KNOWLEDGE-DRIVEN REINFORCEMENT LEARNING FOR DEMAND SIDE MULTI-BUILDING ENERGY MANAGEMENT

PHD THESIS DEFENSE

SHARATH RAM KUMAR
JOINT PHD CANDIDATE
UNIVERSITÉ GRENOBLE-ALPES, FRANCE
NANYANG TECHNOLOGICAL UNIVERSITY, SINGAPORE

SUPERVISORS
PROF BENOIT DELINCHANT (GRENOBLE-INP, CNRS, UGA)
ASSOC PROF ARVIND EASWARAN (NTU)
DR RÉMY RIGO-MARIANI (GRENOBLE-INP, CNRS)

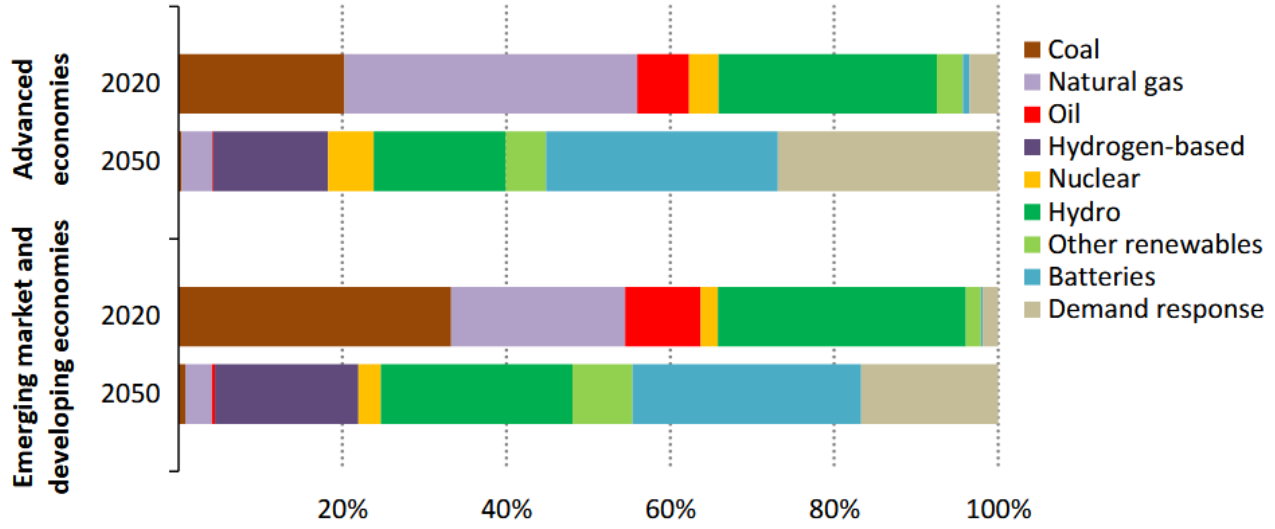
JURY MEMBERS
PROF PATRICK REIGNER (UGA), REVIEWER
PROF RUI TAN (NTU), REVIEWER
PROF DIPTI SRINIVASAN (NUS SINGAPORE), REVIEWER
PROF GILLES GUERASSIMOFF (PARIS-MINES), REVIEWER

SECTION 1

INTRODUCTION

1 INTRODUCTION

Figure 4.18 ▷ Electricity system flexibility by source in the NZE



To meet four-times the amount of hour-to-hour flexibility needs, batteries and demand response step up to become the primary sources of flexibility

Source : IEA. *Net Zero by 2050 – Analysis*. IEA. https://www.iea.org/reports/net-zero-by-2050 (accessed 2024-10-20).

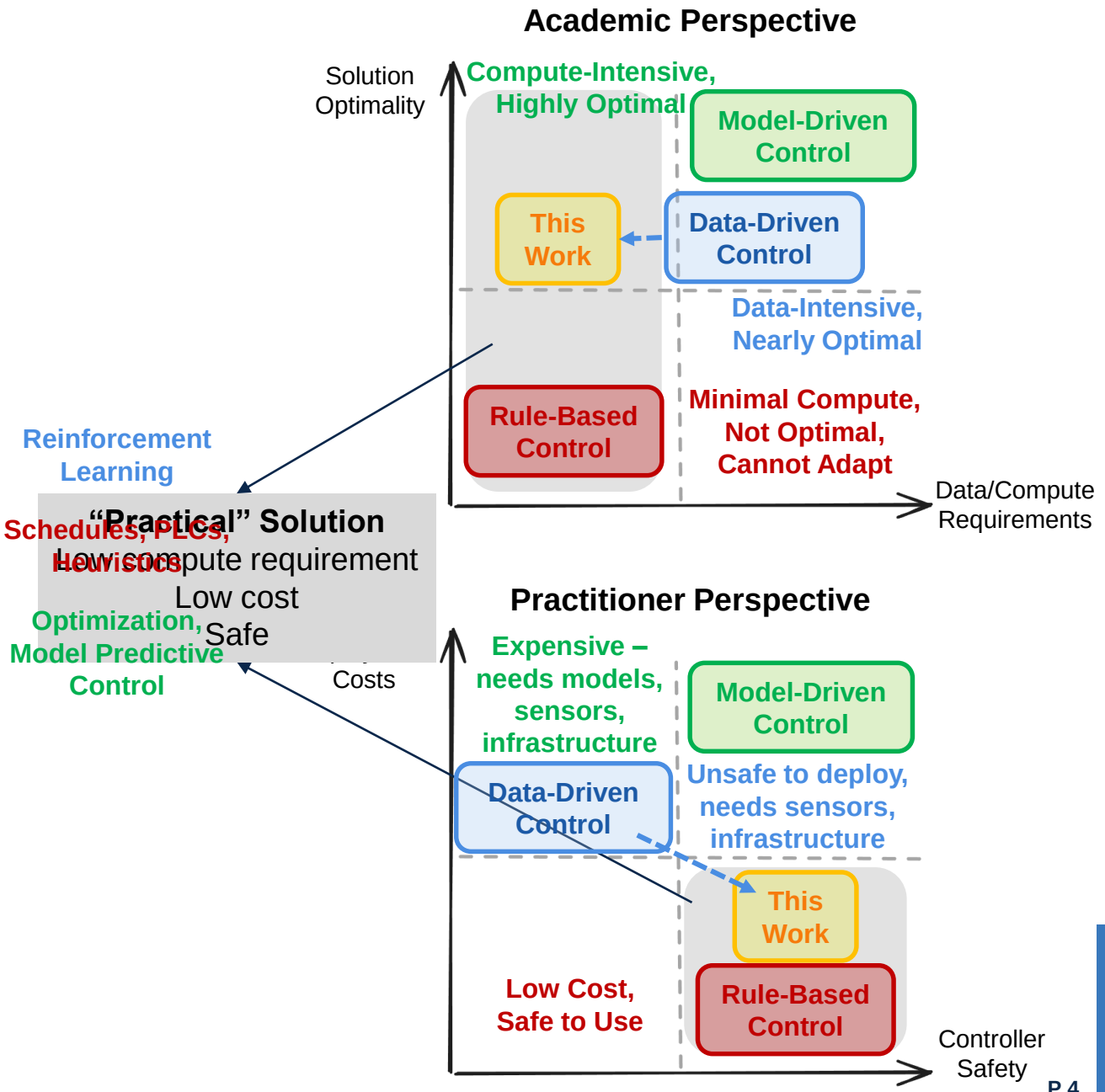
- **Global effort towards massive reduction in emissions – eg Net Zero by 2050 (NZE2050)**
- **Renewables are intermittent – need for flexibility**
  - Demand response (DR) – consumers respond to grid needs
- **DR in practice – building energy management**
  - Grid-aware control of AC, battery systems, appliances
  - 10x required growth for NZE : 25 GW (2020) to 250 GW (2030)

cnrs @CREATE Singapore

THE RESEARCH GAP

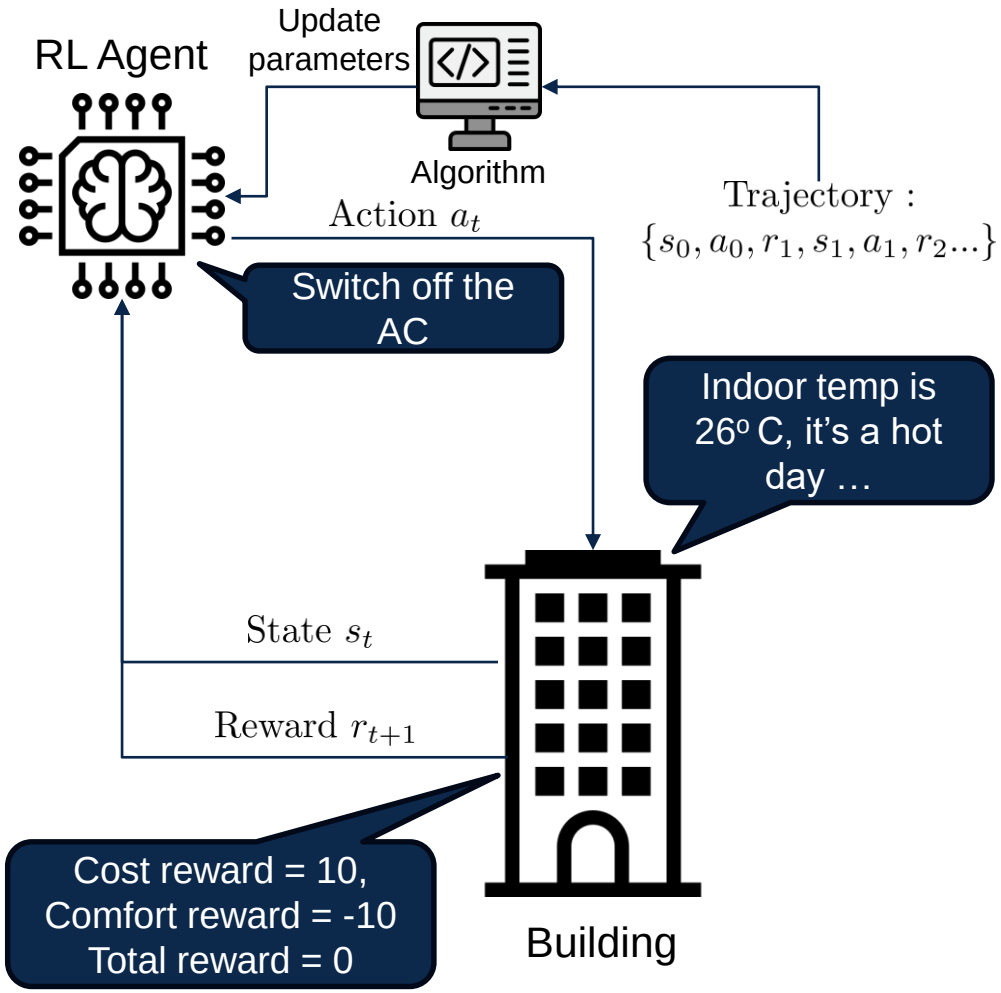
- **Academic and Practitioner priorities are very different**
  - "Abyss between algorithms and applications" (Henze et al, 2024)

- **The research gap –**
  - Practical and scalable control solutions
  - Address real-world challenges



Henze, G. P.; Kircher, K. J.; Braun, J. E. Why Has Advanced Commercial HVAC Control Not yet Achieved Its Promise? *Journal of Building Performance Simulation* **2024**, 1–12. https://doi.org/10.1080/19401493.2024.2429728.

REINFORCEMENT LEARNING AND CHALLENGES

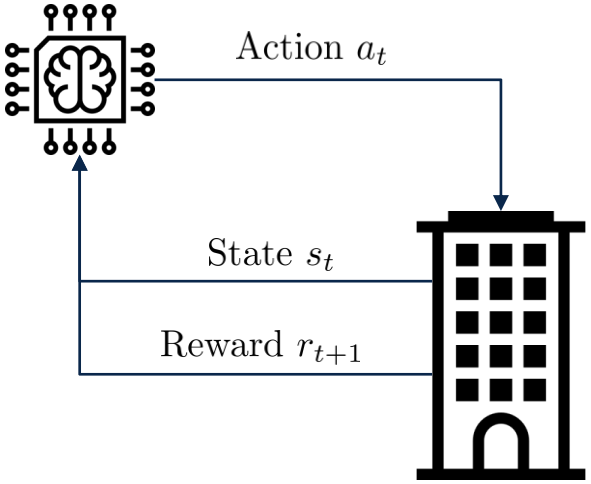


- **Trial-and-error learning through direct interaction and feedback**
- **RL Challenges in BEM (Nagy et al 2023)**
  - Learning challenges
    - Long learning times (millions of interactions ~ years of data)
  - Infrastructure
    - Limited sensors and control equipment in majority of buildings
  - Costs and Benefits
    - Economic costs (return on investment), other benefits (eg better comfort)
  - Safety, Security and Trust
    - Chance for unsafe controls especially during learning phase

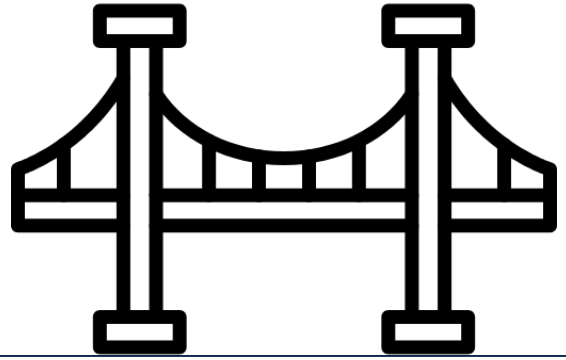
Nagy, Z.; Henze, G.; Dey, S.; Arroyo, J.; Helsen, L.; Zhang, X.; Chen, B.; Amasyali, K.; Kurte, K.; Zamzam, A.; Zandi, H.; Drgoňa, J.; Quintana, M.; McCullogh, S.; Park, J. Y.; Li, H.; Hong, T.; Brandi, S.; Pinto, G.; Capozzoli, A.; Vrabie, D.; Bergés, M.; Nweye, K.; Marzullo, T.; Bernstein, A. Ten Questions Concerning Reinforcement Learning for Building Energy Management. *Building and Environment* **2023**, *241*, 110435. https://doi.org/10.1016/j.buildenv.2023.110435.

FOCUS OF THE THESIS

Reinforcement Learning  Domain Knowledge  Real-World Challenges



Physics, Models, Floorplans, Existing Controls etc.



- Learning Time, Data Requirements
- Limited Sensors or Control Equipment
- Cost & Benefits
- Control safety and acceptability

How can domain knowledge be systematically incorporated into reinforcement learning workflows to address these challenges?

AGENDA

2. Knowledge-Informed Reinforcement Learning
3. Case-study: Single-Building Control
4. Case-study: Multi-Building Coordination
5. Conclusions and Perspectives

SECTION 2

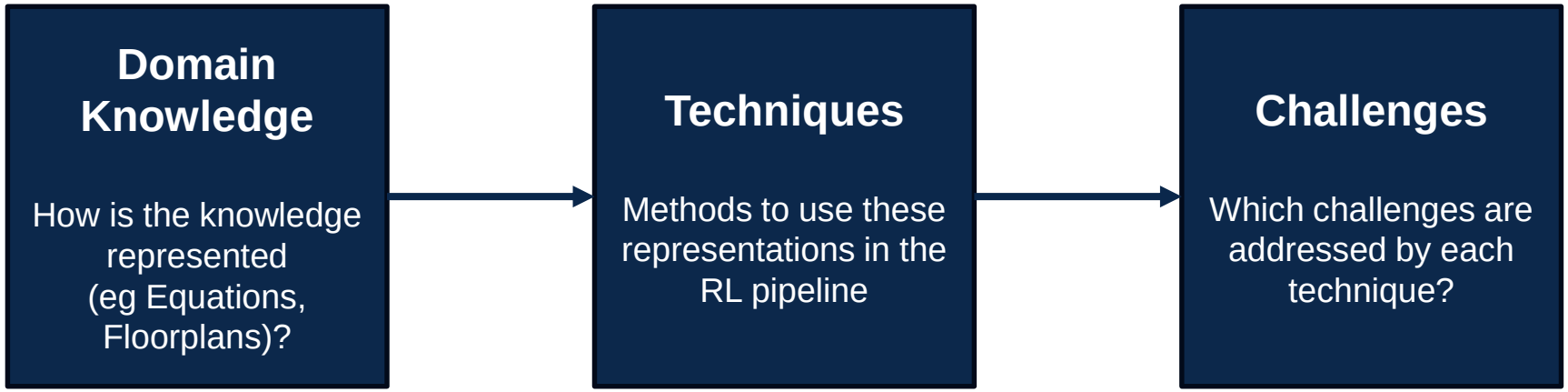
KNOWLEDGE-INFORMED REINFORCEMENT LEARNING

A SYSTEMATIC APPROACH TO DOMAIN KNOWLEDGE INTEGRATION IN RL

2

KNOWLEDGE-INFORMED REINFORCEMENT LEARNING

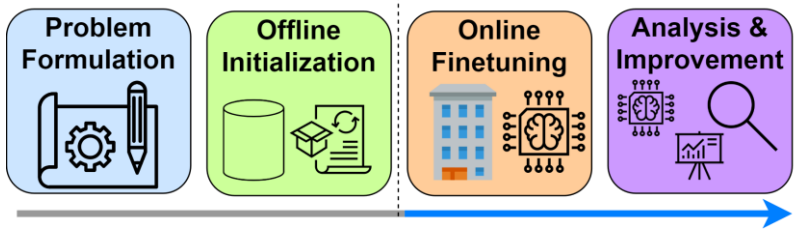
- **Extension of Informed Machine Learning (von Rueden 2023) to RL**
  - Explicit focus on domain-specific challenges



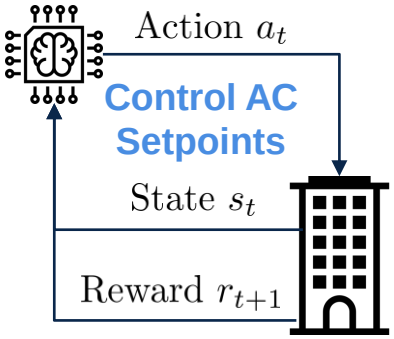
"Which knowledge representations do I need to solve a given challenge?"

"Which challenges can I solve with the knowledge representations I have?"

von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Pfrommer, J.; Pick, A.; Ramamurthy, R.; Walczak, M.; Garcke, J.; Bauckhage, C.; Schuecker, J. Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. *IEEE Transactions on Knowledge and Data Engineering* **2023**, *35* (1), 614–633. https://doi.org/10.1109/TKDE.2021.3079836.



