P-V2B: A Neuro-Symbolic Framework for Leveraging User Persistence in Vehicle-to-Building Charging

Rishav Sen
Vanderbilt University
Nashville, TN, USA
rishav.sen@vanderbilt.edu

Fangqi Liu
Vanderbilt University
Nashville, TN, USA
fangqi.liu@vanderbilt.edu

Jose Paolo Talusan
Vanderbilt University
Nashville, TN, USA
jose.paolo.talusan@vanderbilt.edu

Ava Pettet
Nissan Advanced Technology Center - Silicon Valley
Santa Clara, CA, USA
ava.pettet@nissan-usa.com

Yoshinori Suzue
Nissan Advanced Technology Center - Silicon Valley
Santa Clara, CA, USA
yoshinori.suzue@nissan-usa.com

Ayan Mukhopadhyay
Vanderbilt University
Nashville, TN, USA
ayan.mukhopadhyay@vanderbilt.edu

Abhishek Dubey
Vanderbilt University
Nashville, TN, USA
abhishek.dubey@vanderbilt.edu

Abstract

Vehicle-to-Building (V2B) integration is a cyber–physical system (CPS) where Electric Vehicles (EVs) enhance building resilience by serving as mobile storage for peak shaving, reducing monthly peak-power demand charges, supporting grid stability, and lowering electricity costs. We introduce the Persistent Vehicle-to-Building (P-V2B) problem, a long-horizon formulation that incorporates user-level persistence, where each EV corresponds to a consistent user identity across days. This structure captures recurring arrival patterns and travel-related external energy use, common in employee-based facilities with regular commuting behavior. Persistence enables multi-day strategies that are unattainable in single-day formulations, such as over-charging on low-demand days to support discharging during future high-demand periods. Real-time decision making in this CPS setting presents three key challenges: (i) uncertainty in long-term EV behavior and building load forecasts, which causes traditional control and heuristic methods to degrade under stochastic conditions; (ii) inter-day coupling of decisions and rewards, where early actions affect downstream feasible charging and discharging opportunities, complicating long-horizon optimization; and (iii) high-dimensional continuous action spaces, which exacerbate the curse of dimensionality in reinforcement learning (RL) and search-based approaches. To address these challenges, we propose a neuro-symbolic framework that integrates a constraint-based Monte Carlo Model Predictive Control (MC-MPC) layer with a learned Value Function (VF). The MC–MPC enforces physical feasibility and manages environmental uncertainty, while the VF provides long-term strategic foresight. Evaluations using real building and EV fleet data from an EV manufacturer in California demonstrate that the hybrid framework substantially outperforms state-of-the-art baselines, significantly reducing demand charge and total energy costs, while ensuring feasibility and full compliance with user charging requirements.

ACM Reference Format:
Rishav Sen, Fangqi Liu, Jose Paolo Talusan, Ava Pettet, Yoshinori Suzue, Ayan Mukhopadhyay, and Abhishek Dubey. 2026. P-V2B: A Neuro-Symbolic Framework for Leveraging User Persistence in Vehicle-to-Building Charging. In *Proceedings of 17th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS '26)*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

For commercial and industrial properties, electricity costs are driven by two factors: total energy consumption (in kWh) and peak power (in kW). Of these, the demand charge—a fee levied on the single highest peak of power drawn during a billing cycle, typically a month—can account for over 30% of a building's total electricity bill [26]. This charge creates a significant financial incentive to reduce this monthly peak. The proliferation of electric vehicles (EVs) and bidirectional Vehicle-to-Building (V2B) technology presents a unique solution to this problem. When connected, an EV fleet can act as a dispatchable energy resource (DER), charging during off-peak hours and discharging during peak hours to provide peak shaving for the building and stabilizing the power grid [10, 12, 13].

The V2B problem requires sequential decision-making under uncertainty, i.e., the building must decide when and how much to charge and discharge each EV under exogenous uncertainty (e.g., arrival and departure of EVs). While recent work has tackled this problem through learning-enabled planning (e.g., reinforcement learning [13] or simpler deterministic approximations [28]), prior work has treated daily EV arrivals as random variables that are independent and identically distributed. However, in practice, EV

ICCPS '26, Saint-Malo, France

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

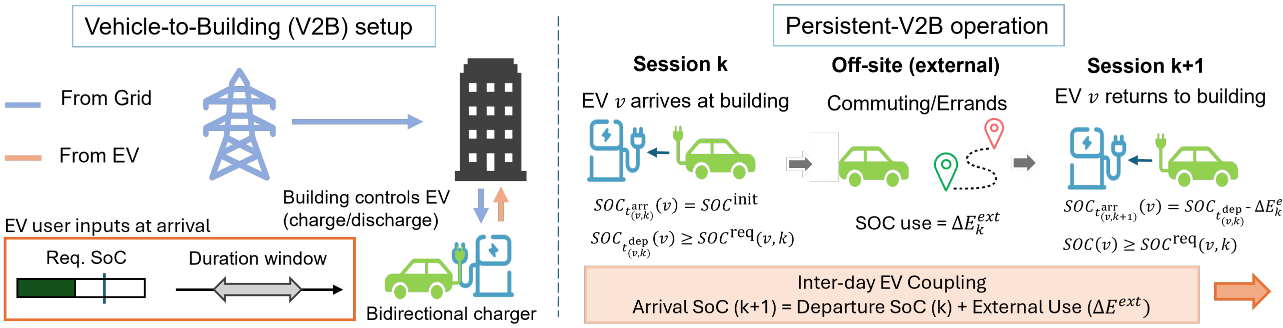
Vehicle-to-Building (V2B) setup
From Grid
From EV
Building controls EV (charge/discharge)
EV user inputs at arrival
Req. SoC
Duration window
Bidirectional charger
Persistent-V2B operation
Session k
EV v arrives at building
$SOC_{t^{\text{arr}}_{(v,k)}}(v) = SOC^{\text{init}}$
$SOC_{t^{\text{dep}}_{(v,k)}}(v) \geq SOC^{\text{req}}(v,k)$
Off-site (external)
Commuting/Errands
SOC use = ΔE_k^{ext}
Session k+1
EV v returns to building
$SOC_{t^{\text{arr}}_{(v,k+1)}}(v) = SOC_{t^{\text{dep}}_{(v,k)}} - \Delta E_k^{ext}$
$SOC(v) \geq SOC^{\text{req}}(v,k)$
Inter-day EV Coupling
Arrival SoC (k+1) = Departure SoC (k) + External Use (ΔE^{ext})

Figure 1: Overview of the persistent V2B (P-V2B) framework for estimating and responding to user behavior in an EV charging network. Electricity flows from the power grid to a building, which allocates energy to connected EVs. At the individual level, observed user behavior (e.g., arrival times, charging preferences) is recorded and aggregated across the pool to form a collective behavioral profile. This pooled data informs predictive models that estimate future user behavior, enabling the system to make proactive energy allocation decisions. The framework supports dynamic negotiation and personalized incentives, ensuring that building-level decisions align with both grid constraints and user needs.

owners arrive at buildings repeatedly over time, e.g., in a large commercial building, the same set of employees arrive regularly over time. We argue that this persistent structure presents an opportunity that can be exploited. Consider that an EV arrives daily at a building. In anticipation of a day on which the building has a high peak, the EV can be charged over the desired state-of-charge (SoC) on the prior day, giving the building more flexibility to discharge the vehicle on the next day for improved peak shaving. Figure 1 shows the operation in a V2B setting, and the possibility for energy cost reduction and peak shaving.

In collaboration with a real-world smart building facility equipped with EV chargers, we propose the Persistent Vehicle-to-Building (P-V2B) problem, which aims to control EV charging decisions using knowledge of user persistence information. This information captures the regularity of user behavior across days, including recurring arrival times, charging durations, and energy needs. Such persistence is common in employee-based facilities where commuters regularly return to the same location, including offices, industrial sites, and schools, and for commercial fleet operators. It is closely linked to observable arrival characteristics such as commuting frequency, departure times, and daily charging demand, which often correlate with commute distance and can be reported or estimated from user data.

Unfortunately, (near) real-time decision-making in the V2B is highly complex and must address several fundamental challenges in the context of CPS planning and control. *First*, EV and building behaviors are highly stochastic: arrival and departure times, charging requirements, and building loads vary unpredictably, causing traditional deterministic or heuristic methods to degrade under uncertainty [16, 28, 33]. *Second*, charging decisions are inter-day coupled, as early actions influence future demand peaks over the monthly billing cycle, creating delayed rewards and complicating long-horizon optimization. *Third*, V2B charging requires allocating continuous charging rates across many EVs in real time, producing a high-dimensional action space that is difficult for predictive and policy-search methods. As a result, existing approaches such as Model Predictive Control (MPC) [16, 33], Monte-Carlo Tree Search (MCTS) [18, 24], and Reinforcement Learning (RL) [1, 13] face limitations in real-time responsiveness and performance stability in persistent and uncertain V2B environments.

