Can reinforcement learning move beyond hype in building control?

Muhammad Saad Ul Haq,[1] Wenjie Ji,[1,*] Yang Geng,[2,3] and Borong Lin[2,3]
[1]School of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China
[2]School of Architecture, Tsinghua University, Beijing, 100084, China
[3]Key Laboratory of Eco Planning & Green Building, Tsinghua University, Beijing, 100084, China
*Correspondence: jiwenjie@bit.edu.cn
Received: March 10, 2026; Accepted: April 16, 2026; Published Online: April 21, 2026; https://doi.org/10.59717/j.xinn-energy.2026.100150

Citation: Saad Ul Haq M., Ji W., Geng Y., et al. (2026). Can reinforcement learning move beyond hype in building control? *The Innovation Energy* **3**:100150.

Dear Editor,

Reinforcement learning (RL) has emerged as a prominent paradigm in literature on intelligent building energy management (BEM). The appeal of RL lies in its promise to move beyond hand-crafted rules and carefully calibrated models. In principle, an RL agent can learn control policies directly through interaction, adapting to changing weather, shifting occupancy, and time-varying electricity prices. As buildings are increasingly expected to function as flexible, grid-interactive resources in support of decarbonization, this promise has fueled a rapidly expanding literature that presents deep RL as a pathway to higher-performance building management systems, offering simultaneous gains in energy efficiency, occupant comfort, and responsiveness to grid signals. At the same time, national-scale analyses of building energy systems present a more grounded picture of where system-level value currently arises. Buildings already account for a dominant share of electricity consumption, and coordinated improvements in efficiency and demand flexibility are projected to yield tens of billions of dollars in annual power system savings and substantial emissions reductions by 2030 using commercially available technologies and established control strategies.[1,2] Against this evidence, the maturity and incremental value of RL-based building control deserves careful scrutiny.

In this article, we contend that the prevailing narrative surrounding RL in building control substantially overstates its readiness for real-world deployment. While RL remains a promising research direction, its current portrayal as a near-term solution for operational building management is difficult to reconcile with available deployment evidence, empirical adoption patterns, and the engineering realities of grid-interactive efficient buildings (GEBs). Our concern is therefore not with the long-term potential of learning-based control itself, but with the strength of present readiness claims relative to the still limited record of operational validation.

THE ALLURE AND THE CLAIMS

The appeal of RL in buildings is clear. RL promises model-free or model-light adaptation, the ability to learn through interaction, and a flexible framework for optimizing across multiple objectives without requiring fully specified control models. In principle, these features are especially attractive in settings characterized by uncertain disturbances, evolving operating conditions, or supervisory decisions that extend across longer horizons. In simulation studies, RL controllers frequently report energy savings relative to rule-based or heuristic baselines, reinforcing the perception that learning-based control is uniquely suited to complex and uncertain building environments.[3,4]

However, systematic reviews paint a more cautious picture. Across the building-control literature, the overwhelming majority of RL studies remain confined to simulation, often over short horizons and under idealized assumptions.[4,5] Real-building demonstrations are still limited, and sustained multi-season operation with full BAS integration remains exceptional rather than representative. Importantly, these limitations should not be attributed solely to RL in the abstract. In many cases, reported weaknesses reflect how the control problem is posed, where the boundary of the learning environment is drawn, and whether RL is assigned to low-level actuation or to a more appropriate supervisory role within a broader control architecture. At the same time, these design choices are not incidental. They shape whether the resulting controller is aligned with the operational realities of buildings, where comfort constraints, equipment protection, fallback logic, and limited tolerance for exploration are fundamental rather than optional.

Recent reviews continue to underscore this tension. More than 85-90% of published RL studies rely exclusively on simulation-based evaluation, only a small minority report any form of field implementation, and explicit treatment of comfort or equipment safety constraints remains relatively uncommon. Nagy et al. explicitly identify unresolved questions regarding safety, benchmarking, reproducibility, and the absence of credible pathways from simulation to deployment in RL-based building control.[4] Similar conclusions emerge from reviews focused on occupant-centric and behavior-driven RL, where stylized behavior models further widen the gap between reported performance and real-world applicability.[5] Figure 1A summarizes the core tension between the promise of RL and the realities of operational building control.