STEPS IN KNOWLEDGE-INFORMED RL APPROACH

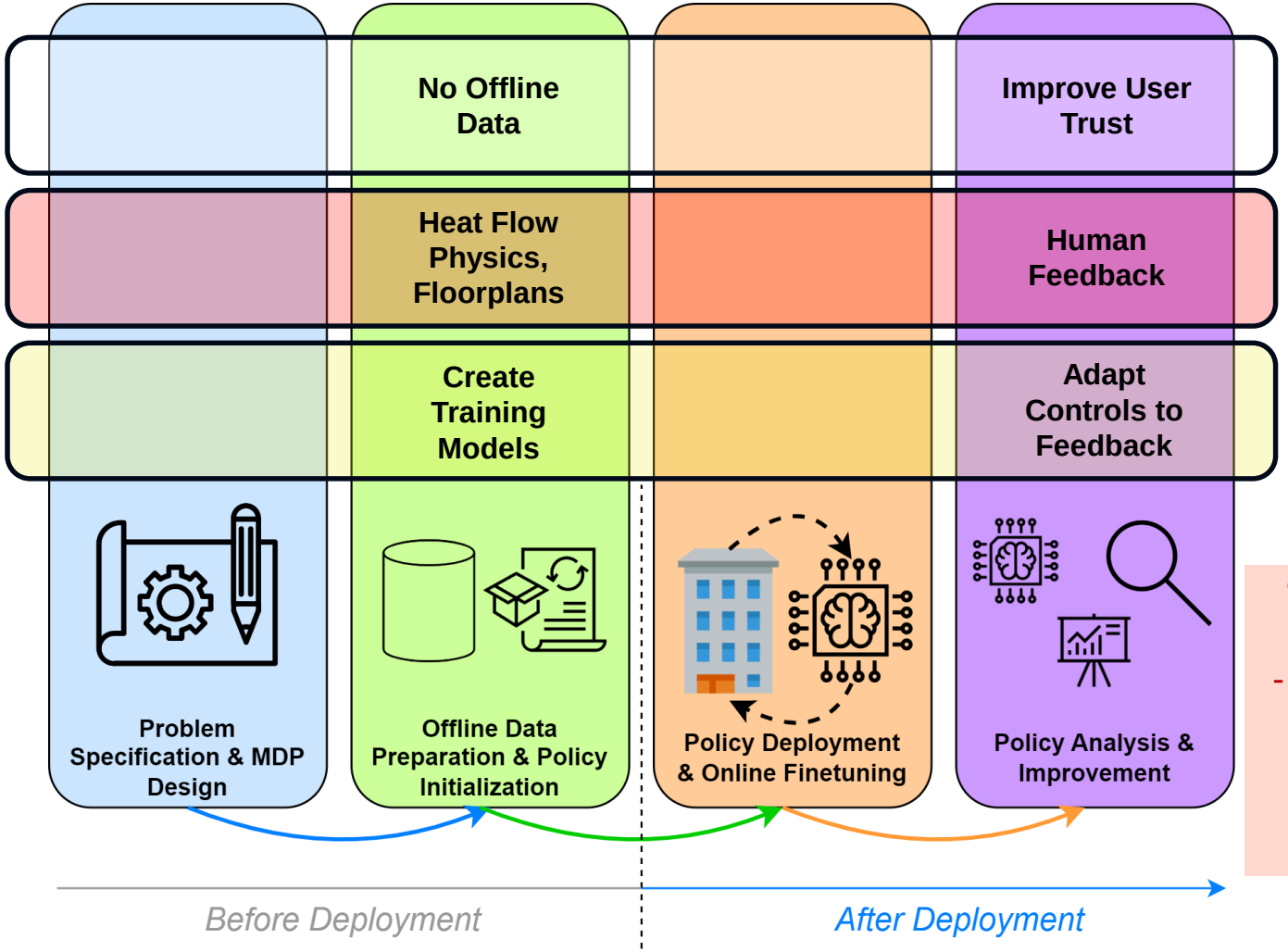


Domain Knowledge

Scientific:
Heat Flow Physics

Expert:
Floorplans

World:
Thermal Comfort



Challenges

Domain Knowledge

Techniques

Thesis Contributions

- *Knowledge-Informed RL Formalism*
- *Taxonomy of Domain Knowledge*
- *Set of Techniques*

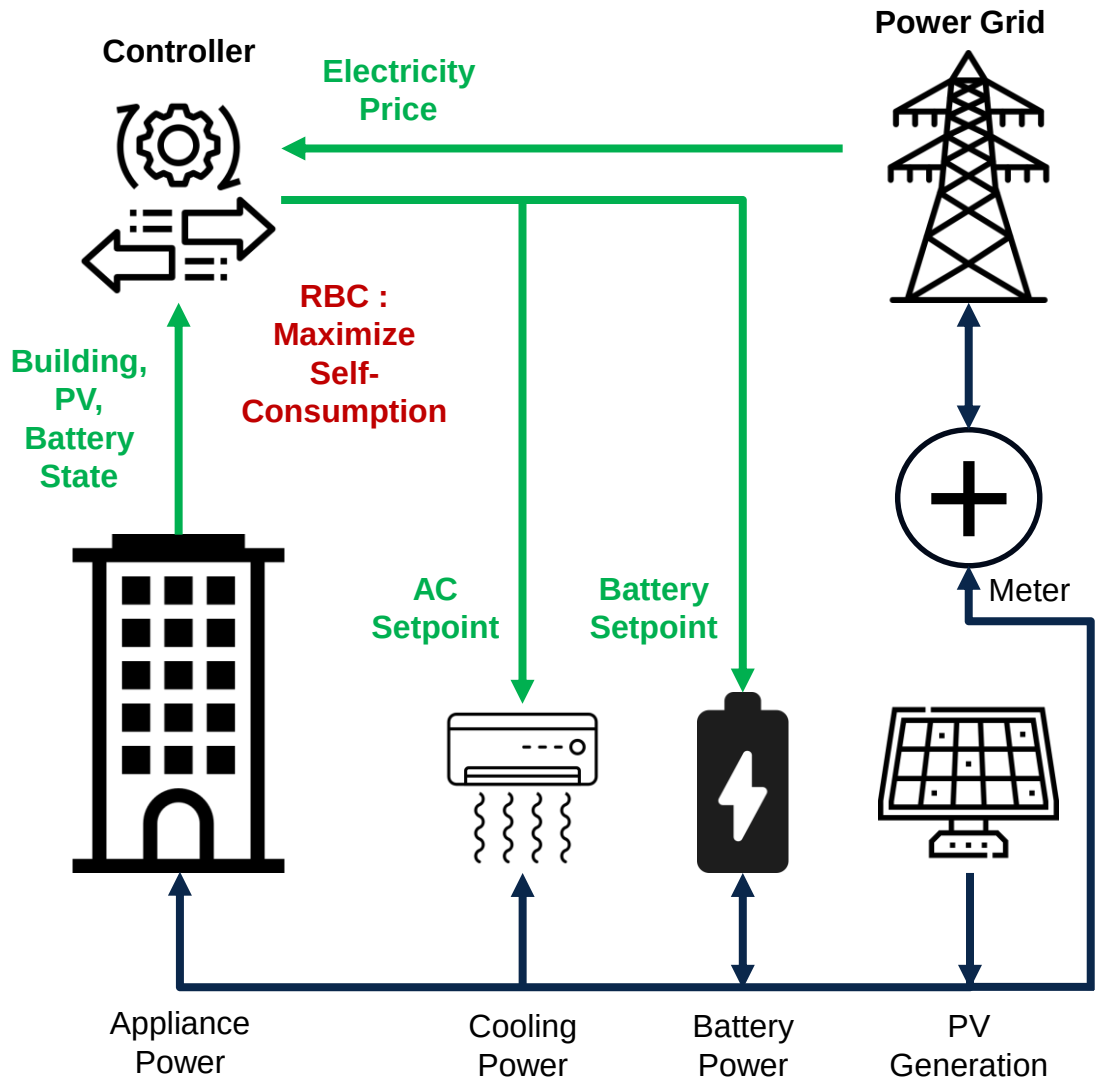
SECTION 3

APPLICATION TO A SINGLE-BUILDING CONTROL PROBLEM

HOW CAN THIS APPROACH BENEFIT PARTICIPANTS IN DEMAND RESPONSE?

3 APPLICATION TO A SINGLE-BUILDING CONTROL PROBLEM

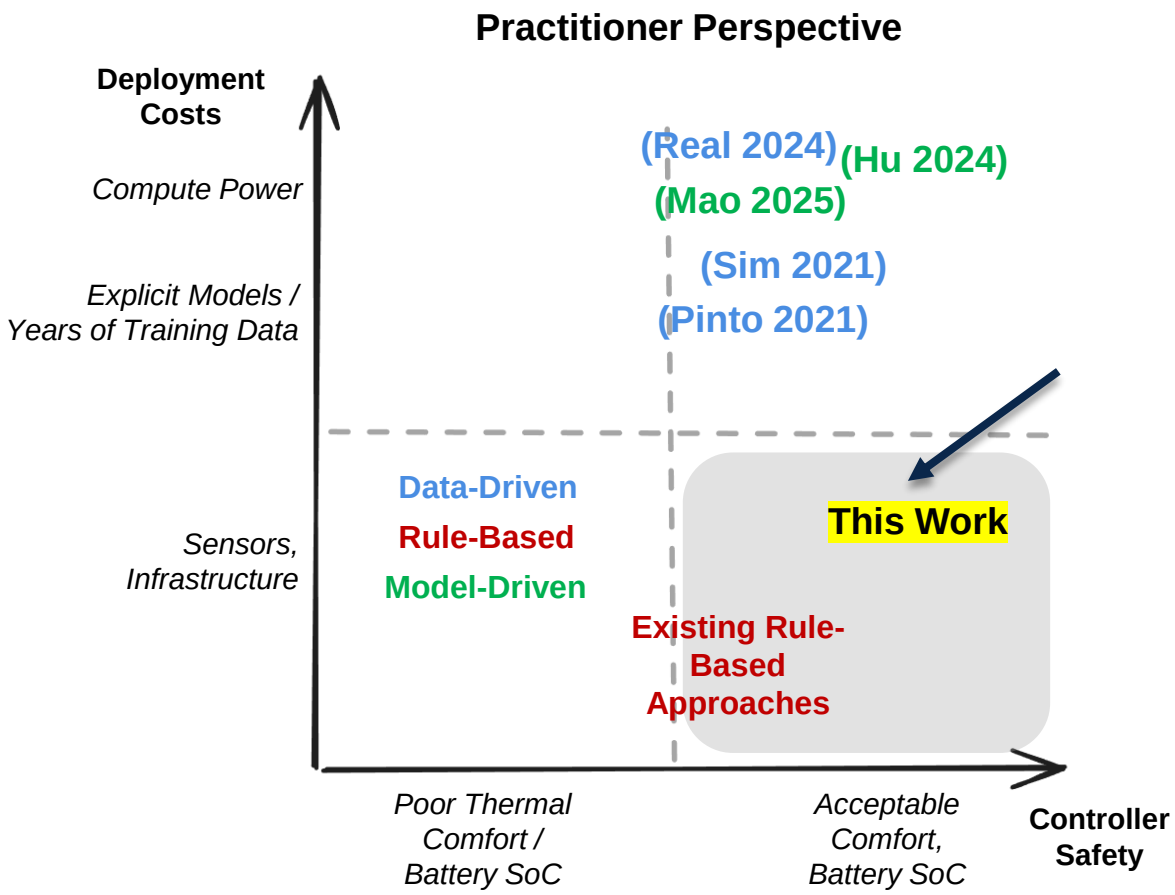
- **Scenario**
  - Grid-connected office building in Singapore with AC, Battery and PV
  - Control AC and Battery based on dynamic electricity prices
  - **Objectives** : Reduce electricity bills, maintain thermal comfort score
- **Considerations**
  - Must obey safety rules
  - No historical data; 1-year deployment window
  - Control period : 15 minutes



cnrs @CREATE Singapore

“Practical” Solution
Low cost
Safe

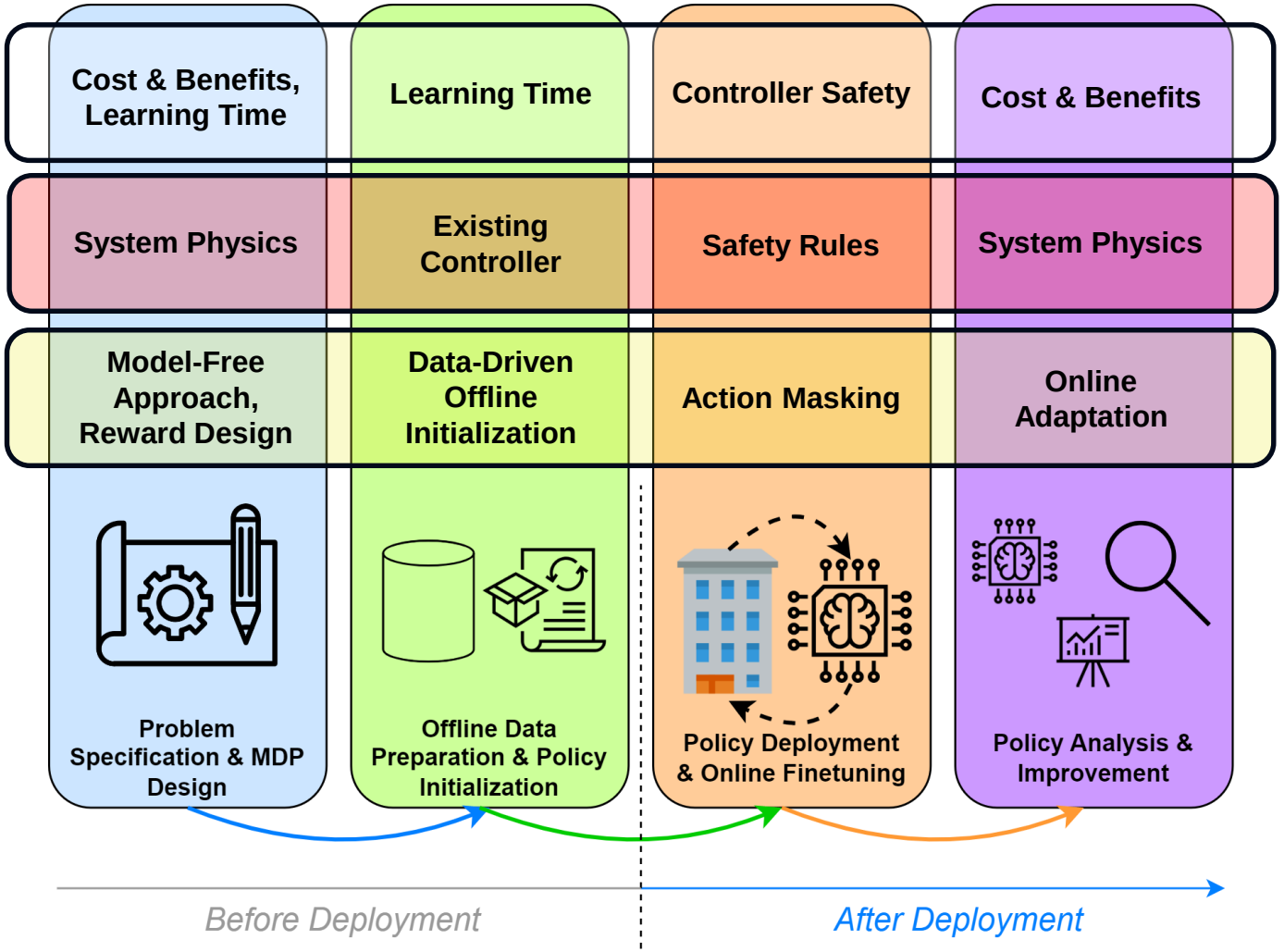
NOVELTY OF THIS WORK



- **Few works focus on combined control of AC + Battery**
- **Existing work typically requires models**
  - Optimization based on forecasts
  - Offline training for RL
- **Main Novelties**
  - ✓ AC + Battery control
  - ✓ No pre-existing model or data required
  - ✓ Respects thermal comfort and SoC limits
  - ✓ Economic savings within realistic timeframe

Khabbazi, A. J.; Pergantis, E. N.; Reyes Premer, L. D.; Papageorgiou, P.; Lee, A. H.; Braun, J. E.; Henze, G. P.; Kircher, K. J. Lessons Learned from Field Demonstrations of Model Predictive Control and Reinforcement Learning for Residential and Commercial HVAC: A Review. *Applied Energy* **2025**, *399*, 126459. https://doi.org/10.1016/j.apenergy.2025.126459.

KNOWLEDGE-INFORMED RL SOLUTION ARCHITECTURE

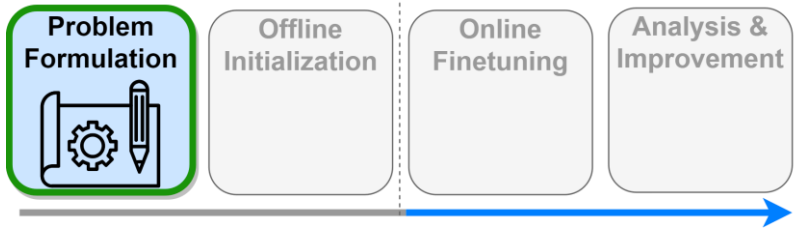


Domain Knowledge Use:

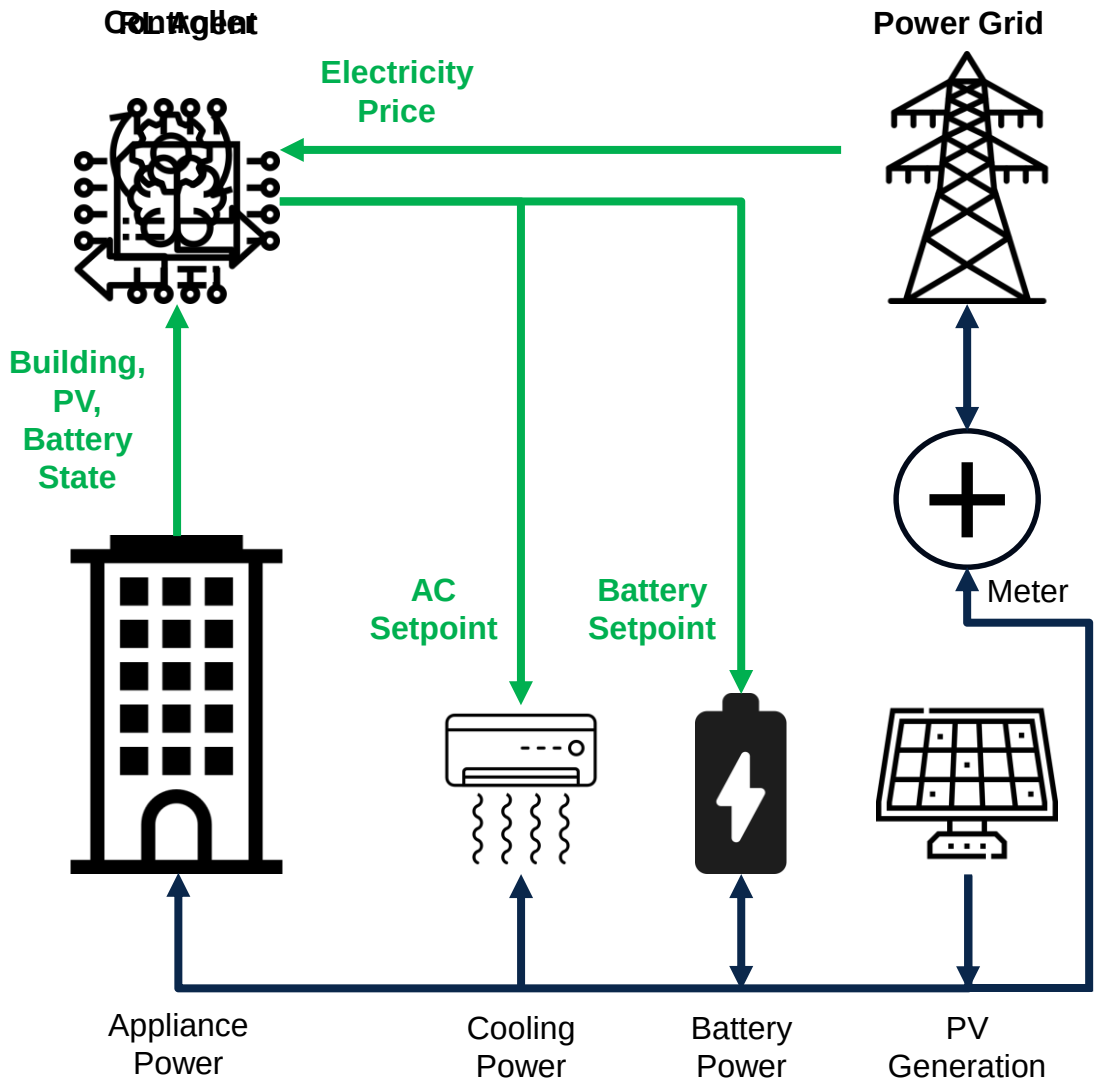
- Encourage specific behaviours
- Learn from existing controller
- Ensure controller safety

Kumar, S. R.; Rigo-Mariani, R.; Delinchant, B.; Easwaran, A. Towards Safe Model-Free Building Energy Management Using Masked Reinforcement Learning. In *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*; IEEE, 2023; pp 1–5. https://doi.org/10.1109/ISGTEUROPE56780.2023.10407781.
Kumar, S. R.; Easwaran, A.; Delinchant, B.; Rigo-Mariani, R. Behavioural Cloning Based RL Agents for District Energy Management. In *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*; ACM: New York, NY, USA, 2022; pp 466–470. https://doi.org/10.1145/3563357.3566165.