To address these challenges, we propose a neuro-symbolic control framework that integrates constraint-based model predictive control (MPC) with a neurally learned value function (VF) that estimates the long-term expected returns from a state of the system.[1] The MPC layer provides short-horizon optimization with explicit physical and user constraints, ensuring feasible charging actions that satisfy all EV energy requirements. The value function layer captures long-term effects that MPC alone cannot anticipate by estimating the remaining monthly demand charge based on the system state at the end of each day. In combination, the neuro-symbolic design leverages the robustness and safety guarantees of MPC while adding the strategic foresight of a learned long-term predictor. This approach overcomes the lack of robustness in traditional Reinforcement Learning (RL) and the limited long-horizon foresight in MPC. We evaluate our framework using real building load data and fleet-level EV usage collected in collaboration with a major EV manufacturer in California. These anonymized commuting patterns allow us to model user persistence and assess the effectiveness of persistence-aware control strategies. Using these real-world datasets, our approach achieves substantial improvements in demand charge reduction and total operating cost, while consistently satisfying all user charging requirements. We summarize our key contributions below:

(1) **A novel formulation of the P-V2B problem.** We introduce the first EV charging control framework that explicitly incorporates user persistence into the interaction between EV fleets and building energy systems (Section 3).
(2) **A neural VF for long-horizon demand-charge prediction.** We design a VF that predicts the impact of current charging decisions on the final monthly demand charge using end-of-day

[1] By "expected," we mean an expectation with respect to the distribution(s) that govern the exogenous uncertainty in the V2B problem.

system states, trained online from MPC-generated trajectories for monthly peak prediction (Section 4.2).

(3) **A neuro-symbolic control framework integrating MC-MPC with VF.** MC-MPC provides robust short-horizon feasibility under uncertainty, while the VF offers long-horizon guidance. A piecewise-linear surrogate embeds the VF into MPC for real-time optimization (Section 4.1).

(4) **Real-world evaluation.** Using EV fleet data from a major California EV manufacturer, our framework achieves substantial improvements over MCTS and RL baselines (Section 6).

2 Related Work

The optimization of V2B systems to mitigate demand charges is a complex, long-horizon stochastic control problem [12, 13]. The core challenge lies in balancing real-time operational uncertainties with the temporally-coupled, month-long demand charge objective [7]. Our work is positioned at the intersection of two primary research areas: V2B optimization and hierarchical control methods.

V2B Optimization via Daily Decomposition The V2B problem has been effectively formulated as a Markov Decision Process (MDP), but the long horizon and sparse reward of the monthly demand charge make it intractable to solve directly. Consequently, the state-of-the-art has converged on temporal decomposition as the primary solution strategy [4].

Advanced methods, whether based on online search [23, 31], reinforcement learning [11, 32], or model predictive control [20, 29], all adopt this approach. They reformulate the intractable month-long problem into a series of tractable, independent *daily* subproblems. This is enabled by a "transient fleet" assumption, which treats each EV's arrival as a new, session-based event, independent of past or future visits. This daily-decomposed approach, however, is blind to the multi-day dependencies of *persistent* users.

Hierarchical Control and Approximate Dynamic Programming Our approach, in contrast, avoids daily decomposition by adopting a hierarchical framework. This methodology is well established for solving large-scale, long-horizon stochastic control problems. In control theory, it is common to use a short-horizon Model Predictive Control (MPC) guided by a terminal cost [14, 15]. This terminal cost serves to approximate the long-term value of the state at the end of the MPC's short planning horizon, ensuring its myopic actions are aligned with a global objective [21].

In operations research, Approximate Dynamic Programming (ADP) provides a formal framework for solving high-dimensional Markov Decision Processes (MDPs) by learning an approximation of the value function (the "cost-to-go") [3, 22]. Similarly, Hierarchical Reinforcement Learning (HRL) decomposes problems by timescale, using a high-level policy to set goals (or value-based guidance) for a low-level, operational policy [6].

The Research Gap: Beyond Daily Decomposition A clear gap emerges at the intersection of these fields. The state-of-the-art in V2B [2, 10] relies on a daily decomposition that is incompatible with the persistent user problem. Simultaneously, the state-of-the-art in control theory provides clear, proven tools for solving long-horizon problems without such decomposition. In EV integration, MPC allows rolling-horizon adaptation to uncertain user behaviors [8]. However, most implementations assume either deterministic inputs or a limited set of forecast scenarios, often with single-stage optimization and no shared control enforcement across futures. In contrast, our MPC method enforces identical first-stage decisions across multiple sampled future scenarios, yielding robustness to both aleatoric (e.g., arrival fluctuations) and epistemic (e.g., load model shifts) uncertainty. This structure aligns with sample average approximation (SAA) theory and provides empirical robustness without over-conservatism [25].

Our approach addresses this gap. To the best of our knowledge, we are the first to formally define the persistent user V2B (P-V2B) problem and propose a solution that explicitly rejects daily decomposition. We do this by applying a hierarchical control framework, uniting a short-term operational MC-MPC with a long-term strategic Value Function (VF). This VF provides the terminal cost, allowing us to solve the user-coupled, month-long optimization problem in a tractable manner.

3 Offline Oracle Formulation of P-V2B

We consider the Persistent Vehicle-to-Building (P-V2B) system, in which each building manages its own EV charging infrastructure. A distinguishing feature of P-V2B is the use of persistence information—predicted EV arrivals, departures, and user requirements over the entire billing period. Such information is often available in office campuses where employees report commuting plans, in commercial fleets with recurring duty cycles, or in residential communities where users communicate expected charging needs in advance. In this section, we adopt a full-information offline (oracle) formulation of the P-V2B problem, assuming complete knowledge of all EV schedules. This oracle setting serves as the ideal benchmark and enables proactive long-horizon coordination that balances daily charging requirements with long-term cost efficiency while respecting user needs.

Persistent EV users and session information. Under the offline oracle formulation, we assume full knowledge of all EV sessions over the billing period $\mathcal{T}$. In the P-V2B setting, EV users exhibit persistent behavior, returning to the facility multiple times within $\mathcal{T}$. Each visit constitutes a charging session. For each EV $v \in \mathcal{V}$, sessions are indexed by $k \in \mathcal{K}_v = \{1, \dots, K_v\}$, with session k occupying the interval $[t^{\mathrm{arr}}_{v,k}, t^{\mathrm{dep}}_{v,k}] \subseteq \mathcal{T}$.

We denote by $\mathrm{SOC}_t(v) \in [0,1]$ the state of charge (SoC) of EV v at time t, expressed as a *fraction* of its battery capacity E^{cap}_v measured in kWh. Each session specifies a required departure SoC $\mathrm{SOC}^{\mathrm{req}}(v,k)$ that must be met by $t^{\mathrm{dep}}_{v,k}$. Between sessions, the EV consumes energy off-site. Let $\Delta E^{\mathrm{ext}}_{v,k} \geq 0$ denote this external energy usage in kWh, which determines the SoC at the next arrival. The first session begins with the initial SoC $\mathrm{SOC}^{\mathrm{ini}}(v)$. In practice, all session-related quantities ($t^{\mathrm{arr}}_{v,k}$, $t^{\mathrm{dep}}_{v,k}$, $\mathrm{SOC}^{\mathrm{req}}(v,k)$, E^{cap}_v, $\Delta E^{\mathrm{ext}}_{v,k}$) must be estimated from historical data or user-reported commuting patterns during online operation.

Chargers and charging control. The building can have N heterogeneous chargers, each being either bidirectional (supports charging and discharging) or unidirectional (only supports charging). Each EV v during session k is mapped to a charger at time t using

$\zeta(v,k)$. Once assigned, an EV remains at the charger through its duration $[t_{v,k}^{\mathrm{arr}}, t_{v,k}^{\mathrm{dep}}]$. The charging rate at each timestep r_t^v (kW) is subject to $\min r^{\zeta(v,k)} \le r_t^v \le \max r^{\zeta(v,k)}, \forall t \in \mathcal{T}$, where $\min r^{\zeta(v,k)}$ is the minimum rate (which can be < 0 for bidirectional chargers), and $\max r^{\zeta(v,k)}$ is the maximum rate of the charger $\zeta(v,k)$ connected to the EV.

EV–Charger assignment and availability. We first determine which charger an EV uses. For fairness and real-world implementability, we adopt a fixed first-come, first-served (FCFS) assignment rule. The mapping $\zeta(v,k)$ assigns EV v to a charger k upon arrival, prioritizing bidirectional chargers when available. Given this assignment, we define an availability indicator a_t^v that specifies whether EV v is physically present at its assigned charger at time t: $a_t^v = 1$ if $t \in [t_{v,k}^{\mathrm{arr}}, t_{v,k}^{\mathrm{dep}}]$ for its assigned charger k, and $a_t^v = 0$ otherwise, determining when charging-rate control is permitted.

Total building load. A building's power use in timestep t is

$$P_t = b_t + \sum_{v\in\mathcal{V}} a_t^v r_t^v, \qquad \forall t \in \mathcal{T}, \tag{1}$$

where b_t is the non-EV building load (in kW) at time t.

SoC dynamics within a session. Let $\mathrm{SOC}_t(v) \in [0,1]$ denote the SoC of EV v at time t while connected to the building, E_v^{cap} its battery capacity (kWh), and Δt the step length (in hours). For any session k and $t \in [t_{v,k}^{\mathrm{arr}}, t_{v,k}^{\mathrm{dep}} - 1]$,

$$\mathrm{SOC}_{t+1}(v) = \mathrm{SOC}_t(v) + \frac{r_t^v \, \Delta t}{E_v^{\mathrm{cap}}}, \qquad 0 \le \mathrm{SOC}_t(v) \le 1. \tag{2}$$

The EV user provides a SoC requirement $\mathrm{SOC}^{\mathrm{req}}(v,k)$ to be fulfilled till $t_{v,k}^{\mathrm{dep}}$ for each session, and we guarantee to reach the required SoC by departure, if possible at the fastest charging rate:

$$\mathrm{SOC}_{t_{v,k}^{\mathrm{dep}}}(v) \ge \mathrm{SOC}^{\mathrm{req}}(v,k) \tag{3}$$

At the start of the first session, $\mathrm{SOC}t_{v,0}^{\mathrm{arr}}(v) = \mathrm{SOC}^{\mathrm{ini}}(v)$, where $\mathrm{SOC}^{\mathrm{ini}}(v)$ is the SoC of EV v at the start of the billing period.