LESSONS FROM GRID-INTERACTIVE EFFICIENT BUILDINGS

The contrast with GEB literature is instructive. National and regional assessments show that substantial system-level benefits can already be achieved through coordinated efficiency and demand flexibility measures grounded in mature technologies and conservative deployment assumptions. Modeling based on hourly system costs and emissions indicates that efficiency and flexible load control could reduce annual power-sector CO_2 emissions by approximately 80 million tons by 2030, while cumulative system savings are projected to reach roughly \$100-200 billion over the corresponding analysis horizon.[1,2]

Importantly, these results are not contingent on autonomous learning-based control. Instead, they rely on predictable mechanisms such as peak load reduction, load shifting to low-cost hours, and improved baseline efficiency, supported by interoperable controls, standardized communication protocols, and robust measurement and verification frameworks. Consistent with this evidence, verified grid-scale benefits to date are dominated by mature efficiency measures, supervisory control, and selected model-based strategies rather than autonomous learning-based strategies. In this context, control intelligence is only one layer within a broader socio-technical system that includes hardware readiness, interoperability, cybersecurity, regulatory alignment, and workforce capability.

The GEB literature repeatedly emphasizes that barriers to deployment are not primarily algorithmic. As summarized in figures, challenges related to interoperability, data availability, cybersecurity, valuation of demand flexibility, and institutional trust dominate the landscape.[1] These factors do not merely complicate RL deployment; they define the practical conditions under which any advanced control method must prove its value. Figure 1B positions RL within this broader control landscape by mapping control approaches against increasing deployment realism and system integration. Established strategies such as rule-based supervisory control and model predictive control cluster in regions associated with verified deployment, commissioning, and grid participation. In contrast, most RL-based approaches occupy regions characterized by higher algorithmic sophistication but limited empirical validation, remaining largely confined to simulation or short-horizon pilots. By contrasting these trajectories, the figure illustrates that progress in algorithmic capability has outpaced demonstrated readiness for operational, grid-interactive building deployment.

WHY REINFORCEMENT LEARNING STRUGGLES IN PRACTICE

Against this backdrop, several recurring limitations undermine claims that RL is ready to serve as a primary control paradigm for buildings. Generalization remains fragile, particularly when policies trained in narrow simulation settings are exposed to seasonal variation, equipment aging, sensor drift, or evolving occupancy patterns.[4] Data efficiency and training burden pose an additional obstacle, since extensive interaction data are costly, slow, and