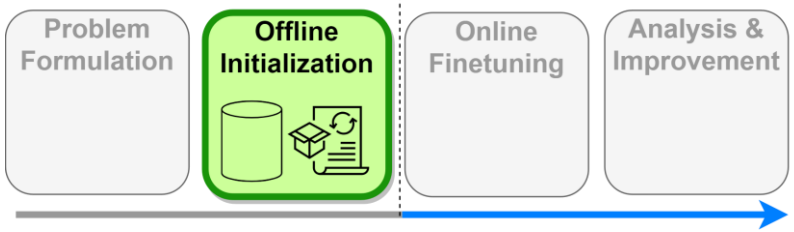
PROBLEM FORMULATION STRATEGIES



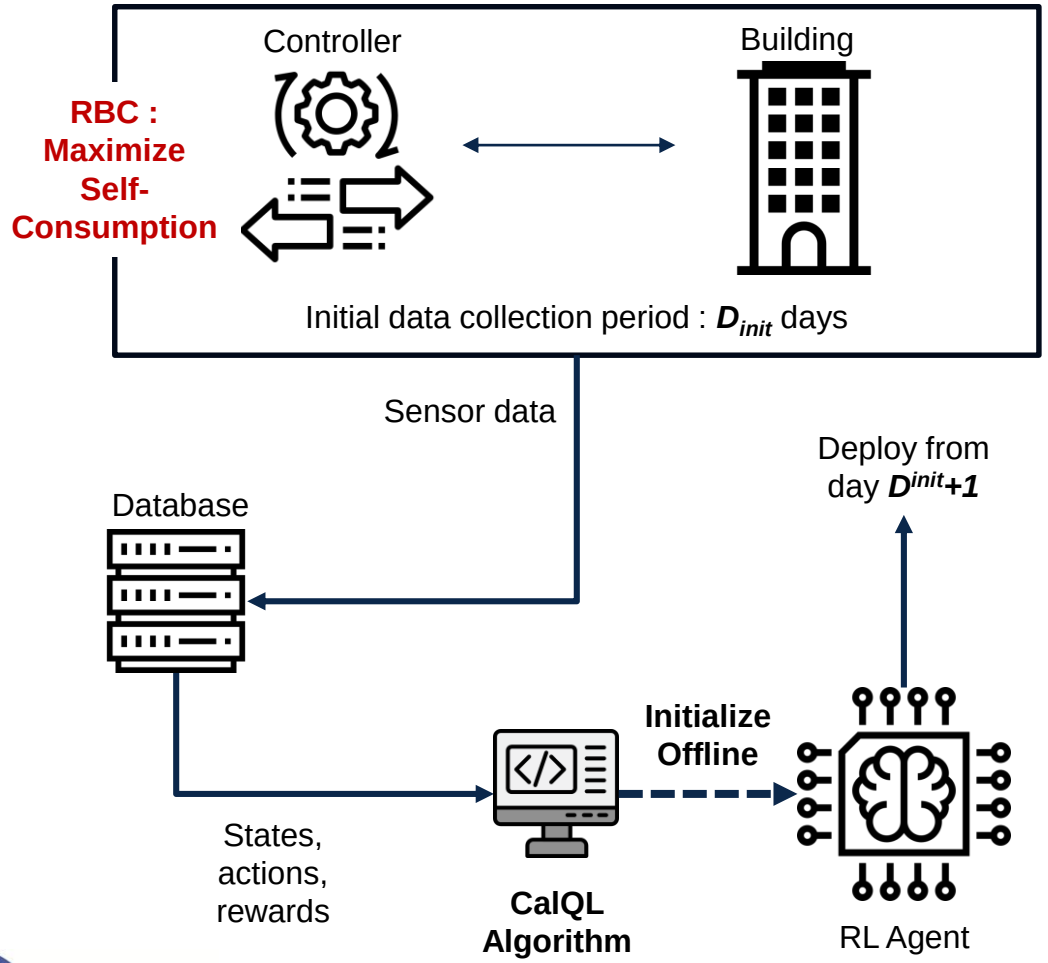
- Fully data-driven : sensor data in, control action out ➔ low cost
- Reward function : penalize electricity cost and thermal comfort deviation, encourage arbitrage
- Algorithm choice : CalQL ➔ suitable for offline learning and online adaptation



Nakamoto, M.; Zhai, Y.; Singh, A.; Mark, M. S.; Ma, Y.; Finn, C.; Kumar, A.; Levine, S. Cal-QL: Calibrated Offline RL Pre-Training for Efficient Online Fine-Tuning; 2023.



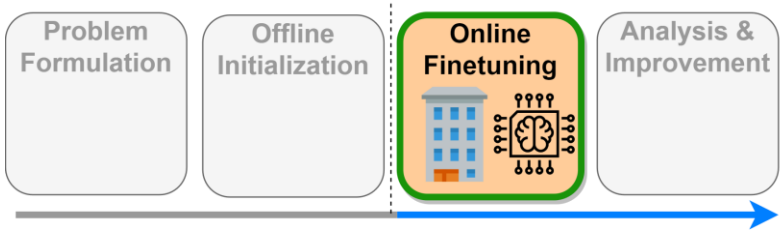
DATA-DRIVEN OFFLINE INITIALIZATION



- **Bootstrap learning agent**
  - Use existing controller
  - Experience collection included in test period
- **Collect data for D_{init} days**
- **Continue learning after deployment**
  - Model update every D_{upd} days

Kumar, S. R.; Easwaran, A.; Delinchant, B.; Rigo-Mariani, R. Behavioural Cloning Based RL Agents for District Energy Management. In *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*; ACM: New York, NY, USA, 2022; pp 466–470. https://doi.org/10.1145/3563357.3566165.

ACTION MASKING FOR SAFETY

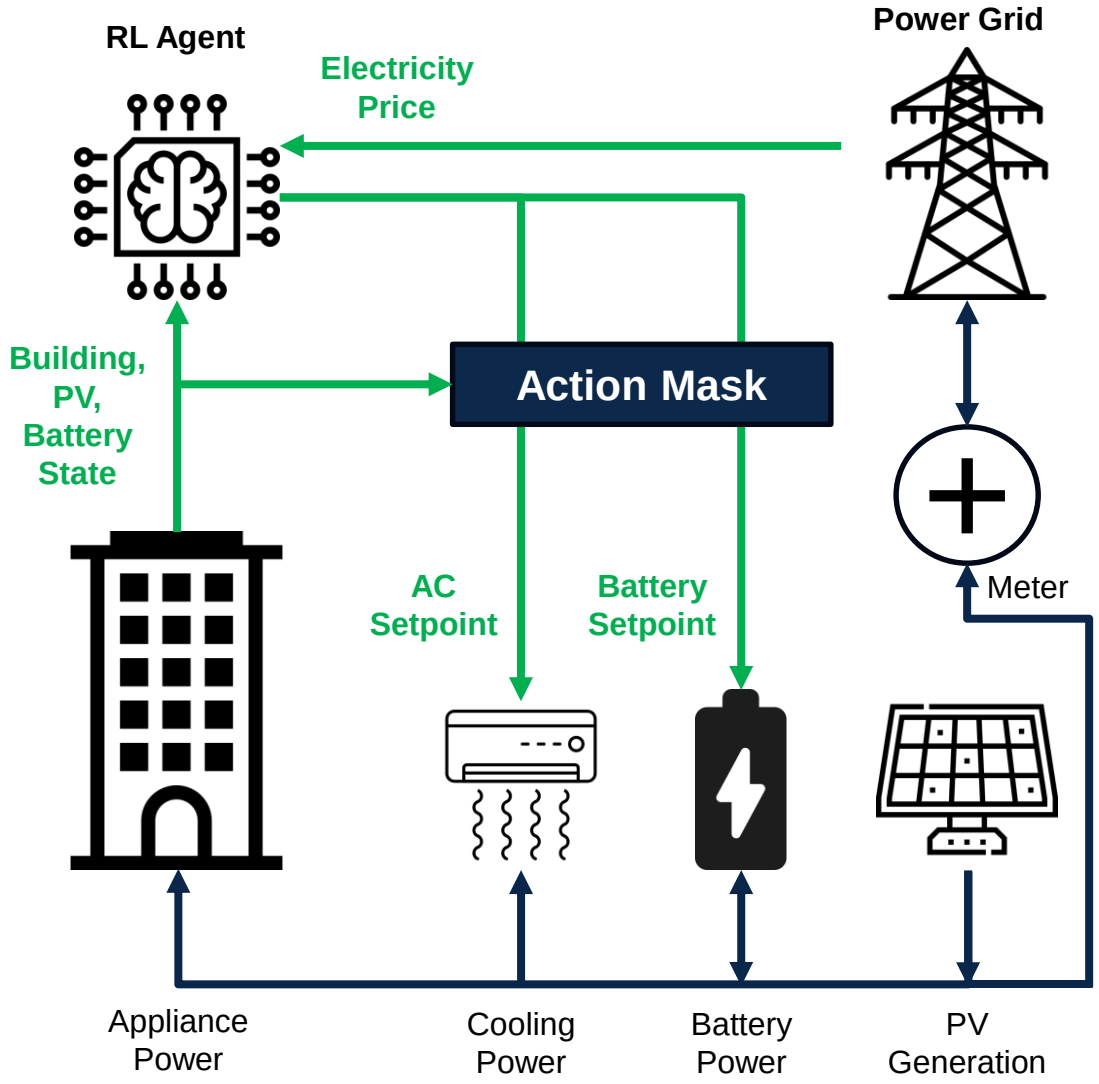


- **Safety rules**
  - Switch off AC outside working hours
  - Maintain occupied indoor temp 20ºC – 27ºC
  - Battery SOC always between 20% - 90%

Condition	Masked (Disallowed) Actions
Indoor Temperature $\geq 26.5^{o}$C During Working Hours	Switch OFF AC
Battery at Lower Limit	Discharge Battery
Battery at Upper Limit	Charge Battery

Rules encoded as state-dependent action mask

Invalid and unsafe actions avoided – agent explores permitted space



Kumar, S. R.; Rigo-Mariani, R.; Delinchant, B.; Easwaran, A. Towards Safe Model-Free Building Energy Management Using Masked Reinforcement Learning. In *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*; IEEE, 2023; pp 1–5. https://doi.org/10.1109/ISGTEUROPE56780.2023.10407781.

KEY RESULTS

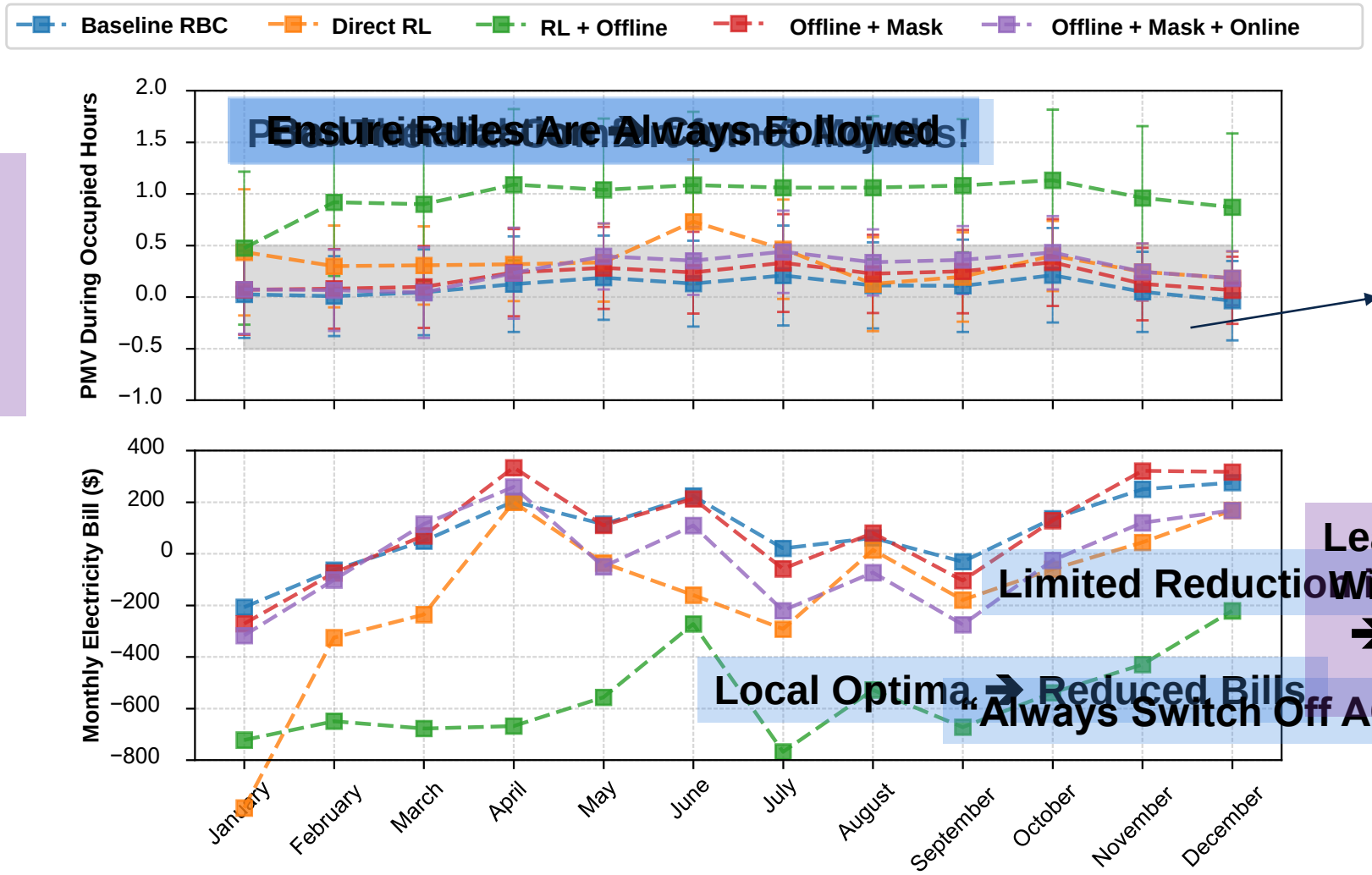
- ✓ 117% Bill Reduction
- ✓ 0.34 Points PMV Improvement
- ✓ Fully Data-Driven Methodology
- ✓ Respects Safety Rules

Challenge	Metric	Existing RBC	Direct RL	RL + Offline Training	RL + Offline + Action Mask
Control Objectives	Annual Electricity Bill ($)	1034.25	-2103.21	-463.28	-174.5
	Thermal Comfort (PMV)	-0.362 ± 0.335	0.25 ± 0.832	0.09 ± 0.474	**0.02 ± 0.421**
Learning Challenges	Training Time/Data Required (Weeks)	0	-	20 weeks	2 weeks
Controller Safety	Safety Rule Violations (Count)	21	2973	1318	395 (Thermal Inertia)
Cost and ROI	Time to 1000$ Savings vs RBC (Weeks)	-	20 weeks	40 weeks	48 weeks

- **Direct RL ➔ local optimum (poor thermal comfort when learning)**
- **Offline Training Only ➔ initial data requirement, no safety guarantees**
- **All techniques ➔ Improved control objectives, respect rules, address challenges**

A CLOSER LOOK : METRIC TRENDS

Always Follows Safety Rules ➔ Maintain Thermal Comfort



Comfortable PMV

Limited Reduction With Bill

Local Optima ➔ Reduced Bills

Learns by Exploring Within Safety Rules ➔ Better Savings Over Time

"Always Switch Off AC to Save Money"

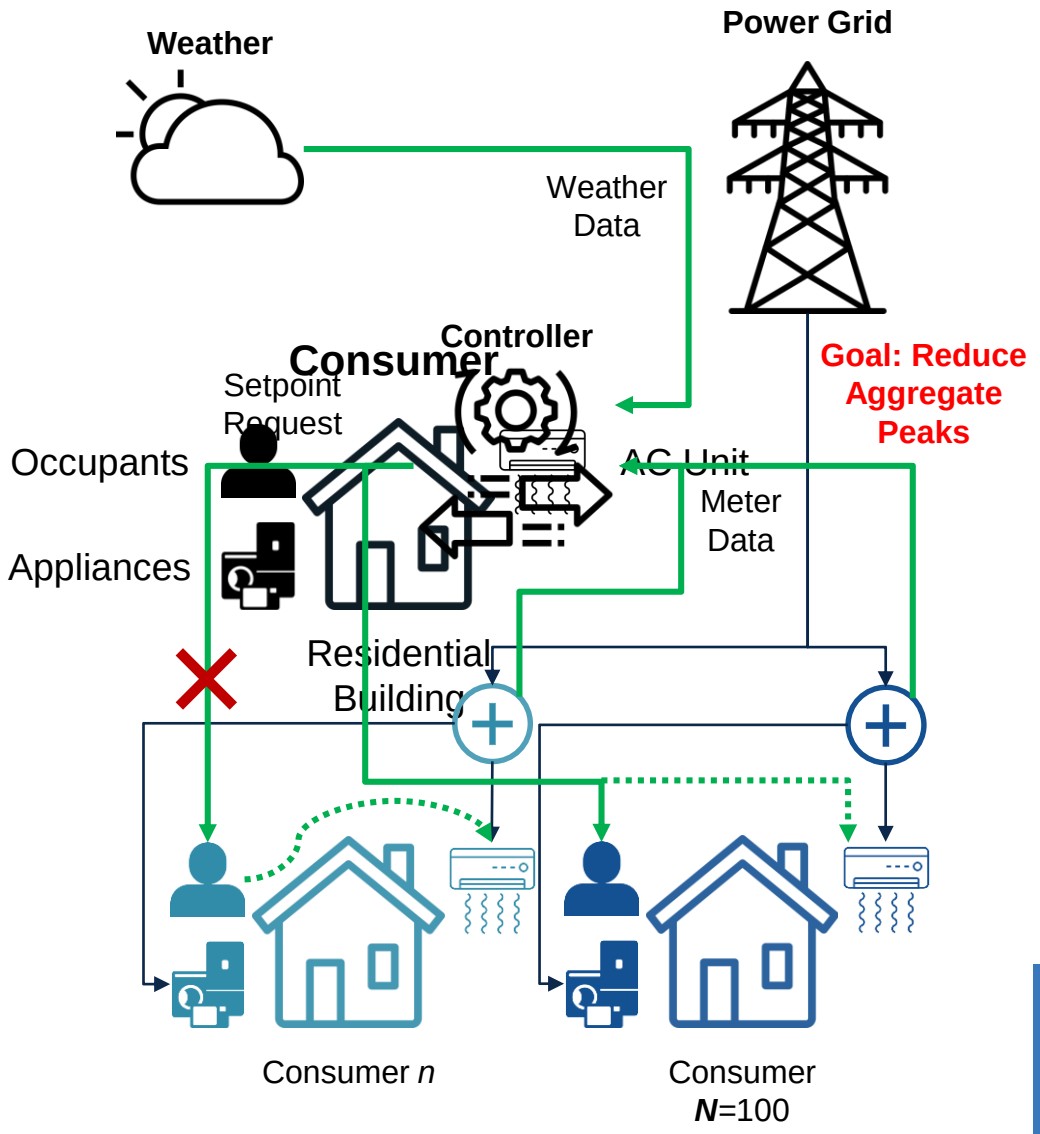
SECTION 4

APPLICATION TO A MULTI-BUILDING CONTROL PROBLEM

HOW CAN THE APPROACH BE USED TO IMPROVE DEMAND RESPONSE PROGRAMS?