Off-site usage between sessions (persistence coupling). Between session k and session $k+1$, the EV departs the site and consumes energy off-site (e.g., driving). The SoC at the arrival of the next session satisfies

$$\mathrm{SOC}_{t_{v,k+1}^{\mathrm{arr}}}(v) = \mathrm{SOC}_{t_{v,k}^{\mathrm{dep}}}(v) - \frac{\Delta E_{v,k}^{\mathrm{ext}}}{E_v^{\mathrm{cap}}}, \qquad \forall k \in K_v. \tag{4}$$

Departure sufficiency for off-site usage. Between session k and $k+1$, EV v consumes $\Delta E_{v,k}^{\mathrm{ext}} \ge 0$ (kWh) off-site. Let E_{cap}^v be the battery capacity. We assume, the SoC required by the user is sufficient top fulfill the off-site SoC usage:

$$\mathrm{SOC}^{\mathrm{req}}(v,k) \ge \frac{\Delta E_{v,k}^{\mathrm{ext}}}{E_v^{\mathrm{cap}}}. \tag{5}$$

Building's Electricity Bill. For the building, the energy cost over the billing period is given by

$$\phi_t^{\mathrm{energy}} = w_t^e \cdot P_t \cdot \Delta t, \tag{6}$$

where w_t^e denotes the time-of-use price.

In addition to energy charges, the demand charge is determined by the maximum observed load:

$$\Phi_\tau^{\mathrm{demand}} = w^d \cdot \max_{k\in[0,\tau]} P_k \tag{7}$$

where w^d is the demand charge rate, the demand charge is accrued from the start of the billing cycle to time step τ.

Thus, the total electricity bill for building combines energy costs and demand charges over the billing cycle:

$$\Phi = \sum_{t\in\mathcal{T}} \phi_t^{\mathrm{energy}} + \Phi_{t_{\mathrm{end}}}^{\mathrm{demand}}. \tag{8}$$

where t_{end} is the end of the billing cycle.

4 Online P-V2B and Neuro-Symbolic Control

In practice, the P-V2B problem must be solved online as a sequential decision-making task, which we model as a Markov Decision Process (MDP). The objective is to find a policy π^* that maximizes the expected total reward (negative total electricity cost) over the monthly billing cycle.

State ($\mathbb{S}_t$). The controller updates all charger power rates at discrete decision times $t \in \mathcal{T}$ with a fixed interval Δt. At each decision time, it gathers the information needed to determine the next charging action. We represent the information available at time t as

$$\mathbb{S}_t = \Big(P_t^{\max,\mathrm{hist}}, b_t, \{\zeta_t(v), \mathrm{SOC}_t(v), \mathrm{SOC}^{\mathrm{req}}(v), \hat{U}_t(v)\}_{v\in\mathcal{V}}\Big),$$

where $P_t^{\max,\mathrm{hist}} = \max_{\tau<t} P_\tau$ is the historical peak load up to time t, b_t is the building load at time t, $\zeta_t(v)$ is the charger currently serving EV v, $\mathrm{SOC}_t(v)$ is its state of charge, $\mathrm{SOC}^{\mathrm{req}}(v)$ is its required departure SoC, and $\hat{U}_t(v)$ denotes predicted future behavior of user v, such as expected arrival and departure times and anticipated off-site energy use.

Actions (A_t). At each decision time t, the controller selects the charging or discharging rates for all connected EVs. The action is $A_t = \{r_t^v\}_{v\in\mathcal{V}}$, where each rate satisfies $r_t^v \in \big[r_{\min}^{\zeta_t(v)}, r_{\max}^{\zeta_t(v)}\big]$, with $r_{\min}^{\zeta_t(v)}$ and $r_{\max}^{\zeta_t(v)}$ denoting the allowable power range of the charger currently assigned to EV v.

State Transition. The next state $\mathbb{S}_{t+1}$ is computed by updating all system components. EV SoC evolves according to Eq. (2). EVs whose sessions end at time t depart, and new arrivals enter following the predicted persistence schedule. Between sessions, SoC updates follow the persistence model in Eq. (4). The building load is updated using the stochastic load model. Together, these updates define the transition $\mathbb{S}_{t+1} = f(\mathbb{S}_t, A_t)$.

Episode Reward (R). The episode reward is defined as the negative total building bill over the billing cycle $\mathcal{T}$, where the bill Φ is computed according to Eq. (8). Thus the episode reward is $R = -\Phi$, reflecting both energy charges and the demand charge incurred under the chosen charging actions.

4.1 The Neuro-Symbolic Control Framework

We now introduce our proposed Neuro–Symbolic (NS) control architecture, which integrates a short-horizon Monte Carlo Model Predictive Control (MC-MPC) module with a long-horizon neural Value Function (VF). This hybrid design is motivated directly by the structure of the P-V2B problem.

The P-V2B problem features a high-dimensional continuous action space and strict physical constraints, making MPC a natural short-horizon symbolic reasoner because it directly optimizes continuous charging rates while guaranteeing feasibility. However, the environment evolves under significant uncertainty: EV arrivals are stochastic, user behavior varies across days, and building load fluctuates with exogenous conditions. To capture these uncertainties, we extend MPC with *Monte Carlo sampling*. At each decision time t, the controller generates N sampled future trajectories of all uncertain quantities, forming a scenario set $\mathcal{F}$. A single MILP is then solved to minimize the expected cost Φ (Eq. (8)) over the predictive horizon $[t,\ t_{\text{eod}}]$, where t_{eod} denotes the end of the current day. This procedure produces a robust first-stage action A_t that performs reliably across all sampled futures.

While MC-MPC effectively optimizes over short horizons, the P-V2B objective is fundamentally long-term, coupling decisions across many days through demand-charge accumulation and persistent constraints. Extending MC-MPC to month-long horizons is computationally intractable. To bridge this gap, we augment the symbolic controller with a value function J_{VF} that predicts the long-term cost-to-go (demand charge, $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}}$) beyond the MPC's finite horizon.

Combined Objective. The jointly optimized action A_t is obtained by solving an MC-MPC problem whose terminal cost is provided by the learned VF J_{VF}. The controller minimizes the expected short-term energy cost, demand charge till the end of the day t_{eod}, and proactive charging, plus the predicted long-term value:

$$\min_{A_t} \mathbb{E}_{f\in\mathcal{F}}\left[\sum_{k=t}^{t_{eod}}\left(\phi_k^{\text{energy}} - w_{\text{pc}}\sum_{v\in V_k} r_k^v \Delta t\right) + \Phi_{t_{\text{eod}}}^{\text{demand}} + J_{VF}(\mathbb{S}_{t_{eod}})\right] \quad (9)$$

subject to all physical and user constraints (Eq 1–5). This objective integrates three elements:

(1) **Short-horizon cost.** The expectation over $f \in \mathcal{F}$ spans N sampled future trajectories, capturing short-term energy cost ϕ_k^{energy} and demand charge up to end-of-day $\Phi_{t_{\text{eod}}}^{\text{demand}}$, effectively incorporating uncertainty.

(2) **Proactive charging.** To build an energy buffer for future peak shaving, we add the reward term $-w_{\text{pc}} \sum r_k^v \Delta t$, encouraging slight overcharging beyond immediate SoC needs. The small coefficient w_{pc} is tuned to avoid new peaks, enabling persistent cross-day coordination to reduce monthly demand charges.

(3) **Learned value function.** The term $J_{VF}(\mathbb{S}_{t_{\text{eod}}})$ is the neural value function, which estimates the long-term cost-to-go from the end-of-day state. Specifically, it predicts the remaining monthly demand charge given $\mathbb{S}_{t_{\text{eod}}}$, providing the long-term foresight that short-horizon MPC alone cannot supply. The structure and training of this value function are detailed in Section 4.2.

4.2 Value Function Definition and Integration

The value function J_{VF} is defined as an estimate of the remaining monthly demand cost conditioned on the system state at the end of each day. It serves as a long-horizon proxy for the demand-charge component of the objective and enables the controller to account for inter-day coupling effects. This subsection describes the construction of the value function input state and the learning target used to train J_{VF}.

Value Function Definition. The value function operates on a compact abstraction of the end-of-day system state. We define

$$\text{VFInput}(\mathbb{S}_t) = \left(t,\ \hat{\text{SOC}}_t^{(1)},\ \hat{\text{SOC}}_t^{(2)},\ \hat{\text{SOC}}_t^{(3)}\right), \quad (10)$$

where t is the current day in the billing cycle (evaluated at the end of day). Each $\hat{\text{SOC}}_t^{(c)}$ represents the aggregated SoC buffer of EV users in cluster c: $\hat{\text{SOC}}_t^{(c)} = \sum_{v\in C^{(c)}} \max\big(0,\ \text{SOC}_t(v) - \text{SOC}^{\text{req}}(v)\big)$, where t is the current day in the billing cycle (evaluated at day's end). To capture the system's persistent energy buffer, we cluster EV users into three groups using their historical arrival frequency and average stay duration. For each cluster c, we compute the total excess SoC: $\hat{\text{SOC}}_t^{(c)} = \sum_{v\in C^{(c)}} \max\big(0,\ \text{SOC}_t(v) - \text{SOC}^{\text{req}}(v)\big)$, which aggregates the SoC above required levels for all users in $C^{(c)}$. These three values summarize the cross-day charging buffer available for future demand-charge reduction.

