energy Storage* **2024**, *99*, 112835. [CrossRef]
12. Gao, S.; Yang, L.; Li, Y.; Liu, S.; Zhang, H.; Arens, E.; Zhai, Y. Metabolic rate in children and adolescents: Tabulate values for common activities and comparisons with standards and adult values. *Build. Environ.* **2023**, *244*, 110804. [CrossRef]
13. Bajaj, S.; Verma, M.; Sharma, H.B.; Ramaiya, K.; Bahendeka, S.; Kalra, S. Southeast Asian, African, and Middle East expert consensus on structured physical activity—Dance, exercise, and sports. *Adv. Ther.* **2025**, *42*, 1692–1715. [CrossRef]
14. Garay, J.L.; Sebe, J.G.; Strickland, J.; Graves, L.; Voss, M. A Use of resting metabolic rate ratio as a relative energy deficiency in sports indicator in female athletes. *Curr. Dev. Nutr.* **2025**, *9*, 106007. [CrossRef]
15. Li, P.; Yan, X.; Qian, T.; Shen, Y.; Alhazmi, M. Biodegradable Transient Data Centers: A Multi-Objective Optimization Framework for Sustainable Computing in Extreme Environments. *IEEE Trans. Ind. Appl.* **2026**, 1–9 . [CrossRef]
16. Bhattacharya, A.; Vasisht, S.; Adetola, V.; Huang, S.; Sharma, H.; Vrabie, D.L. Control co-design of commercial building chiller plant using Bayesian optimization. *Energy Build.* **2021**, *246*, 111077. [CrossRef]
17. Li, H.; Wang, Z.-H.; Zhang, B. How social interaction induce energy-saving behaviors in buildings: Interpersonal & passive interactions vs public & active interactions. *Energy Econ.* **2023**, *118*, 106515.
18. Huckebrink, D.; Finke, J.; Bertsch, V. How user behaviour affects emissions and costs in residential energy systems—The impacts of clothing and thermal comfort. *Environ. Res. Commun.* **2023**, *11*, 115009. [CrossRef]
19. Appau, W.M.; Nubi, T.G.; Aziabah, S.A. Adaptive HVAC cooling control and its impact on energy cost optimization in multi-tenanted shopping malls in Ghana. *Int. J. Constr. Manag.* **2026**, 1–17.. [CrossRef]
20. Pezhmani, Y.; Oskouei, M.Z.; Rezaei, N.; Mehrjerdi, H. A centralized stochastic optimal dispatching strategy of networked multi-carrier microgrids considering transactive energy and integrated demand response: Application to water–energy nexus. *Sustain. Energy Grids Netw.* **2022**, *31*, 100751. [CrossRef]
21. Wang, H.; Xu, J.; Zhao, C.; Lu, Z.; Cheng, Y.; Chen, X.; Zhang, X.-P.; Liu, Y.; Chen, X. Transformloc: Transforming mavs into mobile localization infrastructures in heterogeneous swarms. In *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*; IEEE: New York City, NY, USA, 2024; pp. 1101–1110.
22. Li, P.; Gu, C.; Cheng, X.; Li, J.; Alhazmi, M. Integrated energy-water systems for community-level flexibility: A hybrid deep Q-network and multi-objective optimization framework. *Energy Rep.* **2025**, *13*, 4813–4826. [CrossRef]
23. Wu, H.; Chen, J.; Vaishnav, P.; Sun, M.; Craig, M.T. Technological improvements in EV batteries offset climate-induced durability challenges. *Nat. Clim. Change* **2026**, *16*, 460–467. [CrossRef]
24. Wu, H.; Kong, Q.; Huber, M.; Sun, M.; Craig, M.T. Climate change will increase high-temperature risks, degradation, and costs of rooftop photovoltaics globally. *Joule* **2026**, *10*, 102218. [CrossRef]
25. Wu, H.; Sun, M.; Craig, M.T. Updating Global Green Hydrogen Production Costs and Configurations under Future Climates. *Innovation* **2026**, *7*, 101303. [CrossRef]