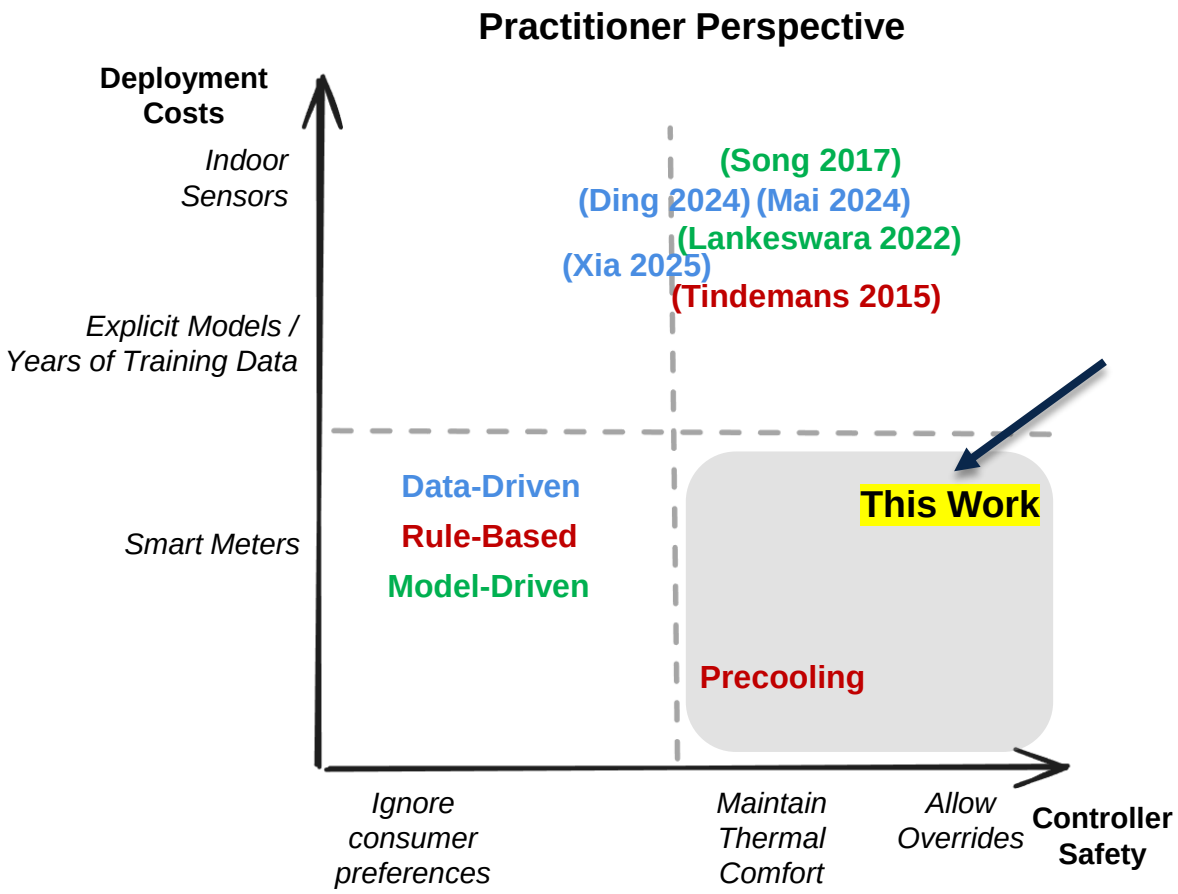
4 APPLICATION TO A MULTI-BUILDING DEMAND RESPONSE PROBLEM

- **Coordinate 100 residential consumers in Singapore**
  - **Goal** : Aggregate peak load reduction
  - **Scope of Control** : Air-conditioner setpoints (hourly)
- **Consumers can override aggregator signals**
  - Each consumer has a preferred indoor setpoint
  - Overrides based on local factors (thermal comfort)
- **Limited data available**
  - Only smart-meter data accessible
  - 2 months of historical data available (without DR)
  - Evaluation Period : 1 month



“Practical” Solution
Low cost
Safe

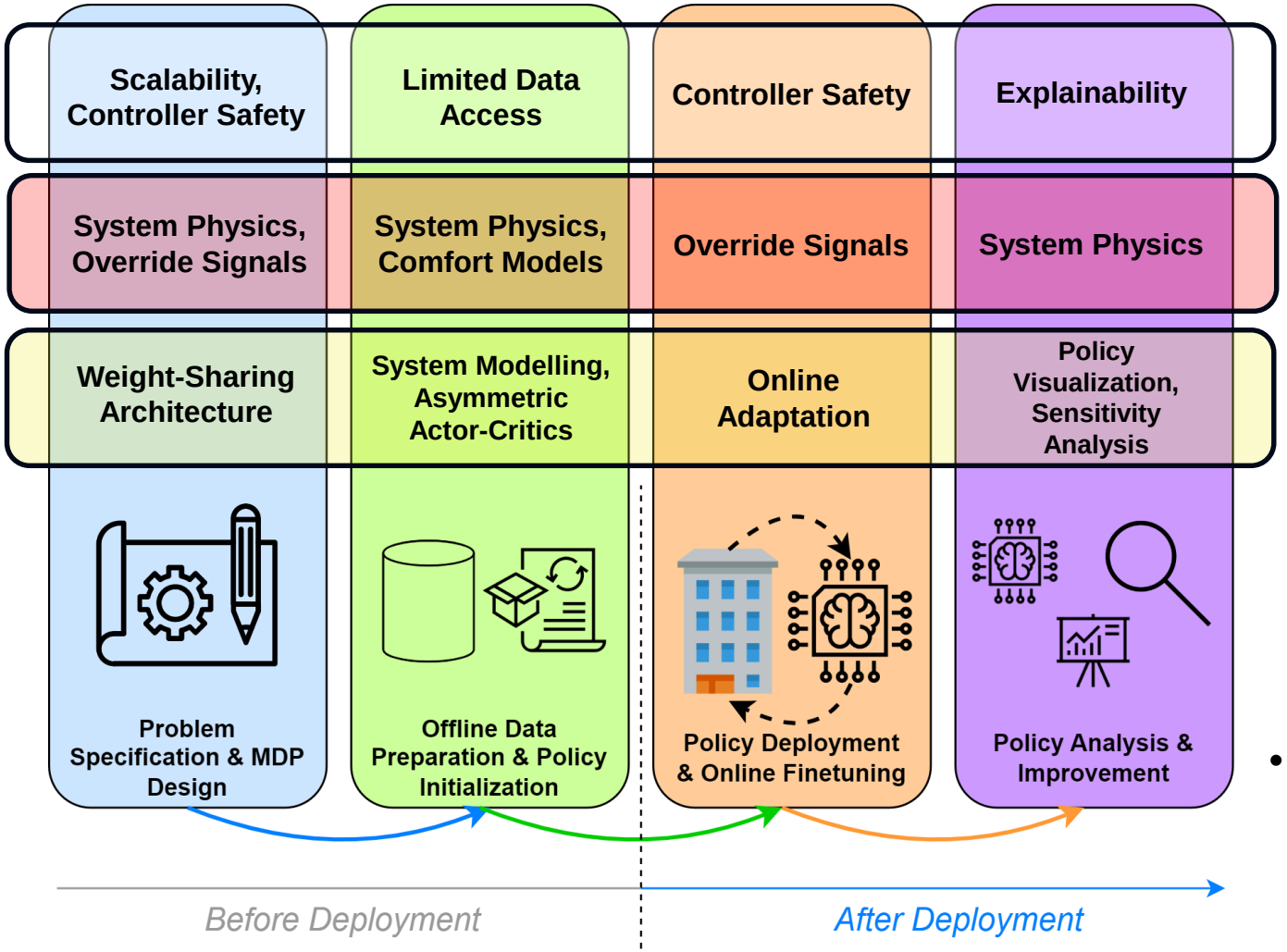
NOVELTY OF THIS WORK



- **Existing work requires indoor sensors (eg Indoor Temperature)**
  - Implementation costs, data privacy issues
- **Thermal comfort depends on several factors (Sarran 2021, ASHRAE-55)**
  - Ideal case ➔ allow consumers to override (Kaspar 2024)
- **Main Novelties**
  - ✓ AC setpoint control of 100 consumers
  - ✓ Only use smart meter data
  - ✓ Explicitly allow consumer overrides

Broberg, T.; Persson, L. Is Our Everyday Comfort for Sale? Preferences for Demand Management on the Electricity Market. *Energy Economics* **2016**, *54*, 24–32. https://doi.org/10.1016/j.eneco.2015.11.005.

KNOWLEDGE-INFORMED RL SOLUTION ARCHITECTURE

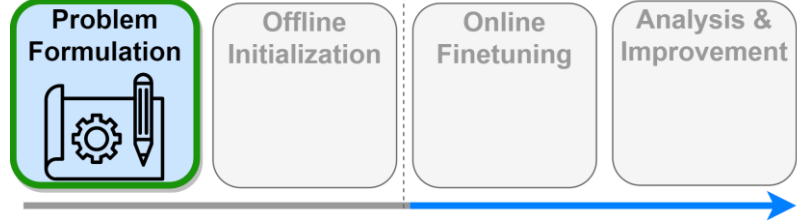


Domain Knowledge Use:

- Scalable solution architecture
  - Respect data limitations
- Design a simulator for RL training

Kumar, S. R.; Rigo-Mariani, R.; Delinchant, B.; Easwaran, A. Towards Safe Model-Free Building Energy Management Using Masked Reinforcement Learning. In *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*; IEEE, 2023; pp 1–5. https://doi.org/10.1109/ISGTEUROPE56780.2023.10407781.
Kumar, S. R.; Easwaran, A.; Delinchant, B.; Rigo-Mariani, R. Behavioural Cloning Based RL Agents for District Energy Management. In *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*; ACM: New York, NY, USA, 2022; pp 466–470. https://doi.org/10.1145/3563357.3566165.

PROBLEM FORMULATION STRATEGIES



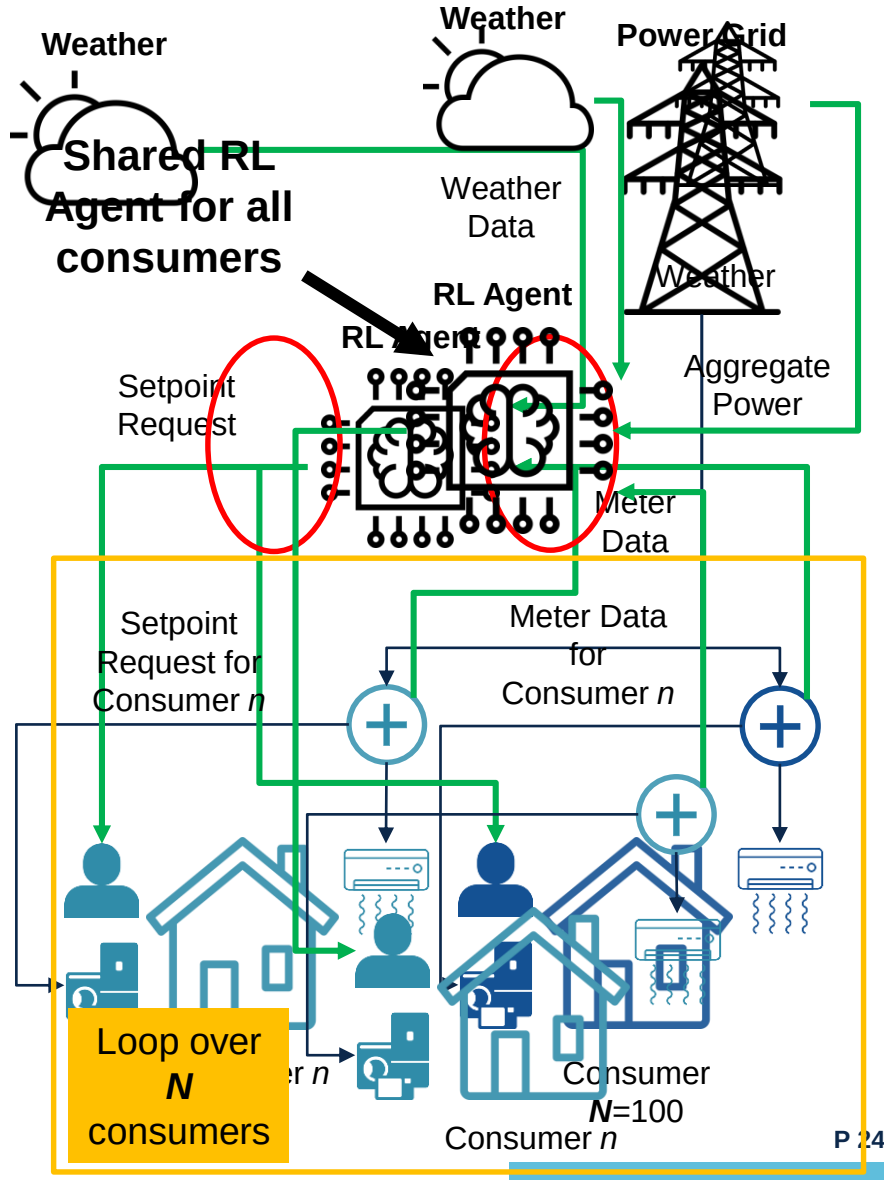
- **Design state, action and architecture of solution**
- **Main Considerations**

Centralized Architecture
N (=100) inputs and outputs
"Curse of Dimensionality"

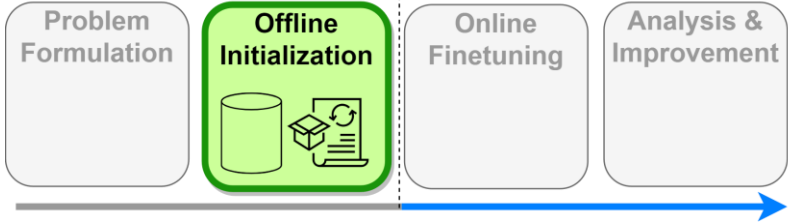
  - True state cannot be measured (eg $T^{in}_{n,t}$)
  - Only use front-of-meter signals in deployment
  - Scalable architecture required

- **Shared set of weights for all consumers**

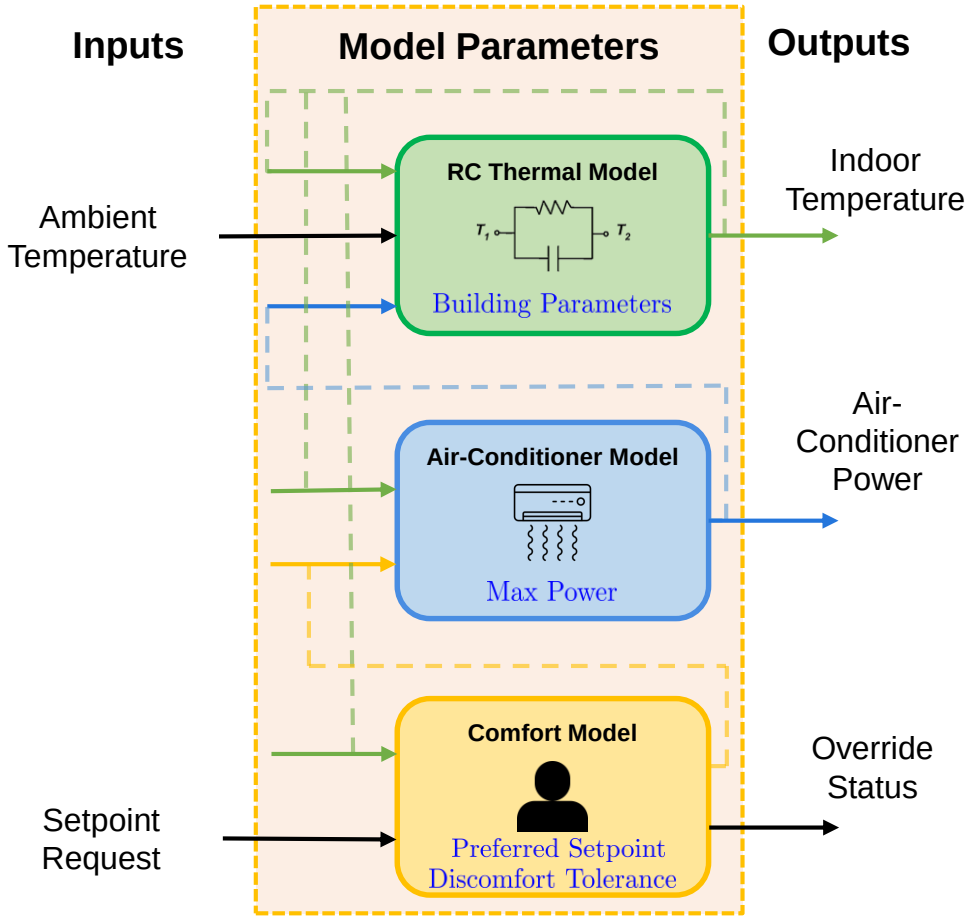
  - No dependence on ***N***
  - Coordination through reward function



Charbonnier, F.; Morstyn, T.; McCulloch, M. D. Scalable Multi-Agent Reinforcement Learning for Distributed Control of Residential Energy Flexibility. *Applied Energy* **2022**, *314*. https://doi.org/10.1016/j.apenergy.2022.118825.

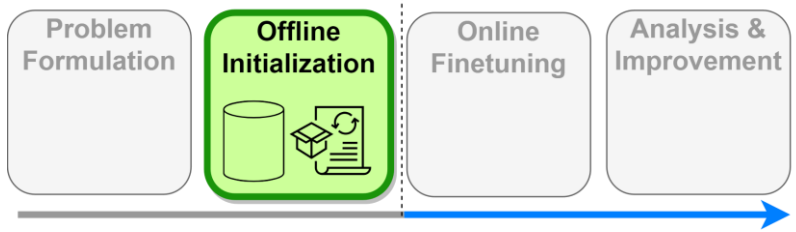


HYBRID SYSTEM MODELLING - APPROXIMATE CONSUMER MODELS (ACM)

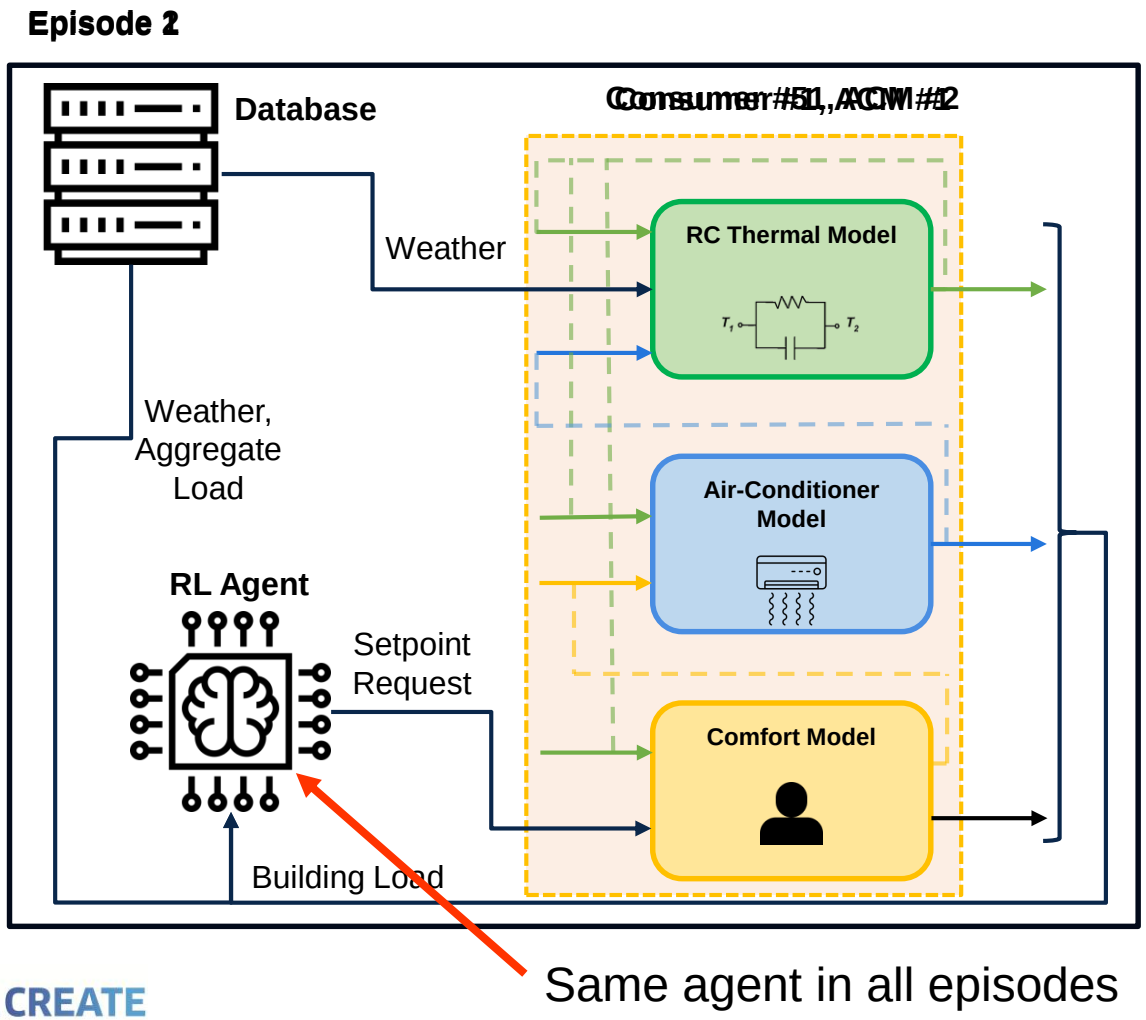


- **Simplified model of system dynamics**
  - Iterative predictions given weather and control signals
- **Fit to historical smart meter data**
- ***M* models for each consumer**
  - For different typical values of preferred setpoint
  - eg 22ºC, 24ºC, 26ºC

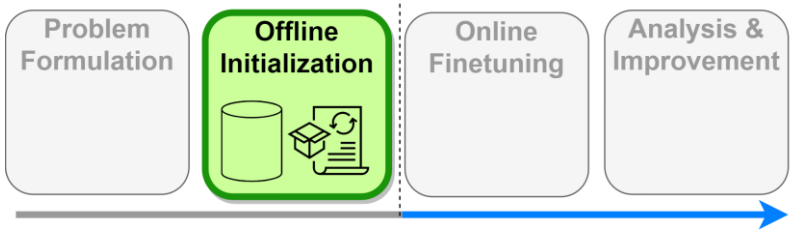
Ryu, J.; Kim, J.; Hong, W.; De Dear, R. Quantifying Householder Tolerance of Thermal Discomfort before Turning on Air-Conditioner. *Energy and Buildings* **2020**, *211*, 109797. https://doi.org/10.1016/j.enbuild.2020.109797.