The value function is trained to predict the remaining monthly demand cost from the current abstract end-of-day state. Its learning target is the final monthly demand charge that will be incurred from this state onward. Thus, $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}} = J_{VF}(\text{VFInput}(\mathbb{S}_t))$, with the reward signal defined as the eventual demand charge at the end of the billing cycle, conditioned on $\mathbb{S}_t$.

Neural Network-MILP Integration. The value function J_{VF} must be embedded directly into the MC–MPC objective so that long-term cost can influence short-horizon optimization (Eq. (9)). In principle, a ReLU-based neural network can be incorporated exactly inside the MILP using Big-M constraints to encode each activation. However, this approach scales poorly: even moderately sized networks produce large mixed-integer programs that are too slow for real-time decision making. To balance accuracy and tractability, we approximate the trained value function with a high-fidelity *Piecewise Linear (PWL) surrogate*. For each day t in the billing cycle, each of the three SoC-buffer features $\left(\hat{\text{SOC}}^{(1)}, \hat{\text{SOC}}^{(2)}, \hat{\text{SOC}}^{(3)}\right)$ is discretized into a one-dimensional grid (e.g., 20 points per dimension), and the trained network is evaluated on all grid points. *Special Ordered Sets of Type 2 (SOS2)* constraints ensure that, within each dimension, the MILP activates only two adjacent grid points, producing a local 1D linear interpolation. Combining these along the three axes yields a trilinear PWL approximation of J_{VF} [9, 30]. This interpolation is linearized using *McCormick envelopes* [17], a standard approach for handling multilinear terms in MILPs. The resulting surrogate provides an accurate representation of the neural value function while keeping the optimization small enough to solve within seconds, enabling reliable real-time MC–MPC control.

4.3 MC-MPC Execution with Action Refinement

The controller executes online as in Algorithm 1. At each step t:

(1) Newly arrived EVs are assigned to chargers using a First-Come, First-Served (FCFS) rule.

(2) A set $\mathcal{F}$ of N future trajectories is generated from the stochastic environment model (detailed in Section 6). Each trajectory provides a sampled realization of EV arrivals, SoC requirements, and building load for the remainder of the current day.

(3) The neuro-symbolic MILP in Eq. (9) is solved to obtain a single action A_t that is shared across all sampled futures.

(4) A fast-timescale refinement step adjusts A_t to produce $\tilde{A}_t$, improving the action based on the estimated monthly peak-power boundary before execution.

The Action Refinement at the final step is a post-processing heuristic applied after solving the neuro-symbolic optimization. The intuition is to exploit any remaining power headroom without increasing the monthly demand charge. We use an estimated monthly peak $\hat{P}^{\max}$, obtained from a statistical oracle built from historical data, to define a conservative peak-power boundary. If the current building power P_k is below this boundary, the controller can safely increase charging rates without affecting the monthly demand cost. Using this available slack capacity, the refinement step increases charging for eligible EVs as follows:

$$\overline{r}_t^v = r_t^v + \min\left(\frac{P^{\text{diff}}}{|\mathcal{V}_t|}, \max r^{C(v,k)}\right), \text{ if } SOC_t(v) + \overline{r}_t^v \le SOC^{\text{req}}(v,k) \quad (11)$$

where $P^{\text{diff}} = \hat{P}^{\max} - P_k$ is the available power headroom.

```
Algorithm 1: P- V2B Charging Optimization
Input: Initial state $S_0$, number of samples $N$, billing period
    $T$, estimated monthly peak power $\hat{P}^{max}$
Output: Charging decisions $A_t$
for $t = 0$ to $\mathcal{T}$ do
    Check EV departures and free assigned chargers;
    Assign chargers to new arrivals using FCFS;
    Observe current state $\mathbb{S}_t$;
    Future trajectory set $\mathcal{F} \leftarrow \emptyset$;
    for $i = 1$ to $N$ do
        Sample trajectory $f$ from the generative model
          given $\mathbb{S}_t$, simulate from $t$+1 to $t_{eod}$, and add $f$ to $\mathcal{F}$;
    Get $A_t$ by solving optimization in Eq. (9)
    Refine action to $\tilde{A}_t = f_{\text{refine}}(A_t, \hat{P}^{\max})$ by Eq. (11)
    Apply $\tilde{A}_t$ and update system state to $\mathbb{S}_{t+1}$;
```

4.4 Two-Stage Value Function Training

The value function is trained using a combination of offline oracle supervision and online simulation-based value iteration.

(1) **Offline Pre-training.** We first train the value function using supervised data generated from a persistent MILP oracle. Synthetic monthly scenarios are created using a generative model built from real EV manufacturer data (arrival patterns, stay durations, charging requirements, and building load profiles). For each simulated month, we solve the persistent MILP (Eq. 8) subject to all physical and user constraints (Eq. 1–5). At the end of each day t, we extract the abstract value-function state VFInput($\mathbb{S}_t$) (Eq. 10) and pair it with the oracle final monthly demand charge Φ^{demand}. These (VFInput($\mathbb{S}_t$), Φ^{demand}) pairs form a high-quality supervised dataset that trains J_{VF} to approximate the mapping VFInput($\mathbb{S}_t$) $\mapsto \hat{\Phi}^{\text{demand}}_{t_{\text{end}}}$.

(2) **Online Fine-tuning.** To make training computationally feasible, we approximate MC–MPC behavior using two lightweight controllers: (i) a **Nominal MPC** using mean forecasts and (ii) a **Robust MPC** using 95th-percentile worst-case forecasts. These two variants bracket typical MC–MPC behavior and produce realistic day-to-day charging actions without the computational burden of full sampling. During simulation, at the end of each day t, (i) we record the abstract state VFInput($\mathbb{S}_t$); (ii) continue the simulation to month-end to obtain the realized demand charge $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}}$; (iii) store (VFInput($\mathbb{S}_t$), $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}}$) in a replay buffer; and (iv) update J_{VF} via supervised regression toward $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}}$, gradually refining its estimate of the long-term cost. This value-iteration refinement aligns the VF with MPC behavior without requiring full MC–MPC runs.

5 Theoretical Guarantees of the Persistent Model Formulation

Let π be a charging policy over a monthly horizon $\mathcal{T}$. The objective function is the total electricity cost Φ. We know $\mathcal{K}$ is the set of all physical and user-session constraints.

Problem 1: Persistent (P-V2B). The optimizer finds the optimal policy π_P^* by solving: $\Phi_P = \min_{\pi \in \Pi_P} \Phi$ where Π_P is the set of all policies satisfying all the user session requirements $\mathcal{K}$ and the persistent coupling constraint (Eq. (4)). This set Π_P contains all physically possible policies.

Problem 2: Daily (D-V2B). The optimizer finds the optimal policy π_D^* by solving: $\Phi_D = \min_{\pi \in \Pi_D} \Phi$ where Π_D is the set of all policies satisfying $\mathcal{K}$ and a daily independence constraint, which *forbids* using the coupling equation.

Theorem 1. *The optimal electricity cost achievable under the Persistent V2B (P-V2B) model is always less than or equal to that of the D-V2B model:* $\Phi_P \le \Phi_D$

Proof. The daily independence constraint is an additional, artificial constraint. Any policy $\pi_D \in \Pi_D$ is, by definition, also a valid policy for the persistent model. Therefore, the set of feasible policies for the daily model is a strict subset of the feasible policies for the persistent model: $\Pi_D \subset \Pi_P$ Minimizing an objective function over a larger set yields a result less than or equal to the minimum over a smaller subset: $\min_{\pi \in \Pi_P} \Phi(\pi) \le \min_{\pi \in \Pi_D} \Phi(\pi)$ Thus, we conclude, $\Phi_P \le \Phi_D$. □

5.1 Stochastic Controller Performance Analysis

Let $J(\pi)$ be the total expected cumulative cost of a policy π over the full horizon $\mathcal{T}$, given an initial state $\mathbb{S}_0$:

$$J(\pi) = \mathbb{E}\left[\sum_{t \in \mathcal{T}} -\phi_t^{\text{energy}}(\pi) + \Phi^{\text{demand}}_{t_{\text{end}}}(\pi) \mid \mathbb{S}_0\right]$$

The optimal cost for the P-V2B problem is $\Phi_P^* = \min_\pi J(\pi)$.
Performance Bound of the Baseline (Daily V2B MC-MPC) The daily V2B MC-MPC (D-MPC) solves the D-V2B problem, introducing a non-negative structural error, ϵ_{model}:

$$\epsilon_{\text{model}} = \Phi_D^* - \Phi_P^* \ge 0$$

The expected performance of the baseline D-MPC, also subject to forecast error $\epsilon_{\text{forecast}}$, is:

$$\mathbb{E}[J_{\text{D-MPC}}] = \Phi_D^* + \epsilon_{\text{forecast}} = (\Phi_P^* + \epsilon_{\text{model}}) + \epsilon_{\text{forecast}}$$

Performance Bound of Our Neuro-symbolic Approach (MC-MPC+VF). Our MPC+VF approach solves for the true optimum

Φ_P^* and does not suffer from ϵ_{model}. Its performance is limited by $\epsilon_{\text{forecast}}$ and the VF approximation error, ϵ_{VF}:

$$\epsilon_{VF} = \max_{S\in\mathbb{S}} |J_{\text{VF}}(S) - \Phi_P^*(S)|$$

The expected performance of our MPC+VF is:

$$\mathbb{E}[J_{\text{MC-MPC+VF}}] = \Phi_P^* + \epsilon_{\text{forecast}} + \epsilon_{VF}$$

Thus our MC-MPC+VF is provably better in expectation if $\mathbb{E}[J_{\text{MC-MPC+VF}}] \le \mathbb{E}[J_{\text{D-MPC}}]$. Given

$$\Phi_P^* + \epsilon_{\text{forecast}} + \epsilon_{VF} \;\le\; \Phi_P^* + \epsilon_{\text{model}} + \epsilon_{\text{forecast}},$$

subtracting common terms on both sides yields the key condition $\epsilon_{VF} \le \epsilon_{\text{model}}$. This proves our MC-MPC+VF controller is guaranteed to outperform the daily baseline (D-MPC) in expectation, so long as the approximation error of our Value Function (ϵ_{VF}) is smaller than the inherent structural error of the daily model (ϵ_{model}). We show this empirically in Section 7.