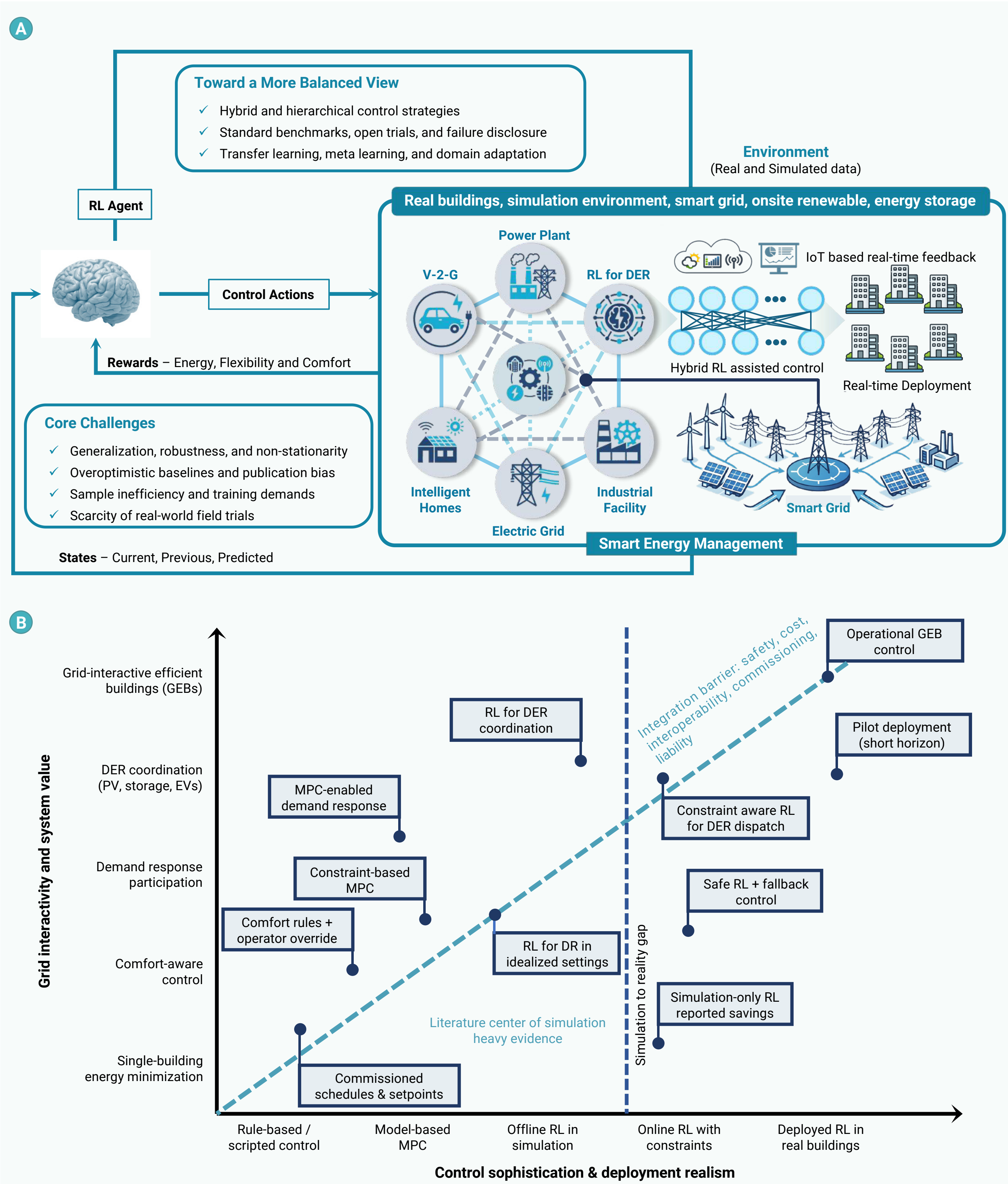

Figure 1. Bridging RL concepts and deployment reality in grid-interactive buildings (A) RL-based smart energy management schematic, highlighting key barriers and practical directions. (B) Evidence-weighted maturity ladder from conventional rule-based HVAC operation and constraint-based MPC to simulation-heavy RL results, then to safe/constraint-aware RL, short-horizon pilots, and ultimately operational GEB control illustrating the persistent sim-to-real gap. Abbreviations: MPC, model predictive control; DR, demand response; DER, distributed energy resources; PV, photovoltaic generation; EV, electric vehicle.

potentially risky to obtain in occupied buildings.[5] Safety and constraint handling also remain unresolved in much of the literature, where comfort and equipment limits are often addressed indirectly through reward design rather than explicit enforcement. At the same time, these shortcomings should not be interpreted as intrinsic to RL in every formulation. They often reflect how the learning problem is posed, where the environment boundary is drawn, and whether RL is assigned to direct low-level actuation or to a more suitable supervisory role. Nor are these burdens unique to RL. In building problems involving long horizons, discrete actions, uncertainty, and coordinated operation of multiple assets, model-based control can also require significant approximation to remain tractable.[1] The relevant question is therefore not which paradigm appears stronger in principle, but which formulations have demonstrated reliable performance under real deployment constraints. From that standpoint, current evidence remains limited.[4,5]

TOWARD A BALANCED WAY FORWARD

Critiquing overstatement does not imply rejecting RL. RL remains attractive where dynamics are difficult to model explicitly, where supervisory decisions extend across longer horizons, or where multiple assets and objectives must be coordinated under uncertainty.[3,4] The more credible path forward, however, lies less in framing RL as a standalone replacement for existing control practice and more in hybrid or hierarchical architectures in which learning-based methods operate at supervisory or advisory levels while rule-based or model-based controllers enforce comfort, safety, and equipment constraints. In this sense, RL and MPC are often better understood as complementary rather than competing approaches.