TRAINING PIPELINE USING APPROXIMATE CONSUMER MODELS

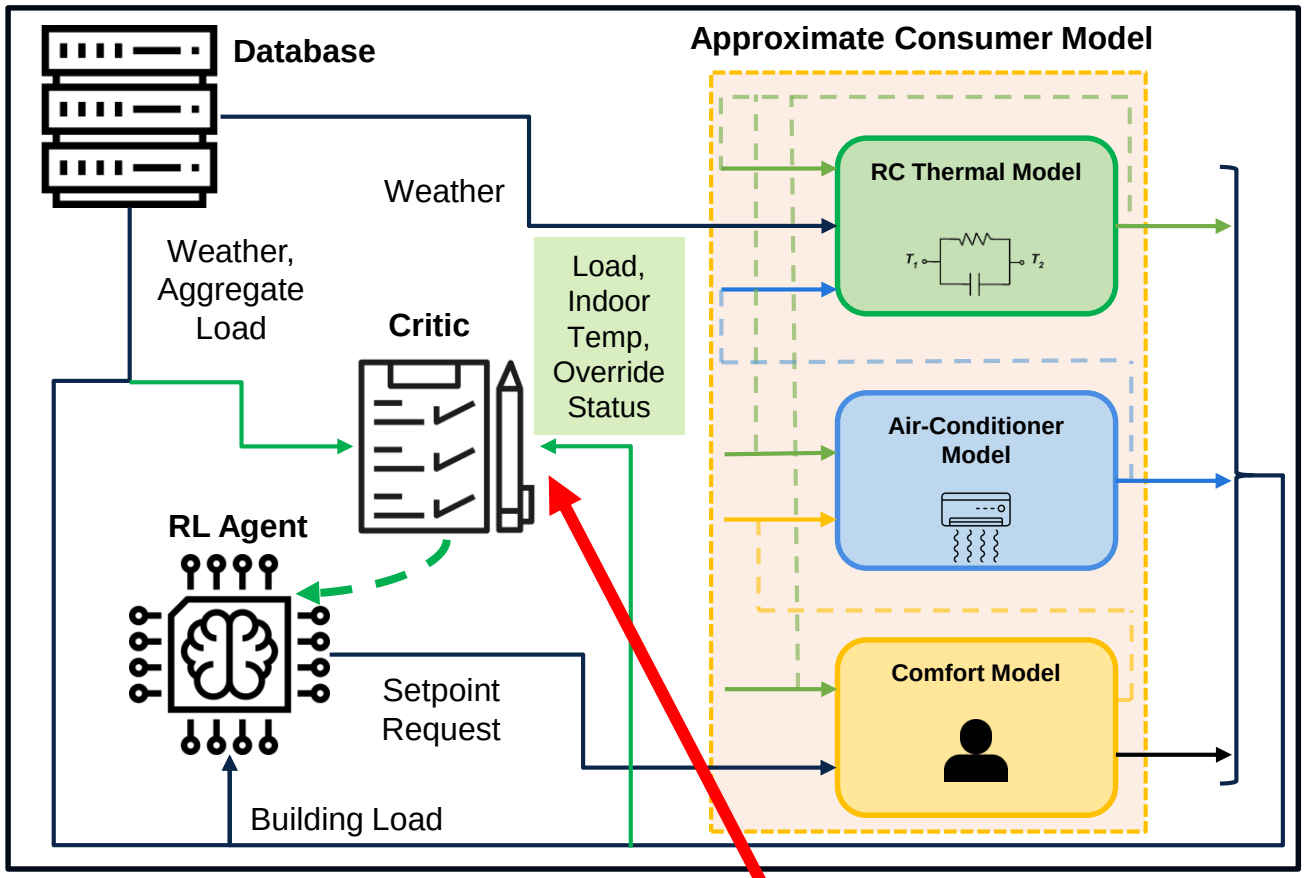


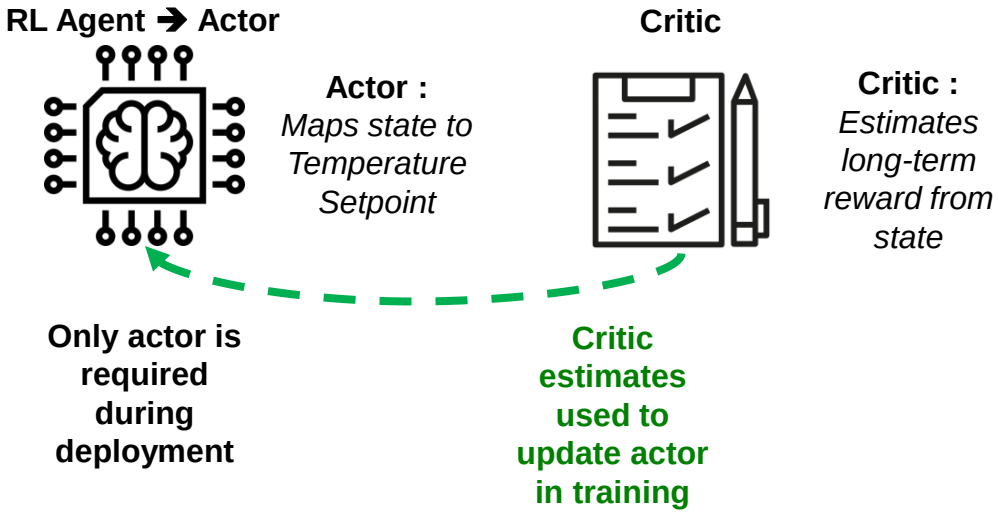
- **Training in simulation with ACMs**
- **Episodic training**
  - 1 episode ➔ 1 day (24 time steps)
  - Change ACM after each episode
  - Total training : 2000 episodes (~5.5 simulated years)
- **Goal :**
  - Make one agent robust to all ACMs



TRAINING PIPELINE USING APPROXIMATE CONSUMER MODELS

Training Episode





- **Use "hidden" inputs in training**
  - Model-generated outputs to critic only
  - "Split-Input" actor-critic architecture

Model-estimated inputs to critic only
(sensors are not available in real system)

ALTERNATE STRATEGIES FOR COMPARISON

Agent	Strategy	Explanation
Baseline	Consumers always use their preferred setpoints	No Control
Precooling	Cool all buildings prior to peaks	Rule-based Heuristic Strategy (commonly used)
Centralized	Single large agent with ***N*** outputs	Directly train 1 large agent, only on meter data
Full-Input	All inputs available, with shared weights	Assume sensors exist for indoor measurements
Split-Input	Split-Input architecture with shared weights	Proposed Solution

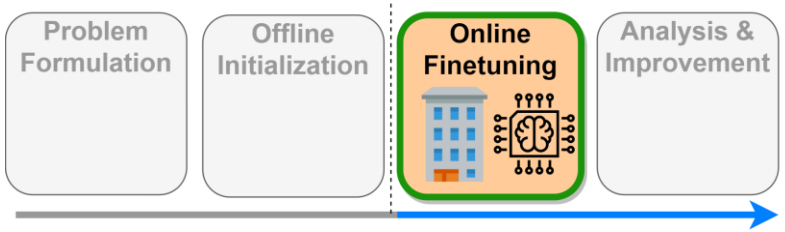
Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. arXiv August 28, 2017. http://arxiv.org/abs/1707.06347 (accessed 2024-10-15).

KEY RESULTS

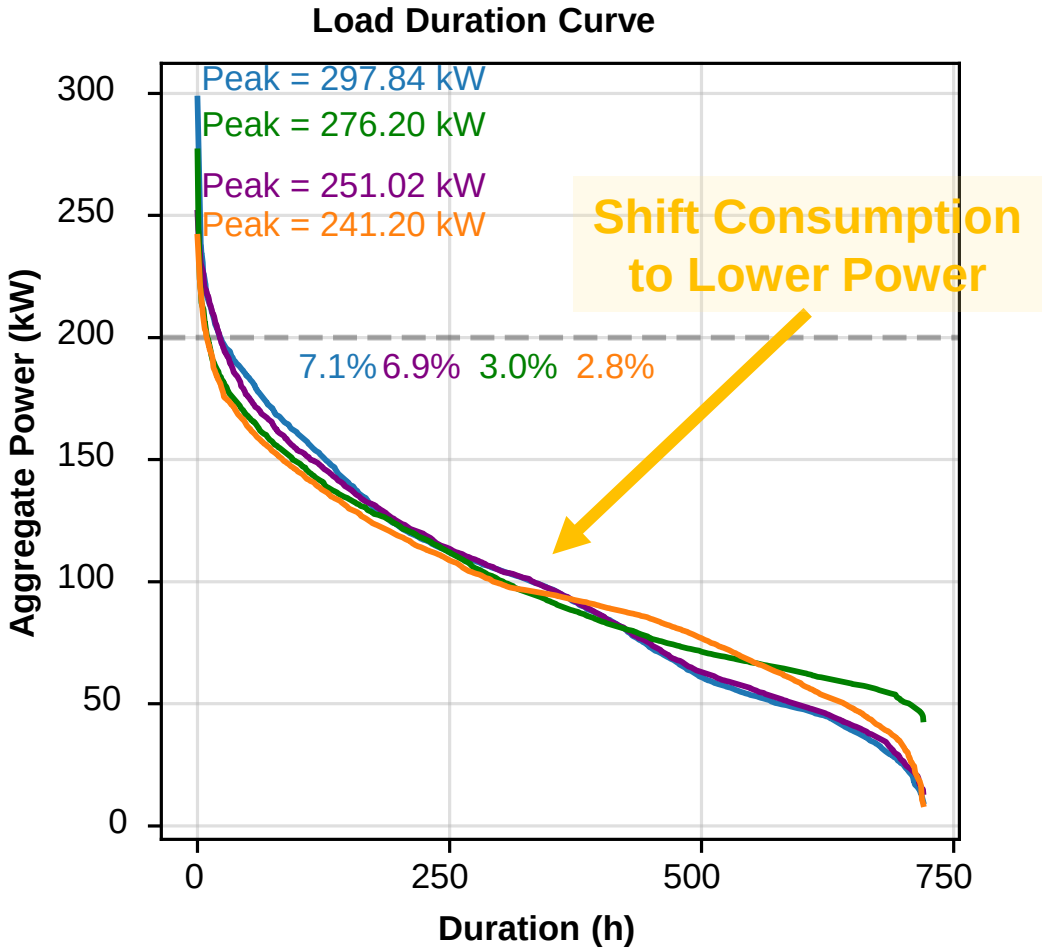
- ✓ 59% Peak Load Reduction
- ✓ Allows Consumer Overrides
- ✓ Only Smart-Meter Data Needed
- ✓ Scalable Architecture

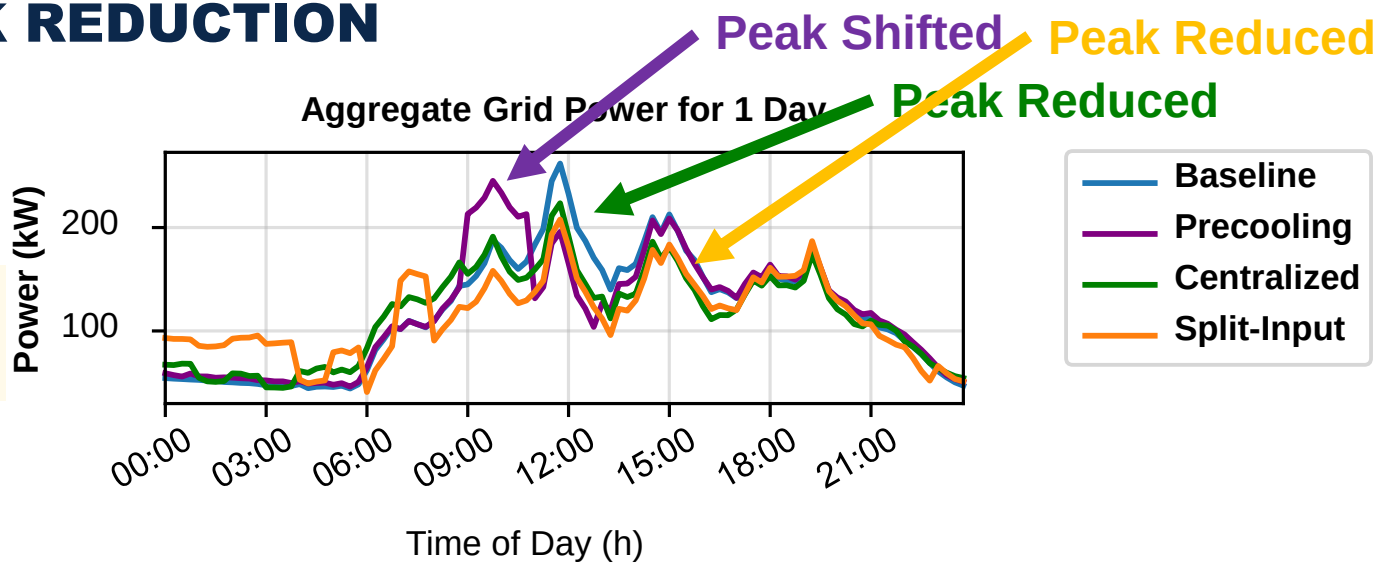
Challenge	Metric	No Control (Baseline)	Precooling (Heuristic)	Centralized PPO	Full-Input PPO	Split-Input PPO
Control Objectives	Energy At Peak Loads (MWh)	5.002	4.843	2.139	**1.830**	2.051
Scalability	Solution Complexity (dependence on *N)*	Independent	Independent	Input, Output Dimension ∝ N	Independent	Independent
Limited Data Access	Needs Indoor Sensors? (Yes/No)	No	No	No	Yes	No
Controller Safety	Fraction of Overrides	-	0.019	0.104	0.174	0.189

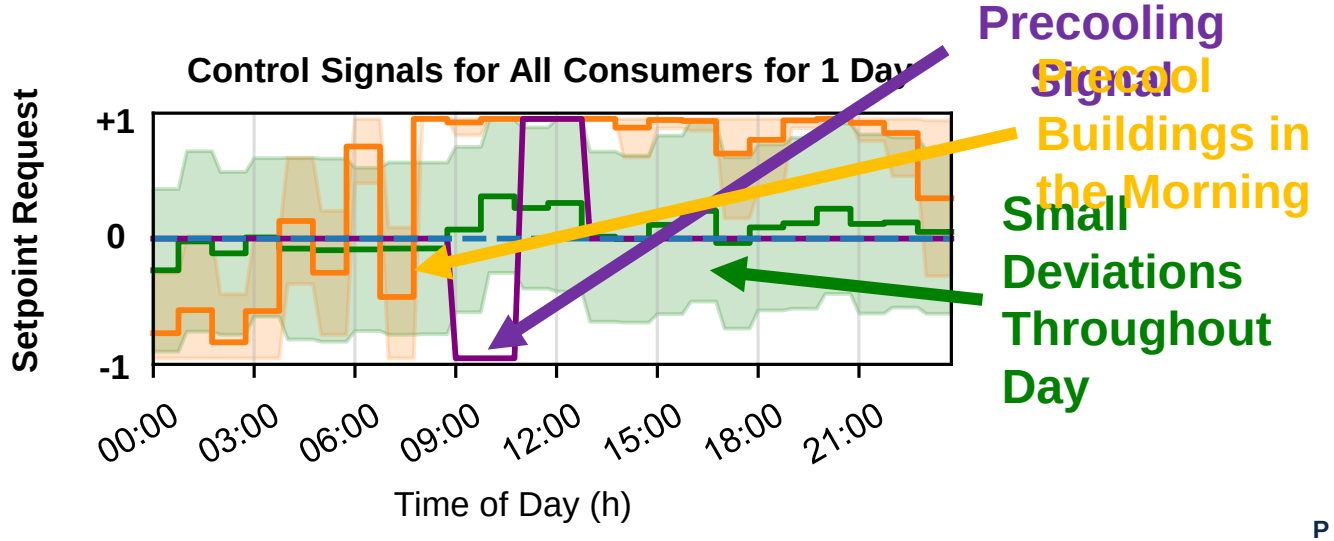
- **Each technique provides advantages**
- **Shared weights ➔ benefits of centralized PPO, no dependence on *N***
- **Asymmetric Actor-Critic ➔ peak reduction without indoor sensors**
- **In combination ➔ meet control objectives + real-world challenges**

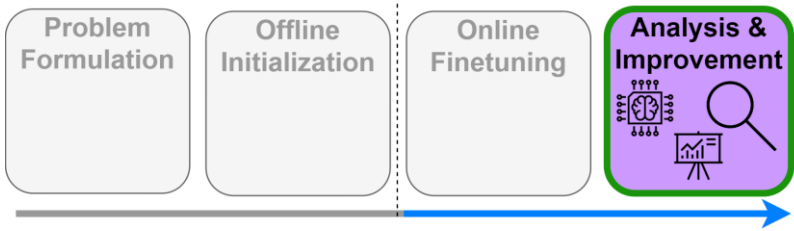


A CLOSER LOOK : PEAK REDUCTION



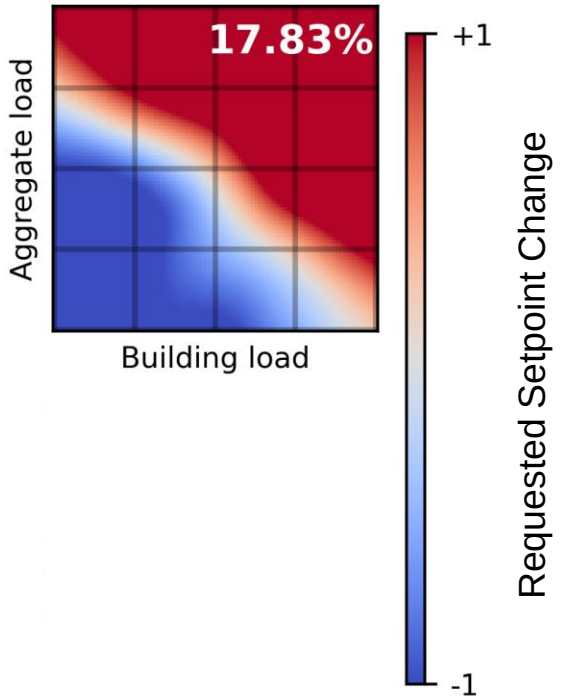






IMPACT OF SPLIT-INPUT ARCHITECTURE

RL Agent Policy Space



Without Split-Input

- **Generalist policy**
  - Request bigger deviation as building and aggregate load increases
- **Effects of asymmetric actor-critic**
  - Sharper decision boundaries
  - Weather dependence accounts for overrides

SECTION 5

CONCLUSIONS AND PERSPECTIVES

5

KEY CONTRIBUTIONS OF THE THESIS

- **A systematic framework for knowledge integration in RL**
  - **Knowledge-Informed RL for Building Energy Management**
  - Taxonomy of domain knowledge representations (eg Physics, Models, Floorplans ..)
- **Specific techniques to incorporate domain knowledge and meet challenges**
  - Maps knowledge representations to domain-specific challenges
- **Case studies in relevant application scenarios**
  - Single-building participant in DR (ie, the consumer's perspective)
  - Multi-building coordination for DR (ie, the aggregator's perspective)

CONCLUSIONS AND FUTURE WORK

- **Knowledge-Informed RL can address specific challenges for BEM**
  - Practitioner-friendly framework
  - Provides benefits for scalability, safety, learning challenges, infrastructure
  - Demonstrated in typical case-studies
- **Limitations and future work**
  - Translate from simulation to real-world trials (work ongoing)
  - Comparison with model-predictive control approaches

LIST OF PUBLICATIONS

1. S. Ram Kumar, A. Easwaran, B. Delinchant, and R. Rigo-Mariani, “Deep Reinforcement Learning for Coordinated Air-Conditioner Control in Groups of Buildings Using Smart Meter Data” – *Journal Submission, Under Review*, Jul 2025.