6 Experimental Setup

To validate our proposed hierarchical framework, we design a comprehensive simulation study to compare its performance against a suite of baselines, including state-of-the-art daily-decomposed (transient) methods and persistence-aware oracles. We extend an existing V2B simulation environment (based on [27]) to support the persistent user model. All the MILPs are solved using IBM ILOG CPLEX Optimization Studio [5].

Chargers: We consider a building with 15 bidirectional chargers, having a maxmimum charging rate of 20 kW, and a minimum rate of -20 kW. Any charging action below 0 is discharging the EV.

EV Charging Profile: Based on data from our EV manufacturer partner, a piecewise linear SoC curve is used for modeling realistic EV charging speeds, and is divided into three regions. Power (kW) is 20 kW for SOC$\le$ 83%, then decreases linearly: $-\frac{4}{3}\cdot \text{SOC} + 130$ for 83%–90%, and $-\text{SOC} + 100$ for 90% − 100%. We model discharge as a constant-power process, so the battery's SoC falls linearly over time. All the EVs $v \in \mathcal{V}$ have a battery capacity E^v_{cap} of 60 kWh.

Dataset: We construct a test-bed using a year's data from 2024. Each month includes 200 stochastic samples, which are divided into 170 training and 30 testing samples. These months capture the commuting variations of the employees and reflect the seasonal variation in building load conditions, including differences in heating, cooling, and occupancy patterns. The months also capture shifts in the timing and magnitude of peak power spikes, which are critical for stress-testing energy allocation strategies. Figure 2 illustrates these variations.

The EV manufacturer provided realistic multi-day driver behavior data by sampling from generative models based on individuals' real mobility data to avoid privacy concerns. For each user, we assume the number of weekdays between workplace arrivals can be modeled with a geometric distribution. Each user's off-site energy usage $\Delta E^{\text{ext}}_{v,k}$ is modeled via a gamma regression conditioned on the days between workplace visits (mean log likelihood across users: -1297.53), arrival time each day is modeled via a Gaussian mixture model (mean log likelihood across users: -3.86), and stay durations are modeled via a Weibull accelerated failure time model conditioned on arrival time (mean log likelihood across users: -787.67). Each generated sample contains 15 persistent users whose individualized probabilistic models create the cross-session coupling from Eq. (4). Summary statistics for the EV users are shown in Figure 3. We use real-world time-of-use pricing ($0.18/kWh from 6AM to 10PM, otherwise $0.13/kWh), and demand charge rate ($11.67/kW) from Silicon Valley Power [26] in Santa Clara, California.

We utilize a real-world aggregated consumption profile from our EV manufacturer partner's lab facility in Santa Clara as the basis for the building load data. This profile serves as a deterministic forecast baseline (b^{base}_t), representing the known component of the load used by the controller for short-horizon planning. Because the real building load is not perfectly predictable, we construct the load that occurs in the simulation by introducing uncertainty: for every sample, independent, uniformly-distributed noise, within a ±5% margin to the baseline load at each time step. This resulting fluctuating load serves as the ground truth for the controller.

Forecast Generation. For each control step, the MC-MPC framework requires stochastic scenarios for building load and EV behavior. Building load estimations use the noisy profile generation described above, while electricity prices follow known pre-published utility rates. EV user forecasts are sampled using the same generative behavior models described in the dataset section to estimate EV user behavior $\hat{U}$ which provides arrival, departure times, off-site usage (SoC requirements).

Estimated Peak Power. To provide a reference value for MC-MPC demand charge control during online optimization, we estimate the target peak power $\hat{P}^{max}$ for each month. The training episodes are solved using the P-V2B oracle MILP formulation, which minimizes the total monthly electricity cost Φ subject to constraints in Eq. 1 - 5. We then compute $\hat{P}^{max}$ as the 99th percentile of the peak power values across the training episodes for that month. This approach ensures that $\hat{P}^{max}$ captures typical high-load behavior without being overly sensitive to rare outliers.

EV SoC Requirement: The requirement is to support off-site usage while ensuring SoC remains above the 10% minimum threshold.

Hyperparameters. We tuned the MC-MPC sample size ($\mathcal{F}$) and the proactive charging reward (w_{pc}) via grid search. $\mathcal{F}$ was searched from 5 to 30, and w_{pc} from 0 to 5. We selected the best-performing values of $\mathcal{F} = 10$ and $w_{\text{pc}} = 0.8$ for all experiments.

Parameter	Description	Min	Max	Step	Best
$\mathcal{F}$	Sample size	5	30	5	10
w_{pc}	Proactive charging reward	0	5	0.2	0.8

We train a VF for each month, exploring network sizes $[16, 16]$, $[32, 32]$, $[64, 64]$, learning rates $10^{-3}-10^{-5}$, and batch size 64.

Hardware Used. All the experiments were performed on a 32-core 5.0 GHz machine with 128 GB of RAM.

6.1 Models for Comparison

We evaluate a comprehensive set of six controllers to isolate the contributions of persistence and our proposed approach. For clarity, the daily-decomposed V2B formulation commonly used in prior work is referred to as "Daily V2B" throughout the experiments.

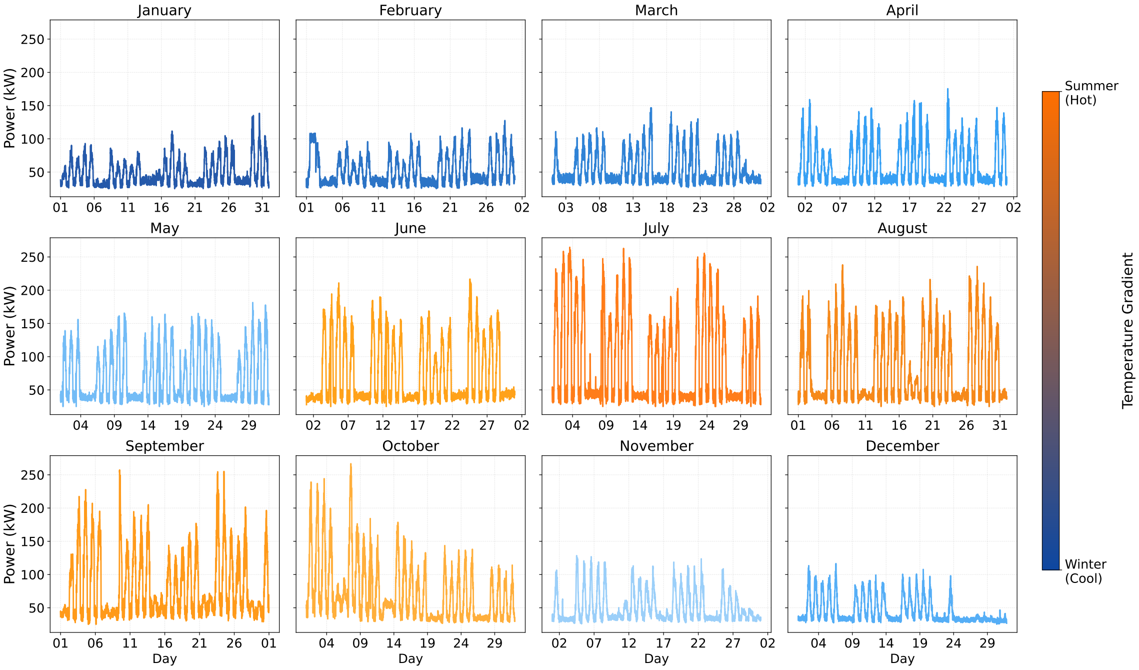

Figure 2: Building electrical load profiles for all 12 months of 2024, highlighting seasonal and temporal variability. Winter months (Jan–Apr, Dec) show lower baselines with moderate peaks, while summer months (Jun–Oct) exhibit higher peak demands due to cooling. Color gradients reflect temperature progression. Daily patterns reveal consistent morning and evening peaks across seasons. These profiles provide realistic test scenarios for evaluating charging coordination algorithms.