Advances in safe and constrained RL, transfer learning, offline training, and standardized benchmarking platforms may further narrow the gap between methodological development and operational requirements. But these advances must still be validated in real buildings over meaningful time horizons before strong deployment claims can be justified.[4,5] Transparent reporting of training cost, failure modes, operator interventions, and negative results, alongside longer-horizon field trials, remains essential. Without such evidence, claims of readiness risk outpacing what has actually been demonstrated in practice.[1,2]

CHALLENGES AND OUTLOOKS

The trajectory of RL in BEM reflects a recurring pattern in applied artificial intelligence research, wherein methodological progress advances more rapidly than operational evidence. Apart from the limited extent of real-world validation, several technical challenges remain unresolved. Generalization across buildings and climatic conditions is still weak, reward design frequently acts as a surrogate for explicit constraint enforcement, and training performance remains highly sensitive to environment formulation, data availability, and simulator fidelity. These limitations are particularly critical in building applications, where control actions are required to remain stable, interpretable, and robust under slow yet uncertain operating conditions.

The outlook, therefore, is not one of rejection but of further refinement. Future progress is likely to depend on RL formulations that are more data efficient, more constraint aware, and better aligned with the supervisory nature of building operation. In this regard, recent studies have increasingly moved beyond early value-based methods such as Q-learning toward actor-critic and policy-optimization approaches, including advantage actor-critic, asynchronous advantage actor-critic, deep deterministic policy gradient, soft actor-critic, and proximal policy optimization, along with transfer-learning and meta-learning strategies aimed at reducing retraining burden and improving adaptability. If these directions mature in parallel with more realistic evaluation practices, RL may yet emerge as a credible component of advanced building control. Until then, its role is more appropriately understood as promising and evolving rather than operationally established.

REFERENCES

1. U.S. Department of Energy. (2021). A National Roadmap for Grid-Interactive Efficient Buildings. DOE.
2. MimsFrick N. and Schwartz L. (2019). Time-Sensitive Value of Efficiency. Lawrence Berkeley National Lab.
3. Ding C., Zhang X., Li Y., et al. (2024). Data-driven control strategies for complex energy systems under uncertainty. *Nat. Commun.* **15**:5916. DOI:10.1016/j.apenergy.2021.118148
4. Nagy Z., Henze G., Dey S., et al. (2023). Ten questions concerning reinforcement learning for building energy management. *Build. Environ.* **241**:110435. DOI:10.1016/j.buildenv.2023.110435
5. Yu H., Tam V.W.Y. and Xu X. (2024). A systematic review of reinforcement learning application in building energy-related occupant behavior simulation. *Energy Build.* **312**:113825. DOI:10.1016/j.enbuild.2024.114189

FUNDING AND ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Grant No. 52578107, 52425801, 52130803 and 52394223), and the Young Elite Scientist Sponsorship Program by China Association for Science and Technology (No. YESS20240622). Academic Papers of the 28th Annual Meeting of the China Association for Science and Technology.

AUTHOR CONTRIBUTIONS

Conceptualization, M.S.U.H. and W.J.; methodology, M.S.U.H. and W.J.; investigation, M.S.U.H.; formal analysis, M.S.U.H.; visualization, M.S.U.H.; writing – original draft, M.S.U.H. and W.J.; writing – review & editing, M.S.U.H., W.J., Y.G., and B.L.; supervision, W.J., Y.G., and B.L.; resources, W.J., Y.G., and B.L.; project administration, W.J.; funding acquisition, W.J., Y.G., and B.L. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.