2. S. Ram Kumar, A. Easwaran, B. Delinchant, and R. Rigo-Mariani, “Improving Demand Response Programs Using Override Signals with Reinforcement Learning,” in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, in E-Energy ’25. New York, NY, USA: Association for Computing Machinery, Jun. 2025, pp. 603–611. doi: 10.1145/3679240.3734657.

3. S. R. Kumar, A. Easwaran, B. Delinchant, and R. Rigo-Mariani, “Real-time Retail Electricity Pricing Using Offline Reinforcement Learning,” in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, in e-Energy ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 454–458. doi: 10.1145/3632775.3661964.

4. S. R. Kumar, R. Rigo-Mariani, B. Delinchant, and A. Easwaran, “Towards Safe Model-Free Building Energy Management using Masked Reinforcement Learning,” in *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/ISGTEUROPE56780.2023.10407781.

5. S. R. Kumar, A. Easwaran, B. Delinchant, and R. Rigo-Mariani, “Behavioural cloning based RL agents for district energy management,” in *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York, NY, USA: ACM, Nov. 2022, pp. 466–470. doi: 10.1145/3563357.3566165.

APPENDICES

REWARD FUNCTION DEFINITION – SINGLE BUILDING

$$f^{scale}(x, x^-, x^+) = \text{clip}_{[-1,0]}[\frac{x - x^+}{x^+ - x^-}]$$

$P_t^{bess,arb}$: Battery power for arbitrage

C_1, C_2, C_3 : Scaling constants

Two formulations with different styles of domain knowledge integration

$$R^A(b_t, T_t^{in}) = C_1 \cdot f^{scale}(b_t, 50, -50) + \mathbf{I}^{occupied} \cdot C_2 \cdot f^{scale}(|T_t^{in} - 24.0|, 3, 0)$$

- Trade-off cost and comfort
- Assumes 24°C is comfortable

$$R^B(b_t, T_t^{in}, \text{RH}_t^{in}, \lambda_t, P_t^{bess}) = C_1 \cdot f^{scale}(b_t, 50, -50) + \mathbf{I}^{occupied} \cdot C_2 \cdot f^{scale}(|\text{PMV}|, 1, 0) + C_3 \cdot f^{scale}(|P_t^{bess} - P_t^{bess,arb}|, 5, 0)$$

- Directly encourage arbitrage
- Use PMV in thermal comfort term

REWARD FUNCTION DESIGN

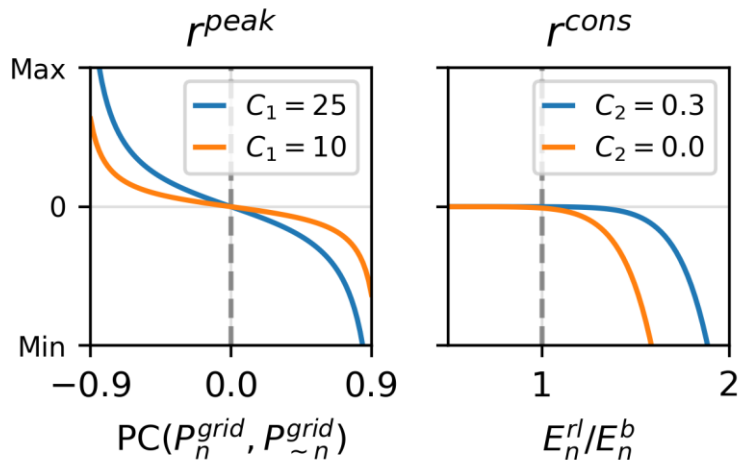
Peak reduction ➔ minimize **correlation** between consumption of building ***n*** and the rest of the cluster

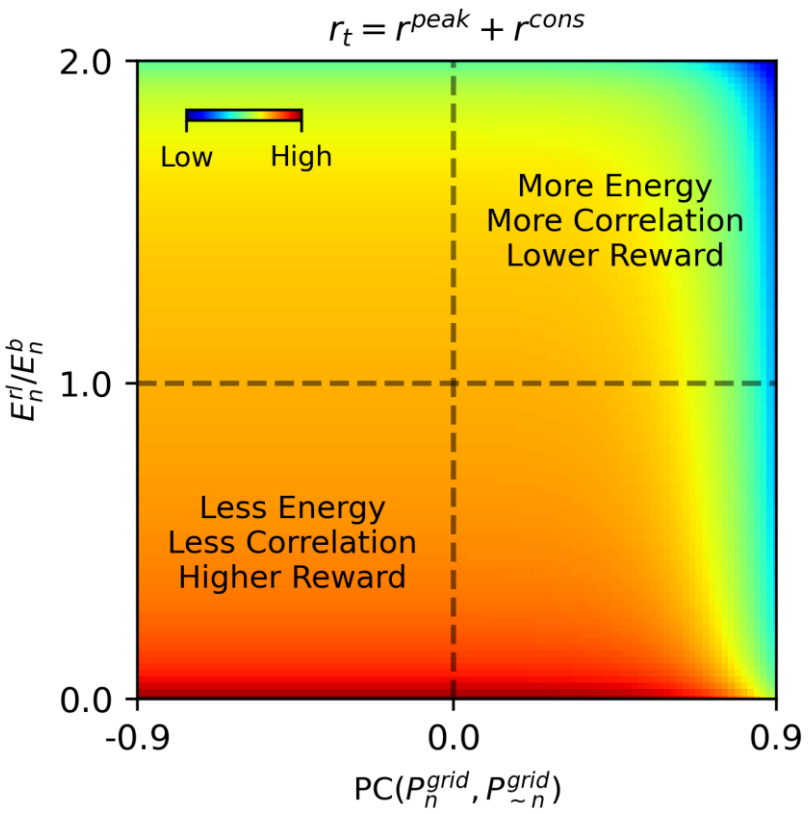
$$r^{peak} = C_1 \tan\left(\frac{-\pi \cdot \mathbf{PC}(P^{grid}_{n,t-1:t-24}, P^{grid}_{\sim n,t-1:t-24})}{2}\right)$$

Energy term ➔ do not consume more energy than baseline case (ie, without DR)

$$r^{cons} = -\left[\frac{E^{rl}_n}{E^b_n} - C_2\right]^{10}$$

$$r_t = \begin{cases} r^{peak} + r^{cons}, & \text{end of training episode} \\ 0, & \text{otherwise} \end{cases}$$

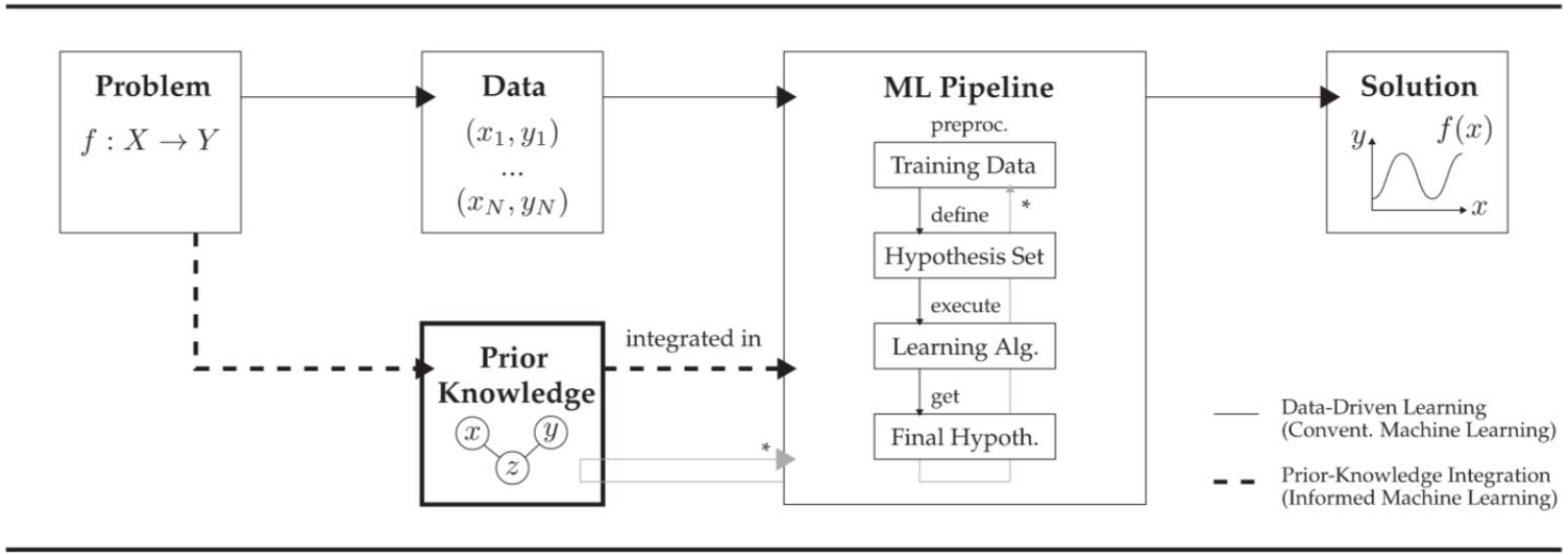




PC: Pearson coefficient
E^{rl}_n: Energy under RL policy, E^b_n: Energy without DR

2

KNOWLEDGE-INFORMED REINFORCEMENT LEARNING

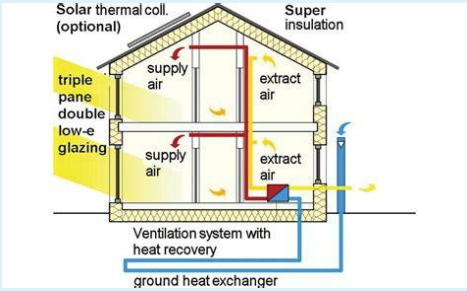


- **Key Contribution : Extension of "Informed Machine Learning" (von Reuden, 2023) to RL for BEM**

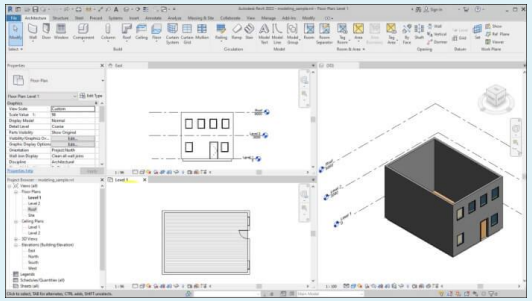
  - How is domain knowledge **represented** in the area of interest?
  - What **methods** can be used to integrate it in the RL pipeline?
  - Which **domain-specific challenges** can be addressed?

von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Pfrommer, J.; Pick, A.; Ramamurthy, R.; Walczak, M.; Garcke, J.; Bauckhage, C.; Schuecker, J. Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. *IEEE Transactions on Knowledge and Data Engineering* **2023**, *35* (1), 614–633. https://doi.org/10.1109/TKDE.2021.3079836.

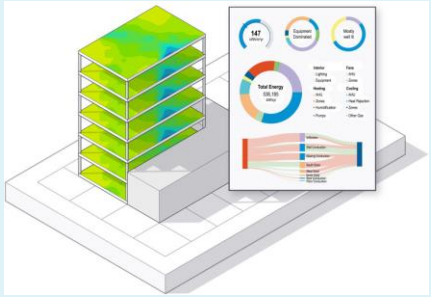
DOMAIN KNOWLEDGE REPRESENTATIONS IN BEM



Building and Device Physics



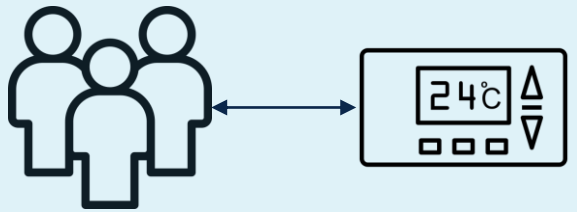
Semantic Models (Floorplan, BIM)



Computational Models (EnergyPlus)

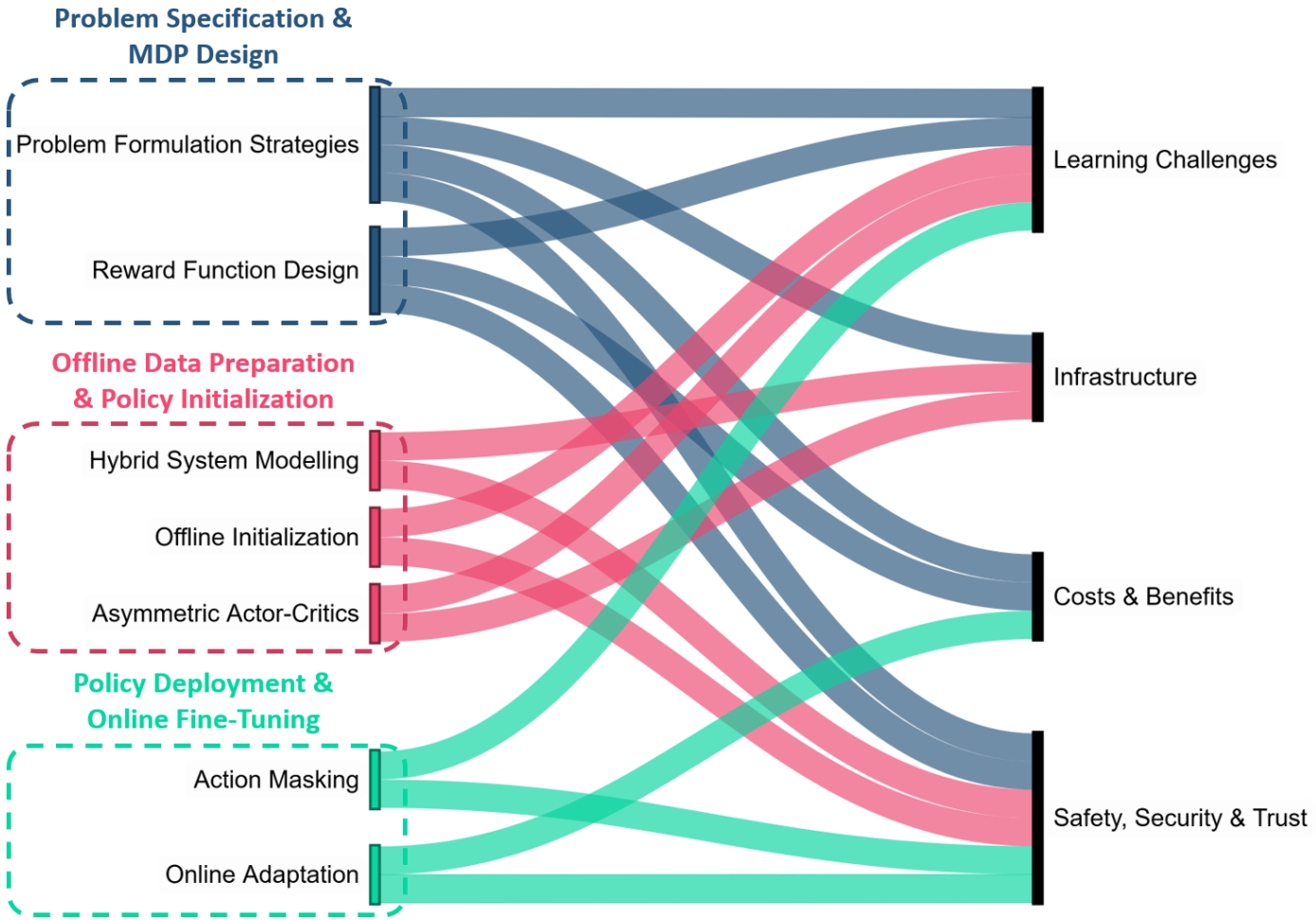
System Constraints

Existing Operational Strategies



User Interaction and Feedback

“TECHNIQUES” TO INCORPORATE KNOWLEDGE REPRESENTATIONS



- **Each technique**
  - Requires specific knowledge representations
  - Is applied at specific points in the workflow
  - Addresses one or more challenges

“Which knowledge representations do I need to solve a given challenge?”

“Which techniques can I use with the knowledge representations I have?”

COMPARISON TO PHYSICS-INFORMED REINFORCEMENT LEARNING

Knowledge-Informed RL

Which knowledge,
How to integrate,
Which challenge?

Physics-Informed RL
How to integrate **physics** into RL?

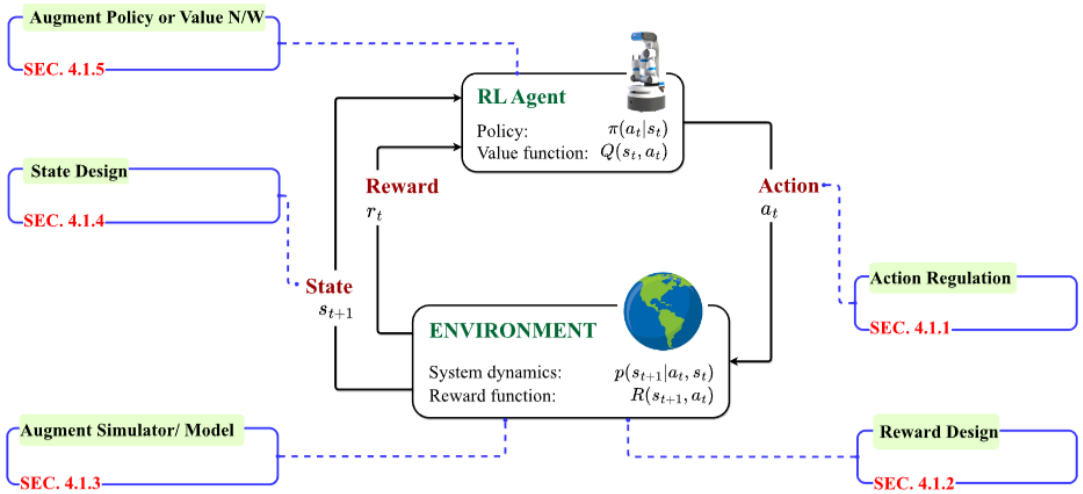


Fig. 4. Map of physics incorporation (PI) in the conventional Reinforcement Learning (RL) framework.

- **System Physics is one type of domain knowledge**
  - Relevant across many domains
- **Knowledge-Informed RL ➔ generalization**
  - Other domain knowledge sources (eg Human Feedback)
  - Challenge-focused
  - Domain specific

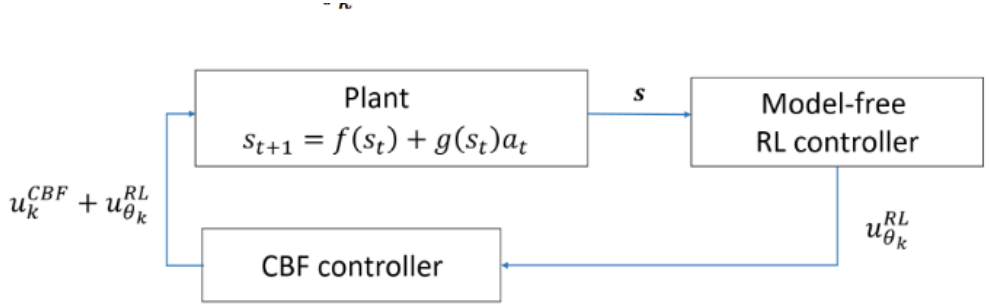
Banerjee, C.; Nguyen, K.; Fookes, C.; Raissi, M. A Survey on Physics Informed Reinforcement Learning: Review and Open Problems. SSRN 2023. https://doi.org/10.2139/ssrn.4597487.