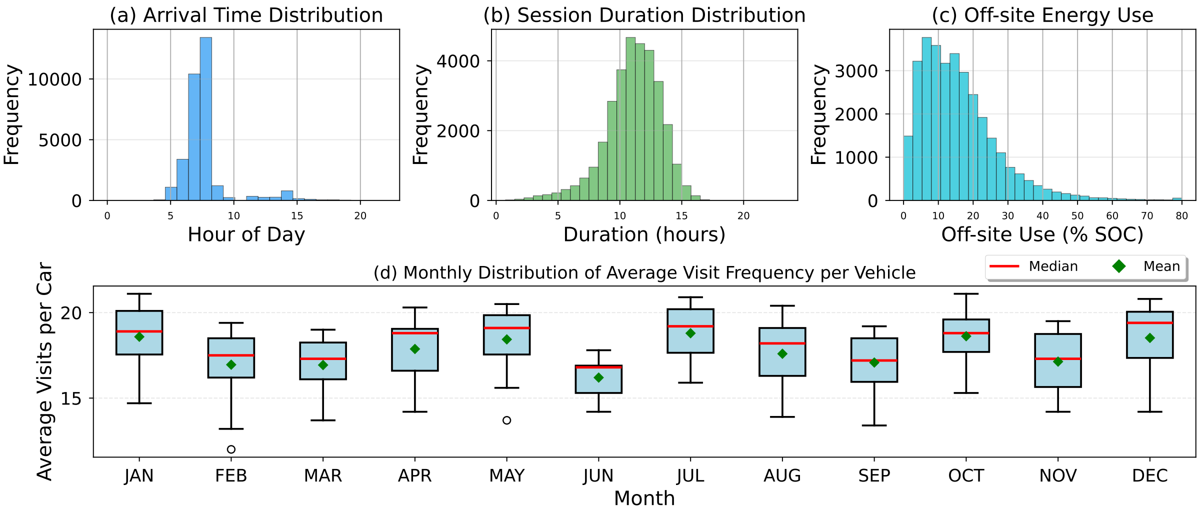

Figure 3: Aggregate EV charging statistics across all scenarios. (a) Peak arrivals occur between 7–9 AM, reflecting workplace charging. (b) Session durations follow a normal distribution centered at 10–12 hours. (c) Off-site energy use is highly variable, indicating diverse driving patterns. (d) Monthly visit frequency is consistent (17–19 visits/vehicle), with June lowest (16.9) and slight increases in January, July, and December.

Proposed Model: Neuro-symbolic MC-MPC+VF (MPC+VF). Our proposed framework from Section 4, which combines a daily-horizon MC-MPC with the strategically-trained Value Function J_{VF}. This is the Neuro-symbolic P-V2Bmodel.

Oracle 1: P-V2B MILP (P-MILP). A full-horizon, deterministic MILP with perfect foresight of all persistent user behavior for the

Table 1: Peak power (kW) measured over a year, where reduced peak demand denotes improved system performance. Gray rows correspond to oracle MILP solutions included for reference only. Lower is better.

Policy	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean
P-MILP	79 ± 2	95 ± 12	90 ± 1	98 ± 2	116 ± 2	144 ± 3	198 ± 5	154 ± 2	147 ± 2	173 ± 5	78 ± 1	77 ± 1	121 ± 3
D-MILP	105 ± 11	102 ± 5	117 ± 3	122 ± 3	140 ± 2	174 ± 2	228 ± 4	178 ± 3	170 ± 3	206 ± 3	97 ± 7	85 ± 2	144 ± 4
MC-MPC+VF	**113 ± 3**	**117 ± 3**	**128 ± 4**	**132 ± 3**	**150 ± 4**	**190 ± 2**	**248 ± 6**	**191 ± 3**	**187 ± 3**	**225 ± 4**	**102 ± 2**	**85 ± 2**	**156 ± 3**
D-RL	115 ± 3	120 ± 8	140 ± 3	145 ± 13	168 ± 0	200 ± 1	252 ± 8	202 ± 9	216 ± 18	246 ± 18	126 ± 15	119 ± 17	171 ± 9
D-MCTS	122 ± 4	134 ± 5	149 ± 3	150 ± 5	142 ± 5	204 ± 2	260 ± 5	209 ± 8	207 ± 2	247 ± 6	128 ± 2	119 ± 2	173 ± 4
D-LLF	133 ± 3	122 ± 3	146 ± 7	171 ± 3	176 ± 4	218 ± 5	261 ± 4	234 ± 5	257 ± 4	264 ± 4	122 ± 3	108 ± 4	184 ± 4
D-FC	147 ± 3	136 ± 2	161 ± 3	185 ± 3	190 ± 3	232 ± 5	275 ± 4	248 ± 5	274 ± 5	279 ± 4	136 ± 3	121 ± 4	199 ± 4

entire month. This represents the theoretical optimum and serves as our primary lower bound.

Oracle 2: Daily V2B MILP (D-MILP). A full-horizon, deterministic MILP that has perfect foresight of all arrivals, but assumes all users are transient (i.e., daily decomposed, and it cannot perform multi-day strategic optimization).

Baseline 1: Daily V2B RL (D-RL). An end-to-end, monolithic Deep Deterministic Policy Guidance (DDPG) based Reinforcement Learning agent trained on daily episodes to directly optimize the daily decomposed V2B problem [13].

Baseline 2: Daily V2B MCTS (D-MCTS). A state of the art online search technique for daily-decomposed V2B, which also has a decentralized version for better scalability [24].

Baseline 3: Heuristics. We include two non-predictive heuristics: (1) **Daily V2B Least Laxity First (D-LLF)**: A smart, non-persistence-aware heuristic [19]. (2) **Daily V2B Fast Charging (D-FC)**: Charge all EVs, v at maximum rate to $\text{SOC}^{\text{req}}(v, k)$ for each session $k \in \mathcal{K}_v$, representing the real-world charging standard.

6.2 Evaluation Metrics

We compare all models across the full dataset (50 samples for 12 months) using the following metrics, which capture the trade-off between cost and service quality. **Peak Power (kW):** A key metric for grid resilience, representing the peak power demand ($P^{\max}$) achieved by the controller. **Total Monthly Bill ($):** This is the economic validation metric, summing energy costs and the final demand charge.

6.3 Experimental Plan

Experiment 1: Baseline Comparison. We run all the models on the complete suite of controllers, including our proposed MC-MPC+VF framework, on the full dataset (30 samples over 12 months). The primary goal is to assess the practical performance of our neuro-symbolic approach against state-of-the-art methods in terms of grid resilience (Peak Shaving) and economic validation (Total Monthly Bill). This comparison is critical for demonstrating that the MC-MPC+VF model can significantly outperform daily-decomposed (daily V2B) control strategies. The results will be analyzed across a year of building load and user profiles to validate robustness under varying demand conditions and temporal variability.

Experiment 2: Ablation Analysis. To assess the contribution of the VF and proactive charging components in our MC-MPC+VF model, we perform targeted ablations on each feature.

1. MC-MPC+VF\VF. The daily-decomposed MC-MPC approach described in section 4, without value function integration. We test the performance our the controller, focusing on proactive charging, and solving for the modified objective:

$$\min_{A_t} \mathbb{E}_{f\in\mathcal{F}} \left[\sum_{k=t}^{t_{\text{eod}}} \left(\phi_k^{\text{energy}} - w_{\text{pc}} \sum_{v\in V_k} r_k^v \Delta t \right) + \Phi_{t_{\text{eod}}}^{\text{demand}} \right] \tag{12}$$

2. MC-MPC+VF\{VF, Proactive Charging (PC)}. The MC-MPC works without any value function approximation and proactive charging. It operates on a 24-hour horizon with no terminal cost and treats all users as transient, solving the MPC according to the following objective:

$$\min_{A_t} \mathbb{E}_{f\in\mathcal{F}} \left[\sum_{k=t}^{t_{\text{eod}}} \left(\phi_k^{\text{energy}} \right) + \Phi_{t_{\text{eod}}}^{\text{demand}} \right] \tag{13}$$

This policy solves a daily V2B problem without foresight beyond the current day. Month-long MC-MPC is omitted due to computational intractability.

Experiment 3: Sensitivity Analysis. To rigorously evaluate the control framework's robustness against forecasting degradation, we subjected the system to two perturbation scenarios using July 2024 as a maximum stress condition. Since the Value Function (J_{VF}) is fixed, the uncertainty is applied by perturbing the Building Load (BL) data used across the entire simulation chain: both the future scenario set $\mathcal{F}$ used by the MC-MPC for decision-making and the final testing building load. We analyzed increasing uniform noise levels of ±10% and ±20% applied simultaneously to both components, moving beyond the nominal ±5% environment. This dual perturbation provides a comprehensive measure of the system's operational performance under realistic, high-volatility conditions.

7 Results

We evaluated the MC-MPC+VF framework against a broad set of controllers (ranging from simple heuristics to perfect-foresight oracles), using four representative months to capture seasonal variation in building load and peak timing. Results show that modeling user persistence via a learned value function significantly improves building cost and peak-shaving performance without sacrificing user satisfaction. Action computation times averaged ∼2 seconds for MC-MPC+VF, ∼1 second for base MC-MPC, a few milliseconds for D-RL, D-LLF, and D-FC, and ∼15 seconds for D-MCTS.

Table 2: Total monthly electricity bill ($) evaluated over representative operating periods, with lower costs reflecting improved system performance. Gray rows correspond to oracle MILP solutions included for reference only. Lower is better.

Policy	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean
P-MILP	7285	7522	8315	8854	10409	10956	14298	12283	11890	11547	7306	6825	9791
D-MILP	7587	7609	8625	9139	10691	11317	14648	12562	12166	11931	7525	6900	10058
MC-MPC+VF	**7717**	**7717**	**8803**	**9288**	**10844**	**11535**	**14922**	**12769**	**12401**	**12197**	**7651**	**6900**	**10238**
D-RL	7718	7831	8912	9425	11041	11636	14958	12856	12728	12415	7877	7301	10392
D-MCTS	7813	8028	10866	9481	8948	11815	15063	12977	12624	12436	7921	7296	10439
D-LLF	7953	7861	8997	9734	11126	11842	15059	13242	13212	12632	7839	7186	10557
D-FC	8118	8039	9178	9909	11309	12026	15239	13411	13406	12817	8013	7324	10732

7.1 Experiment 1: Baseline Comparison

This experiment compares the proposed MC-MPC+VF against state-of-the-art baselines using Reinforcement Learning (RL) [13] and Monte Carlo Tree Search (MCTS) [24] and standard industry practices (Fast Charging).

Peak Shaving Performance. A key goal of the P-V2B framework is to reduce the monthly peak demand charge, the single largest cost factor for commercial buildings. As detailed in the table 1, our approach achieved the lowest mean peak power among all online controllers. The proposed model reduced the mean monthly peak to 156.20 kW (±3.37), representing a significant reduction compared to the most widely used real-world policy, Fast Charging (FC) heuristic (≈ 199 kW), and consistently surpassed complex policy-search methods like the Daily V2B MCTS baseline (≈ 184 kW). This finding confirms that the P-V2B formulation effectively leverages user consistency to achieve superior grid-side flexibility. Crucially, the MC-MPC+VF performance is closest to the theoretical lower bound set by the P-V2B MILP Oracle, for all the months, confirming the success of the hybrid architecture in making near-optimal inter-day strategic decisions under real-time uncertainty.

Economic Impact and User Satisfaction. The significant reduction in peak demand directly translated to substantial financial benefits. In terms of the total monthly bill, from Table 2, the MC-MPC+VF was the highest-performing policy, achieving the lowest mean bill among valid solutions at $10,238.21 (±43.89). This demonstrates an operational cost savings compared to other state-of-the art Daily V2B baseline and significantly undercuts high-cost heuristics. Critically, the MC-MPC+VF model consistently met all user SoC requirements, ensuring a guarantee for off-site usage.

7.2 Experiment 2: Ablation Analysis

An ablation study (Table 3) was conducted to test the core idea that combining MC-MPC with VF enables effective strategic buffering. The buffering is measured by the excess SoC values, which is the energy stored beyond user requirements, each time a user departs. The daily V2B MC-MPC (without the VF or proactive charging) (*Ours\ {VF, PC}*) exhibited a complete myopic nature, maintaining an excess SoC of 0.0% and incurring the highest bill ($10,334.60). Introducing opportunistic overcharging (*Ours\ VF*) provided limited improvements, achieving only 3.80% excess SoC.

The complete MC-MPV+VF (Ours) model achieved the lowest cost ($10,238.21) by maintaining a significantly higher 5.26% excess SoC buffer. This data empirically validates that the VF successfully quantifies the long-term benefit of early charging, effectively advising the MC-MPC+VF to build a large energy reserve strategically. This superior strategic control confirms that the learned Value Function's approximation error (ϵ_{VF}) is less than the inherent structural error of the daily-decomposition model (ϵ_{model}), guaranteeing a better result in expectation.

Table 3: Ablation Analysis of Value Function, and Proactive charging (PC), for a year combined. Lower is better.

Model	Peak Power (kW)	Total Monthly Bill ($)	Excess SoC (kWh)
MC-MPC+VF **(Ours)**	**156.20 ± 3.37**	**10238.21 ± 43.89**	**5.26 ± 1.91**
Ours\ VF	158.13 ± 4.50	10265.53 ± 46.47	3.80 ± 1.87
Ours\ {VF, PC}	163.66 ± 46.58	10334.60 ± 43.04	0.0

7.3 Experiment 3: Sensitivity Analysis

Table 4: Sensitivity analysis on building load estimation, for July 2024 (month with highest peak power). Lower is better.

Model / Noise	Peak Power (kW)	Total Monthly Bill ($)
Baseline Performance (±5% **Internal Noise)**		
MC-MPC+VF **(Ours)**	**248.25 ± 6.49**	**14921.53 ± 46.47**
D-MCTS	260.02 ± 5.16	15063.34 ± 33.19
Perturbation 1: ±10% **Internal Noise**		
MC-MPC+VF (Ours)	**257.82 ± 6.22**	**15043.53 ± 75.90**
D-MCTS	286.89 ± 8.21	15379.42 ± 54.82
Perturbation 2: ±20% **Internal Noise**		
MC-MPC+VF (Ours)	**263.66 ± 9.74**	**15117.60 ± 43.04**
D-MCTS	312.43 ± 7.21	15723.98 ± 65.89

The robustness of the MC-MPC+VF controller was tested using July 2024 (maximum stress condition) by applying high noise (±10% and ±20%) exclusively to the Monte Carlo exploration trajectories, forcing the controller to hedge against volatile predictions, while testing was done against a nominal ±5% ground truth. The results in Table 4 showed the neuro-symbolic framework's superior resilience over the D-MCTS baseline. At ±5% noise, MC-MPC+VF already outperformed D-MCTS (248.25 kW vs. 260.02 kW peak). Under the ±20% noise perturbation, MC-MPC+VF's peak power increased only marginally by 6.2% (to 263.66 kW) and cost rose by 1.3% (to $15, 117.60). In contrast, D-MCTS performance degraded drastically, with its peak power surging by 20.1% (to 312.43 kW) and its total cost rising to $15, 723.98. This disparity confirms that D-MCTS is highly sensitive to prediction errors, whereas the MC-MPC+VF's integrated architecture maintains strong operational performance and firm control on the cost and power use under significant internal uncertainty.

8 Conclusion

We introduced and addressed the Persistent Vehicle-to-Building (P-V2B) problem, a long-horizon stochastic control formulation essential for realistic EV fleet management that tracks EV users' usage. We demonstrated that neglecting this user persistence significantly compromises peak power reduction opportunities. To solve this complex, inter-day coupled problem, we proposed a novel neuro-symbolic framework (MC-MPC+VF) integrating Monte Carlo Model Predictive Control (MC-MPC) with a neurally-learned Value Function (VF).

The MC-MPC+VF framework proved highly effective on real-world data, outperforming all practical baselines (including state-of-the-art Reinforcement Learning and Monte Carlo Tree Search controllers) in both total monthly cost and peak demand reduction. Crucially, it maintained a perfect user-satisfaction record (zero missed SoC requirements). Ablation analysis confirmed that superior performance results from the synergy between the MC-MPC's real-time constraint satisfaction and the VF's long-term strategic guidance. This synergy enables intelligent, non-myopic strategic buffering (proactive charging) to create future energy reserves for peak-shaving. Sensitivity analysis further demonstrated the robustness of the approach to errors in predicting building load. Our findings validate that modeling user persistence is essential, and the proposed neuro-symbolic architecture offers a robust, practical solution that successfully balances formal safety guarantees with strategic foresight. Future work involves extending the framework with advanced VF architectures, adaptive user models, and scaling it for larger, V2G applications.

References

[1] Heba M Abdullah, Adel Gastli, and Lazhar Ben-Brahim. 2021. Reinforcement learning based EV charging management systems–a review. *IEEE Access* 9 (2021), 41506–41531.

[2] Syed Muhammad Ahsan, Hassan Abbas Khan, Sarmad Sohaib, and Anas M Hashmi. 2023. Optimized power dispatch for smart building and electric vehicles with v2v, v2b and v2g operations. *Energies* 16, 13 (2023), 4884.

[3] Dimitri Bertsekas. 2012. *Dynamic programming and optimal control: Volume I*. Vol. 4. Athena scientific.

[4] Enrique Castillo, Roberto Mínguez, A Conejo, and R Garcia-Bertrand. 2006. Decomposition techniques in mathematical programming. *ed: Springer Heidelberg* (2006).

[5] IBM ILOG Cplex. 2009. V12. 1: User's Manual for CPLEX. *International Business Machines Corporation* 46, 53 (2009), 157.

[6] Thomas G Dietterich. 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of artificial intelligence research* 13 (2000), 227–303.

[7] Wladyslaw Findeisen, Frederic N Bailey, Mieczyslaw Brdys, Krzysztof Malinowski, Piotr Tatjewski, and Adam Wozniak. 1980. *Control and coordination in hierarchical systems*. John Wiley & Sons.

[8] S Gupta, A Maulik, D Das, and A Singh. 2022. Coordinated stochastic optimal energy management of grid-connected microgrids considering demand response, plug-in hybrid electric vehicles, and smart transformers. *Renewable and Sustainable Energy Reviews* 155 (2022), 111861.

[9] Akshay Gupte, Shabbir Ahmed, Myun Seok Cheon, and Santanu Dey. 2013. Solving mixed integer bilinear problems using MILP formulations. *SIAM Journal on Optimization* 23, 2 (2013), 721–744.

[10] Zhanwei He, Javad Khazaei, and James D Freihaut. 2022. Optimal integration of Vehicle to Building (V2B) and Building to Vehicle (B2V) technologies for commercial buildings. *Sustainable Energy, Grids and Networks* 32 (2022), 100921.

[11] Hussain Kazmi and Johan Driesen. 2020. Automated demand side management in buildings. In *Artificial Intelligence Techniques for a Scalable Energy Transition: Advanced Methods, Digital Technologies, Decision Support Tools, and Applications*. Springer, 45–76.