CONSTRAINTS IN REINFORCEMENT LEARNING

and the reinforcement learning problem in a CMDP is

$$\pi^* = \arg\max_{\pi \in \Pi_C} J(\pi).$$

Feasible set



- **Model-based approaches**
  - Post-hoc modifications ➔ solve QP based on model to adjust the action
- **Model-free approaches**
  - Penalty terms in reward
  - Primal-dual decomposition ➔ requires some exploration, can provide guarantees
  - **Action Masking** ➔ based on implicit understanding of system

Achiam, J.; Held, D.; Tamar, A.; Abbeel, P. Constrained Policy Optimization. **2017**.

EXISTING WORK WITHIN KNOWLEDGE-INFORMED FRAMEWORK : EXAMPLE

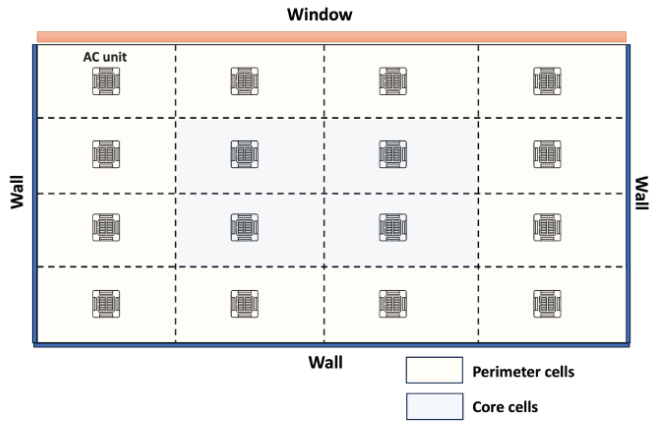


Figure 1: Schematic of an open-plan layout

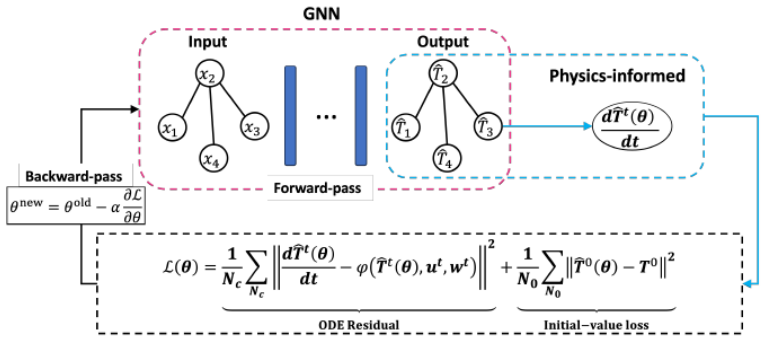
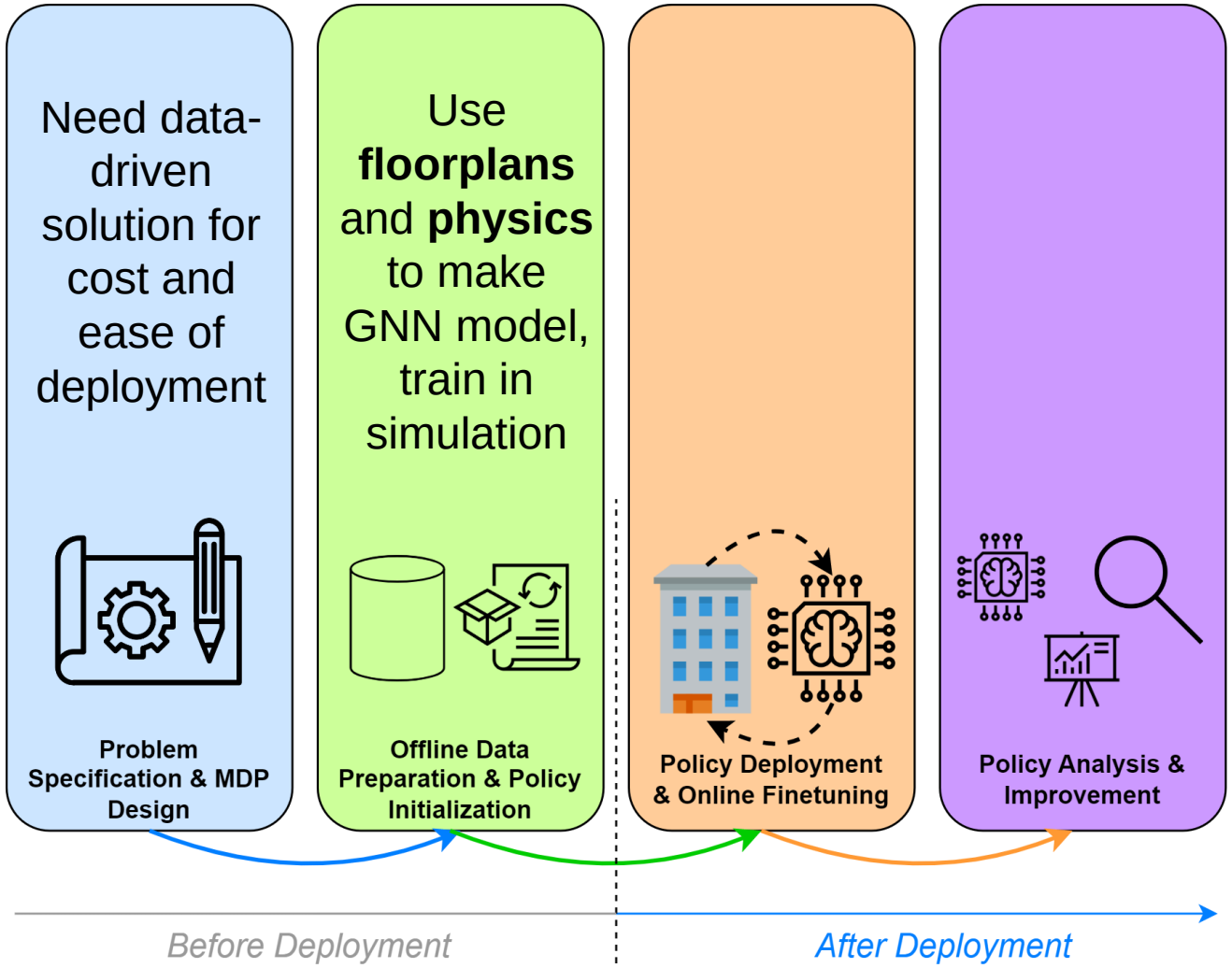


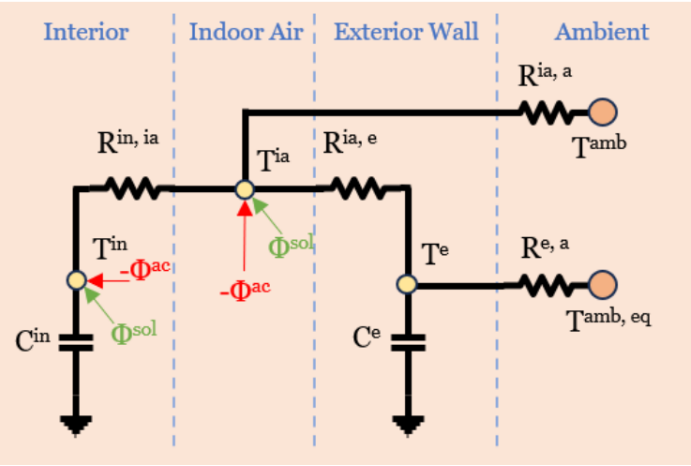
Figure 3: Schematic of PI-GNN architecture



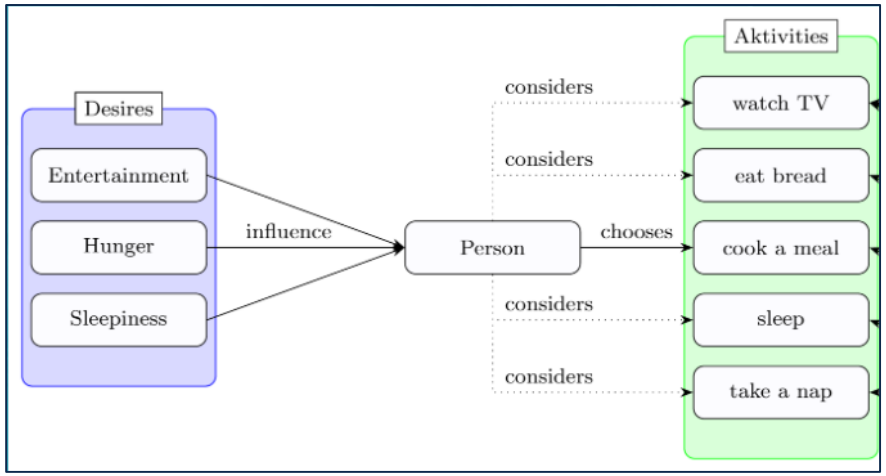
Nagarathinam, S.; Vasan, A. PhyGICS – A Physics-Informed Graph Neural Network-Based Intelligent HVAC Controller for Open-Plan Spaces. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*; e-Energy '24; Association for Computing Machinery: New York, NY, USA, 2024; pp 203–214. https://doi.org/10.1145/3632775.3661957.

SIMULATED CASE-STUDY – MULTI-BUILDING CASE

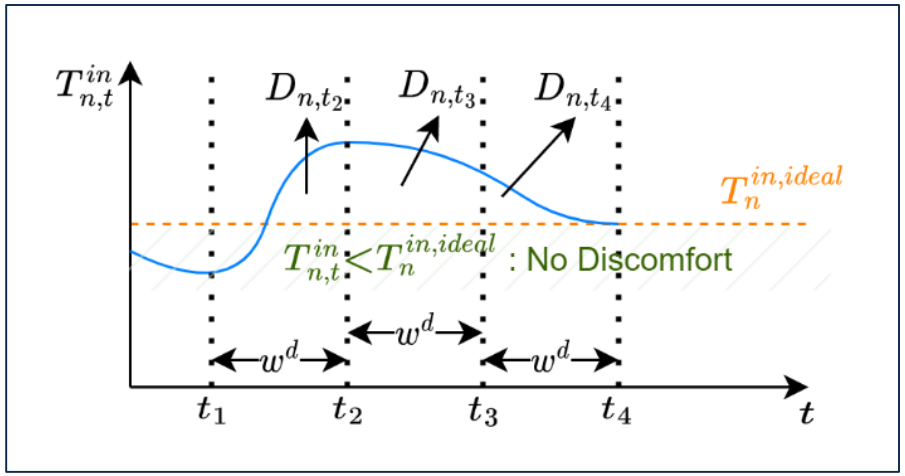
Simulation of 100 consumers in tropical weather conditions – model of consumers' buildings and behaviours



Thermal characteristics of consumer buildings using reduced-order models (4R2C)



Appliance loads (non-controllable) using behaviour simulation (LoadProfileGenerator)



$$o_{n,t} = \begin{cases} 1 & \text{if } D_{n,t} > D^{max}_n \text{ and } T^{set,req}_{n,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$

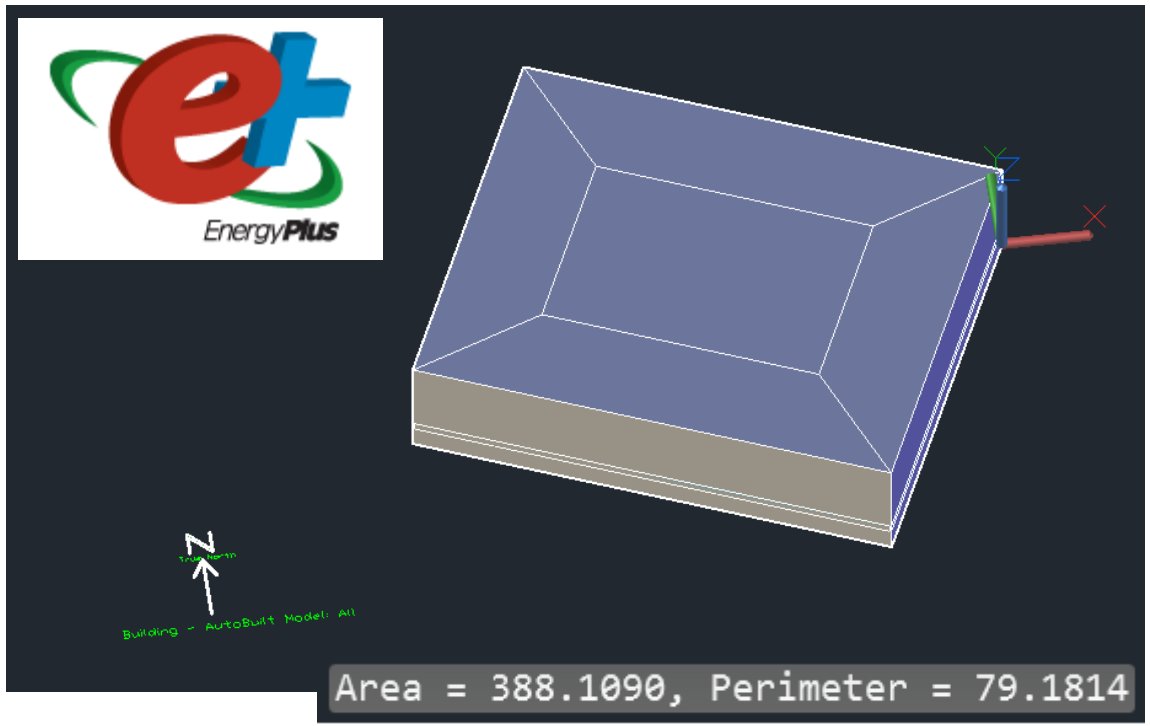
Override behaviour based on cumulative discomfort over a window of time

Harb, H.; Boyanov, N.; Hernandez, L.; Streblow, R.; Müller, D. Development and Validation of Grey-Box Models for Forecasting the Thermal Response of Occupied Buildings. *Energy and Buildings* **2016**, *117*, 199–207. https://doi.org/10.1016/j.enbuild.2016.02.021.

Pflugradt, N.; Stenzel, P.; Kotzur, L.; Stolten, D. LoadProfileGenerator: An Agent-Based BehaviorSimulation for Generating Residential Load Profiles. *JOSS* **2022**, *7* (71), 3574. https://doi.org/10.21105/joss.03574.

Ryu, J.; Kim, J.; Hong, W.; De Dear, R. Quantifying Householder Tolerance of Thermal Discomfort before Turning on Air-Conditioner. *Energy and Buildings* **2020**, *211*, 109797. https://doi.org/10.1016/j.enbuild.2020.109797.

SIMULATED CASE-STUDY – SINGLE-BUILDING CASE



- **Standard reference building in EnergyPlus simulation tool**
  - "Shop With PV And Battery" Example
  - Designed and validated by US DoE team
  - All components simulated in E+ with detailed analytical models
  - Incl shading for PV, lead-acid battery dynamics
  - Weather : typical weather data for Singapore
- **Used as a replacement for real-building**
  - E+ model never used for offline training etc.

COMPARISON TO MODEL-BASED APPROACHES

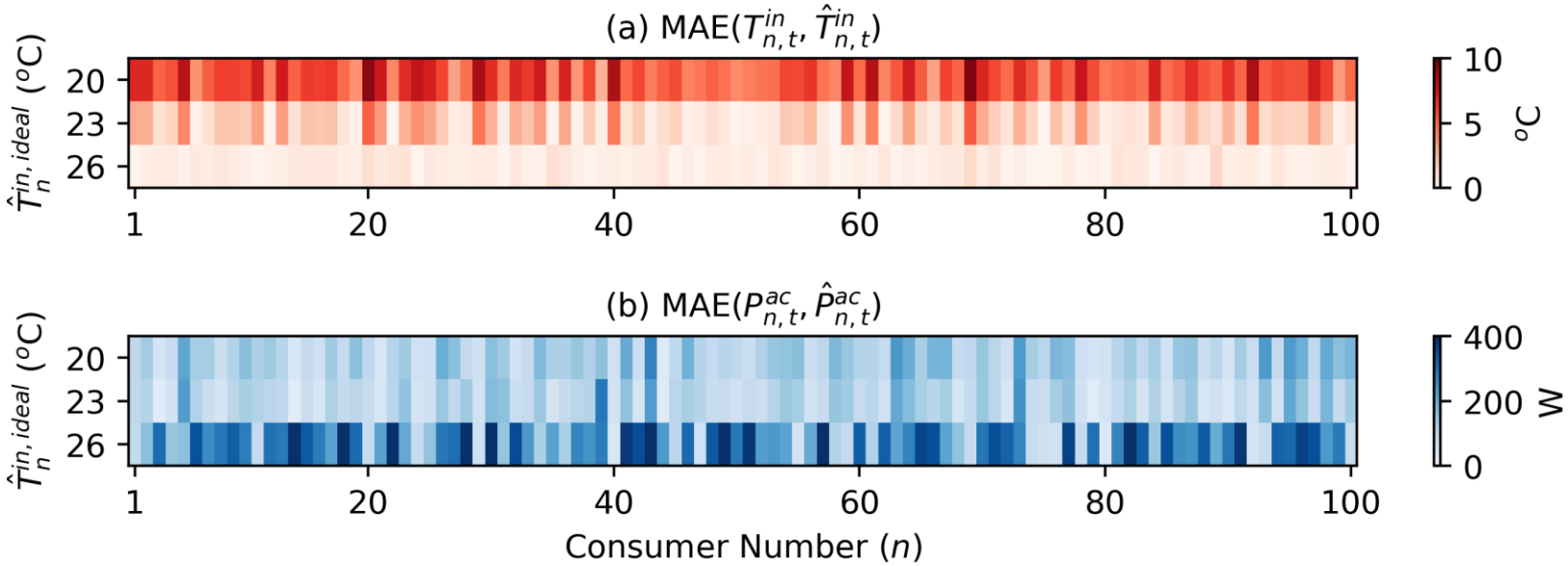
- **General comments**
  - Several established firms (Schneider Electric, Johnson Controls, etc.) have failed to commercialize MPC-based solutions (Henze et al 2024)
  - Models are engineering-intensive, optimization can be compute-intensive
  - Reinforcement learning is unsafe and requires too long to learn
  - Reduced-order & hybrid modelling approaches ➔ promising solution
- **Single-building application**
  - Current work – design and test MPC for comparison
  - Benchmark RL vs existing MPC deployment : OmegaALPES platform, G2ELab
  - Some real-world trials with MPC-based group in NTU
    - MPC model development and deployment ➔ 2-3 years of effort
    - RL (offline initialization) ➔ deployed in 1 month, evaluation ongoing

COMPARISON TO MODEL-BASED APPROACHES

- **Multi-building application**
  - Form of the model is important - explicit analytical model or “black-box” simulator?
  - No optimization-based technique for smart-meter-only context; at the minimum, requires indoor temperature measurements
  - Social factors in residential demand response ➔ difficult to incorporate
  - Solving optimization problem at scale is difficult ➔ (Mai 2024)

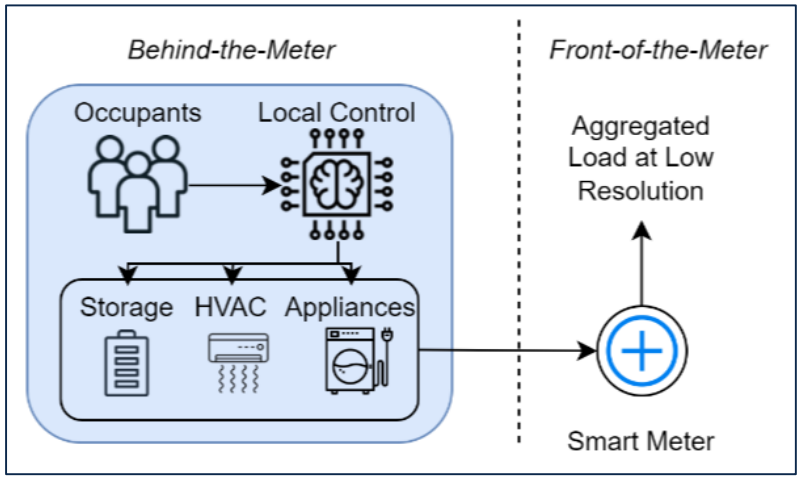
Mai, V.; Maisonneuve, P.; Zhang, T.; Nekoei, H.; Paull, L.; Lesage-Landry, A. Multi-Agent Reinforcement Learning for Fast-Timescale Demand Response of Residential Loads. *Mach Learn* **2024**, *113* (8), 5203–5234. https://doi.org/10.1007/s10994-023-06460-4.