[12] Willett Kempton, Jasna Tomic, Steven Letendre, Alec Brooks, and Timothy Lipman. 2001. Vehicle-to-grid power: battery, hybrid, and fuel cell vehicles as resources for distributed electric power in California. (2001).

[13] Fangqi Liu, Rishav Sen, Jose Paolo Talusan, Ava Pettet, Aaron Kandel, Yoshinori Suzue, Ayan Mukhopadhyay, and Abhishek Dubey. 2025. Reinforcement Learning-based Approach for Vehicle-to-Building Charging with Heterogeneous Agents and Long Term Rewards. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems* (Detroit, MI, USA) *(AAMAS '25)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1345–1353.

[14] David Q Mayne. 2014. Model predictive control: Recent developments and future promise. *Automatica* 50, 12 (2014), 2967–2986.

[15] David Q Mayne, James B Rawlings, Christopher V Rao, and Pierre OM Scokaert. 2000. Constrained model predictive control: Stability and optimality. *Automatica* 36, 6 (2000), 789–814.

[16] Graham McClone, Avik Ghosh, Adil Khurram, Byron Washom, and Jan Kleissl. 2023. Hybrid machine learning forecasting for online mpc of work place electric vehicle charging. *IEEE Transactions on Smart Grid* 15, 2 (2023), 1891–1901.

[17] Garth P McCormick. 1976. Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. *Mathematical programming* 10, 1 (1976), 147–175.

[18] Jan Mrkos and Robert Basmadjian. 2022. Dynamic Pricing for Charging of EVs with Monte Carlo Tree Search. *Smart Cities* 5, 1 (2022), 223–240.

[19] Yorie Nakahira, Niangjun Chen, Lijun Chen, and Steven H Low. 2017. Smoothed least-laxity-first algorithm for EV charging. In *Proceedings of the Eighth International Conference on Future Energy Systems*. 242–251.

[20] Oussama Ouramdane, Elhoussin Elbouchikhi, Yassine Amirat, and Ehsan Sedgh Gooya. 2021. Optimal sizing and energy management of microgrids with vehicle-to-grid technology: A critical review and future trends. *Energies* 14, 14 (2021), 4166.

[21] Alessandra Parisio, Evangelos Rikos, and Luigi Glielmo. 2016. Stochastic model predictive control for economic/environmental operation management of microgrids: An experimental case study. *Journal of Process Control* 43 (2016), 24–37.

[22] Warren B Powell and Huseyin Topaloglu. 2006. Approximate dynamic programming for large-scale resource allocation problems. In *Models, Methods, and Applications for Innovative Decision Making*. Informs, 123–147.

[23] Muhammad Salman, Muhammad Arslan, Shoaib Ahmed Khan, Shah Fahad, Muhammad Imran, and Salman Ullah. 2025. Demand-side management and managing electric vehicles and their optimal charging locations and scheduling in smart grids. In *Handbook on New Paradigms in Smart Charging for E-Mobility*. Elsevier, 375–403.

[24] Rishav Sen, Yunuo Zhang, Fangqi Liu, Jose Paolo Talusan, Ava Pettet, Yoshinori Suzue, Ayan Mukhopadhyay, and Abhishek Dubey. 2025. Online Decision-Making Under Uncertainty for Vehicle-to-Building Systems. In *Proceedings of the ACM/IEEE 16th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2025)* (Irvine, CA, USA) *(ICCPS '25)*. Association for Computing Machinery, New York, NY, USA, Article 20, 12 pages. https://doi.org/10.1145/3716550.3722024

[25] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. 2021. *Lectures on stochastic programming: modeling and theory*. SIAM.

[26] Silicon Valley Power. 21-07-2025. Commercial Rates and Fees. https://www.siliconvalleypower.com/businesses/rates-and-fees

[27] Jose Paolo Talusan, Rishav Sen, Ava Pettet, Aaron Kandel, Yoshinori Suzue, Liam Pedersen, Ayan Mukhopadhyay, and Abhishek Dubey. 2024. OPTIMUS: Discrete Event Simulator for Vehicle-to-Building Charging Optimization. In *2024 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 223–230.

[28] Kevin Tanguy, Maxime R Dubois, Karol Lina Lopez, and Christian Gagné. 2016. Optimization model and economic assessment of collaborative charging using Vehicle-to-Building. *Sustainable Cities and Society* 26 (2016), 496–506.

[29] Dimitrios Thomas, Olivier Deblecker, and Christos S Ioakimidis. 2018. Optimal operation of an energy management system for a grid-connected smart building considering photovoltaics' uncertainty and stochastic electric vehicles' driving schedule. *Applied Energy* 210 (2018), 1188–1206.

[30] Juan Pablo Vielma, Shabbir Ahmed, and George Nemhauser. 2010. Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions. *Operations research* 58, 2 (2010), 303–315.

[31] Feng Ye, Yi Qian, and Rose Qingyang Hu. 2015. A real-time information based demand-side management system in smart grid. *IEEE Transactions on Parallel and Distributed Systems* 27, 2 (2015), 329–339.

[32] Liang Yu, Shuqi Qin, Meng Zhang, Chao Shen, Tao Jiang, and Xiaohong Guan. 2021. A review of deep reinforcement learning for smart building energy management. *IEEE Internet of Things Journal* 8, 15 (2021), 12046–12063.

[33] Yu Zheng, Yue Song, David J Hill, and Ke Meng. 2018. Online distributed MPC-based optimal scheduling for EV charging stations in distribution systems. *IEEE transactions on industrial informatics* 15, 2 (2018), 638–649.

Table 5: Table of Symbols

Symbol	**Description**	**Unit / Type**
Indices and Sets		
$t \in \mathcal{T}$	The entire billing period (e.g., one month)	Time
Δt	Time step length	hours
$v \in \mathcal{V}$	Set of all EV users/vehicles	Set
$k \in \mathcal{K}_v$	Set of all charging sessions for user v	Set
N	Number of chargers	Integer
$\omega \in \Omega$	EV clusters for Value function training	-
$\mathcal{F}$	Set of Monte Carlo-sampled future trajectories	Set
t_{eod}	Short-term planning horizon for the MPC	Time
System Parameters		
E_v^{cap}	Battery capacity of EV v	kWh
$\mathrm{SOC}^{ini}(v, 0)$	Initial SoC of EV v at the start of the billing period	fraction of E_v^{cap}
$\mathrm{SOC}^{req}(v, k)$	Required SoC for EV v at the end of session k	fraction of E_v^{cap}
$\Delta E_{v,k}^{ext}$	Estimated external energy consumption of EV v between sessions	kWh
$[t_{v,k}^{arr}, t_{v,k}^{dep}]$	Arrival and departure time of EV v for session k	Time
$min\ r^{\zeta(v,k)}$	Minimum charging rate of the assigned charger (can be < 0)	kW
$max\ r^{\zeta(v,k)}$	Maximum charging rate of the assigned charger	kW
w_t^e	Time-of-use (TOU) energy price at time t	\$/kWh
w^d	Demand charge rate	\$/kW
w^{pc}	Proactive charging reward	\$/kWh
$\hat{P}^{max}$	Estimated monthly peak power (for action refinement)	kW
State Variables		
$\mathbb{S}_t$	Full system state at time t	Vector
$\bar{S}_t$	Aggregate, strategic state used by the Value Function	Vector
b_t	Base (non-EV) building power load at time t	kW
P_t	Total building power (base load + EV charging) at time t	kW
$P_t^{max,hist}$	Historical peak power observed up to time t	kW
$\mathrm{SOC}_t(v)$	State-of-Charge (SoC) of EV v at time t	% or fraction
a_t^v	Availability indicator (1 if EV v is present at time t)	Binary
$\mathcal{E}_{agg}(t)$	Aggregate energy state for all persistent EVs	kWh or SoC
$\mathrm{SOC}^{excess}(\mathcal{V}_\omega)$	Aggregate excess SoC buffer for an EV cluster ω	fraction of E_v^{cap}
$\hat{U}_t(v, k)$	Probabilistic forecast of user v's future trajectory	Distribution
Control / Action Variables		
A_t	Set of joint charging/discharging rates for all EVs at time t	Vector
r_t^v	Charging or discharging rate for EV v at time t	kW
$\tilde{A}_t$	Refined charging/discharging action after refinement step	Vector
$\zeta(v, k)$	Assignment map of EV v (session k) to a specific charger	Mapping
Cost, Objective, and Policy Functions		
ϕ_t^{energy}	Energy cost at time t	\$
$\Phi_{t_{\mathrm{eod}}}^{demand}, \Phi$	Demand charge and Electricity bill over the billing period	\$, \$
R_t	Reward ($\Phi^{\mathrm{demand}} - \hat{P}^{\mathrm{max}}$) at time t	\$
$J_V(\mathcal{S}_t)$	The neural, long-term strategic Value Function (cost-to-go)	\$
π	A charging policy (a sequence of actions A_t)	Policy
$J(\pi)$	Total expected cumulative cost of a policy π	\$
$P^{max}(\pi)$	The peak demand resulting from policy π	kW
Model-Specific and Analysis Terms		
ϵ_{model}	Structural error introduced by the transient (T-V2B) assumption	\$
$\epsilon_{forecast}$	Error from online forecasting of uncertain variables	\$
ϵ_{VF}	Approximation error of the learned Value Function J_V	\$