MAIN RESULTS : ACCURACY OF ACM



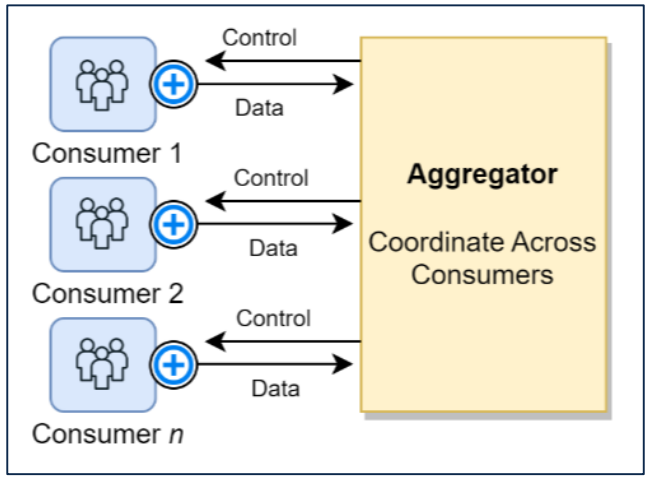
- ACM fit to 2 months of historical data for *M=3* assumed values of $T^{in,ideal}$
- Wide variation in ACM accuracy (MAE ($T^{in}_{n,t}$) ~ 0-10 °C)
- All models follow (simplified) system dynamics

CONTROL PROBLEMS IN DEMAND RESPONSE



- **At the consumer level**
  - Allow control of HVAC systems, battery storage
  - Local objectives – reduce bills, maintain thermal comfort
  - Limited access to data – only front-of-the-meter measurements

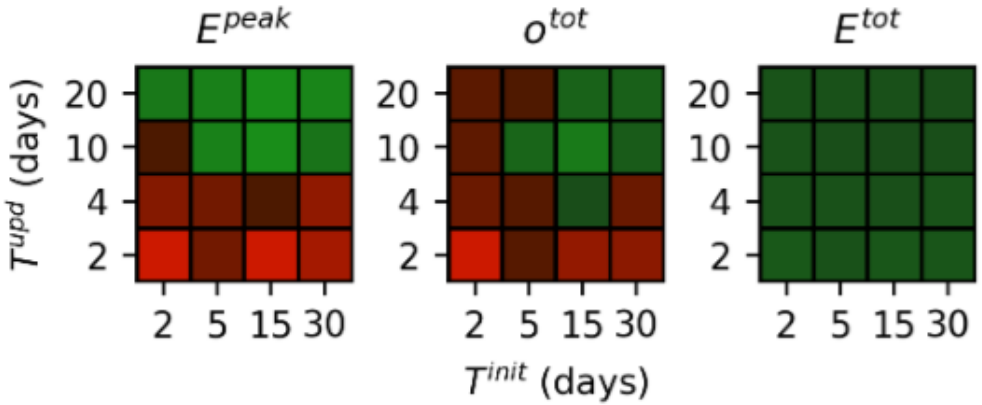
- **At the Aggregator level**
  - Scalable control across many consumers
  - Coordination objectives – peak reduction, load flattening
  - Respect consumer privacy and agency – allow overrides



cnrs @CREATE Singapore

TECHNIQUES IMPACT ON LEARNING STABILITY

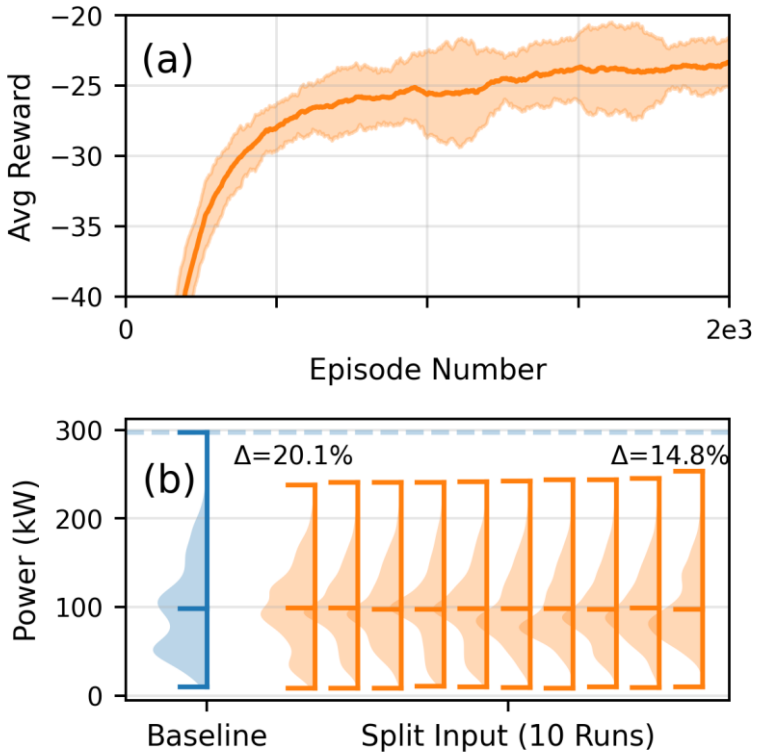
Data-Driven Initialization



(a) π^{init} = Offline RL

More variance and sensitivity to hyperparameters when using only data

Model-Driven Initialization

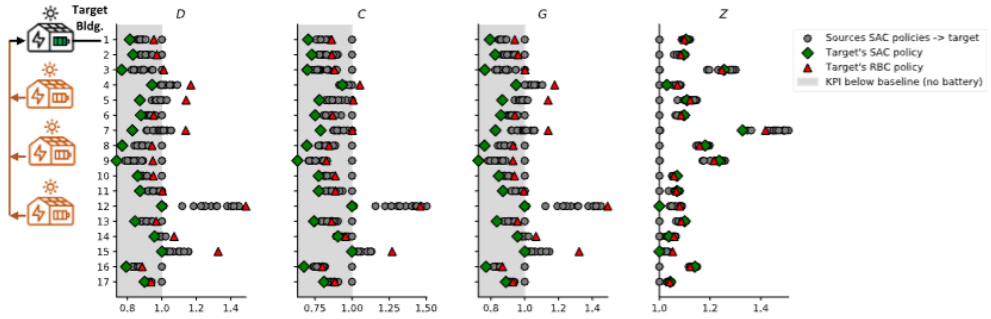


Model-driven is reliable since we can train for long periods and explore many cases

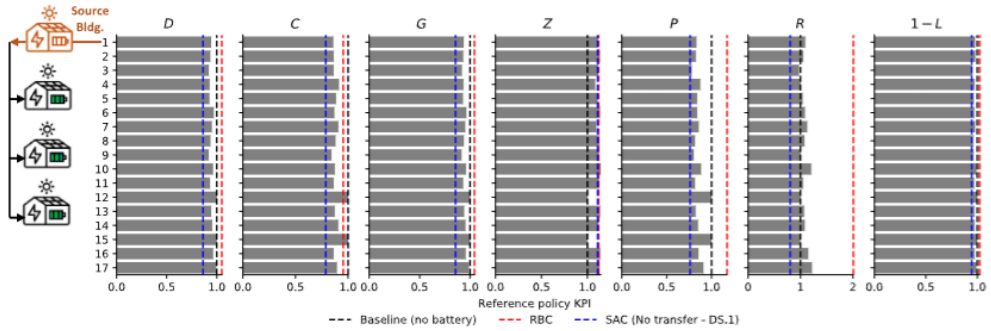
cnrs @CREATE Singapore

TRANSFERABILITY OF REINFORCEMENT LEARNING CONTROLLERS

- **Sensitivity of RL controller to variations in hyperparameters, weather data, building types etc extensively studies ➔ PhD Thesis, Kingsley Nweye (UT Austin)**



(a) Building-level KPIs when one of other 16 buildings' (sources) agents is transferred to a target building.

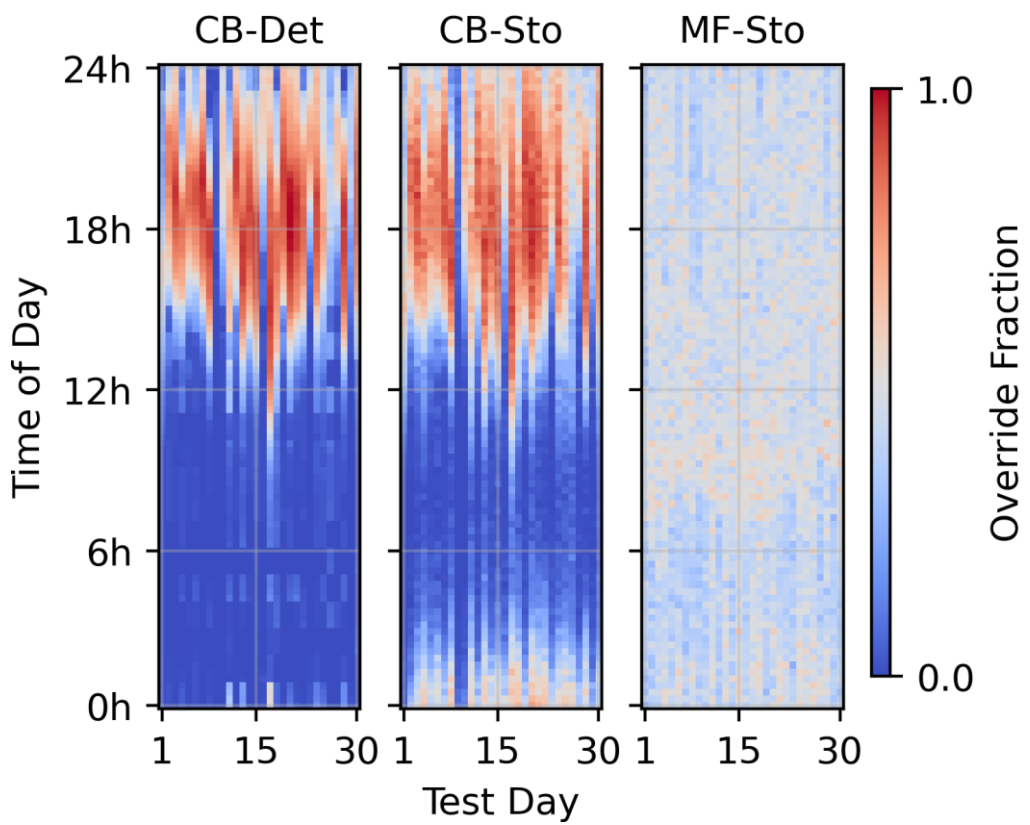


(b) District-level KPIs when a source building's SAC agent is transferred to other 16 buildings (targets).

- **Main findings**
  - Sensitive to hyperparameters and even random seeds
  - Some degree of transferability across buildings and climate zones
  - Poor transfer from hot climate zone data to cool climate zone building and vice versa
  - Similar buildings ➔ able to transfer and fine-tune

Nweye, K.; Sankaranarayanan, S.; Nagy, Z. MERLIN: Multi-Agent Offline and Transfer Learning for Occupant-Centric Operation of Grid-Interactive Communities. *Applied Energy* **2023**, *346*, 121323. https://doi.org/10.1016/j.apenergy.2023.121323.

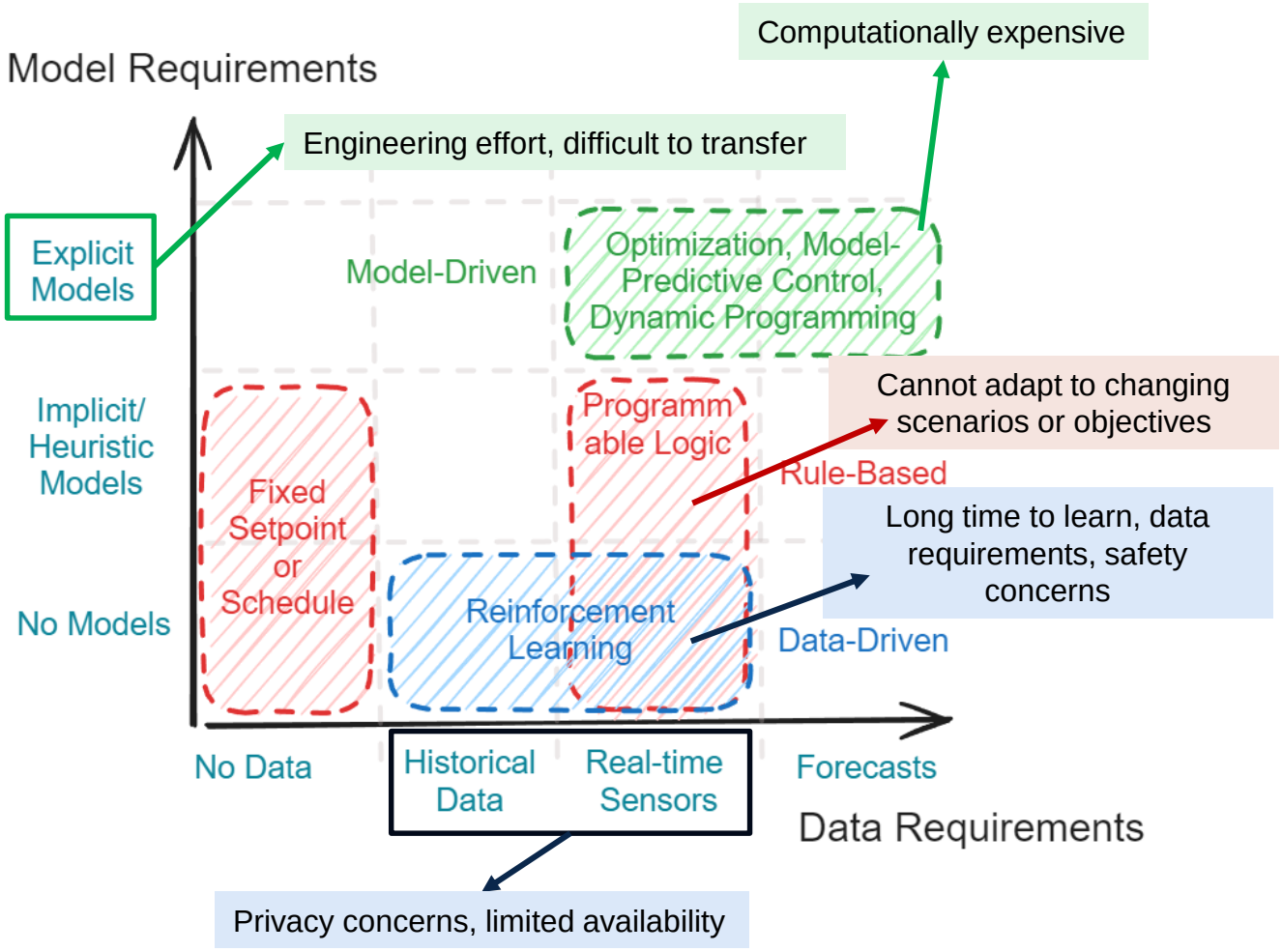
SENSITIVITY TO DISCOMFORT MODELS



Different discomfort models for consumers tested (in deployment phase)

- **RL agent still reduces peaks**
- **In some cases, performance actually improves**
- **Overrides adversely affect peak reduction**
- **The exact type of override behaviour is important**

THE RESEARCH GAP



- **Research focus : algorithms, solution optimality**
  - "Abyss between algorithms and applications" (Henze et al, 2024)
- **The research gap –**
  - Practical and scalable control solutions
  - Address real-world challenges

Henze, G. P.; Kircher, K. J.; Braun, J. E. Why Has Advanced Commercial HVAC Control Not yet Achieved Its Promise? *Journal of Building Performance Simulation* **2024**, 1–12. https://doi.org/10.1080/19401493.2024.2429728.

SAFETY RULES AND EVALUATION METRICS

Metrics	Equation
Electricity Bill	$\sum_t b_t = \sum_t P_t^{grid} \cdot \lambda_t$
Thermal Comfort	$\frac{1}{T} \sum_t \lvert \mathrm{PMV} \rvert_t$

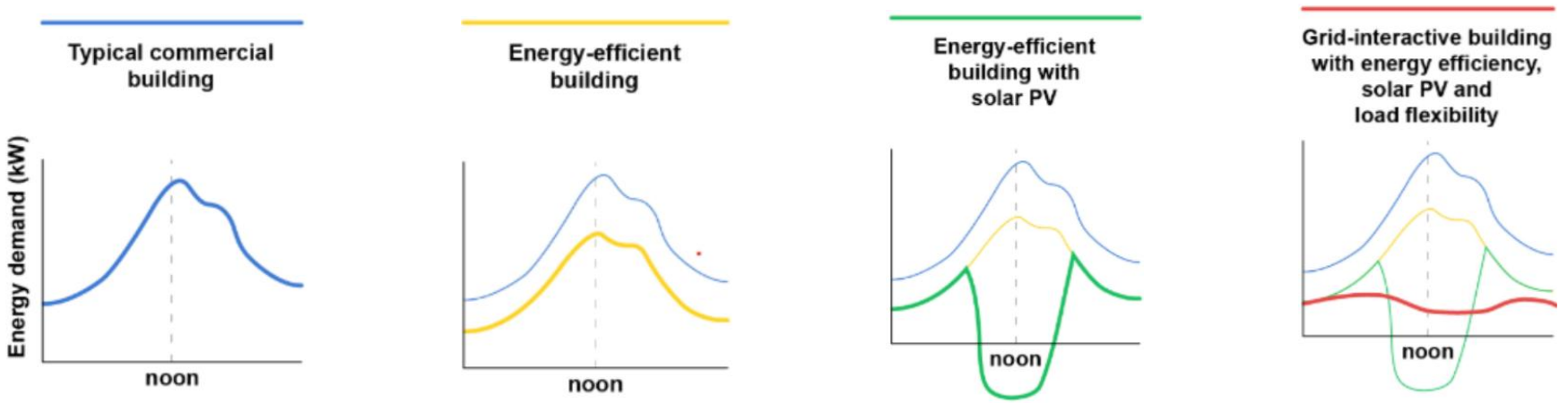
Bill may be negative if surplus energy is exported to the grid

- **Safety rules**
  - Switch off AC outside working hours
  - Maintain occupied indoor temp 20°C – 27°C
  - Battery SOC always between 20% - 90%
- **PMV ➔ Predicted Mean Vote**
  - Measure of thermal comfort
  - -3 = too cold, 0 = ideal, +3 = too hot
  - Ideal range : |PMV| < 0.5 (ASHRAE)

Fanger, P. O. *Thermal Comfort: Analysis and Applications in Environmental Engineering*; Danish Technical Press, 1970.

ASHRAE. *Standard 55 – Thermal Environmental Conditions for Human Occupancy*. https://www.ashrae.org/technical-resources/bookstore/standard-55-thermal-environmental-conditions-for-human-occupancy (accessed 2025-08-22).

BUILDINGS AND THE POWER GRID



EVALUATION METRICS

Metric	Equation
Energy Under Peak Loads E^{peak} (MWh)	$\sum_{n,t} P^{grid}_{n,t} \Delta t$ if $G_t >$ 200 kW
Fraction of Overrides o^{frac}	$\frac{1}{\|N\|\|T\|} \sum_{n,t} o_{n,t}$
Total Energy Consumption E^{tot} (MWh)	$\sum_{n,t} P^{grid}_{n,t} \Delta t$

$G_t = \sum_n P^{grid}_{n,t}$

- **Meter data available for Mar-Apr**
  - Does not include $o_{n,t}$
- **Test duration : 30 days (May)**
  - Metrics are calculated over this period