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

Rishav Sen*, Fangqi Liu*, Jose Paolo Talusan*, Ava Pettet†, Yoshinori Suzue†,
Ayan Mukhopadhyay*, and Abhishek Dubey*
*Vanderbilt University, Nashville, TN, USA
†Nissan Advanced Technology Center - Silicon Valley, Santa Clara, CA, USA

***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 Nissan's Santa Clara office 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.**

***Index Terms*—Vehicle-to-Building, Persistent EV Users, Neuro-Symbolic Control, Monte Carlo Model Predictive Control, Demand Charge Management**

I. 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 which is 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 [1]. 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 [2]–[4].

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 [4] or simpler deterministic approximations [5]), prior work has treated daily EV arrivals as random variables that are independent and identically distributed. However, in practice, EV owners arrive at buildings repeatedly over time, e.g., in a large commercial building, the same set of employees arrive regularly over time. This persistent structure creates a control opportunity: charging above the required SoC on a low-demand day can create discharge flexibility on a later high-peak day. 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. Unlike prior daily-decomposed V2B formulations, P-V2B introduces inter-day coupling through SoC carryover and off-site energy usage, so actions on day d directly affect feasibility and demand charges on day $d+1$, making daily decomposition structurally suboptimal.

Unfortunately, (near) real-time decision-making in the V2B is highly complex and must address several fundamental

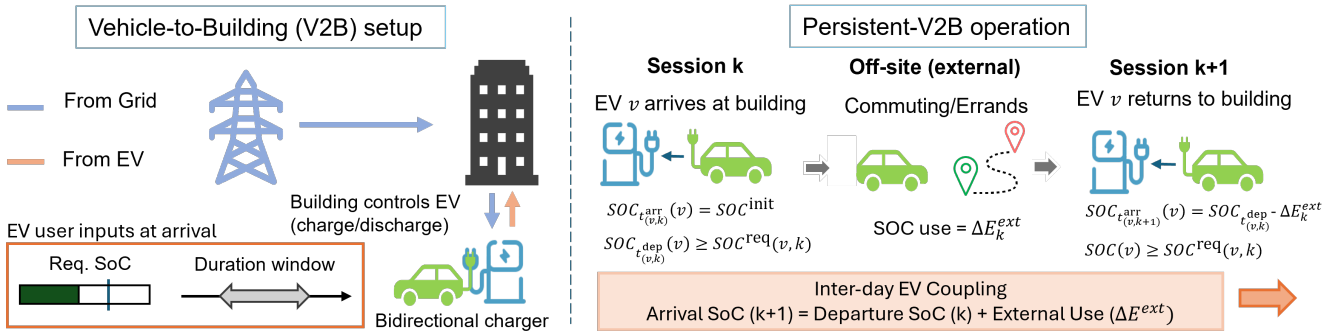

Fig. 1: Persistent V2B framework: user behavior across sessions induces inter-day SoC coupling, enabling proactive charging and later peak shaving.

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 [5]–[7]. *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) [6], [7], Monte-Carlo Tree Search (MCTS) [8], [9], and Reinforcement Learning (RL) [4], [10] face limitations in real-time responsiveness and performance stability in 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. 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 for EV fleets and building energy systems. Unlike prior daily-decomposed V2B formulations, P-V2B captures inter-day coupling through SoC carryover and off-site energy usage, invalidating daily decomposition and enabling new cross-day charging strategies (Section III).

(2) **A neural VF for long-horizon demand-charge prediction.** We design a value function that predicts the impact of current charging decisions on the final monthly demand charge using end-of-day system states, trained online from Monte Carlo sampled trajectories (Section IV-B).

(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 IV-A).

(4) **Real-world evaluation.** Using EV fleet data from Nissan, our framework achieves substantial improvements over MCTS and RL baselines (Section VI).

II. Related Work

The optimization of V2B systems to mitigate demand charges is a complex, long-horizon stochastic control problem [2], [4]. The core challenge lies in balancing real-time operational uncertainties with the temporally-coupled, month-long demand charge objective [11]. 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 [12]. Advanced methods, whether based on online search [13], [14], reinforcement learning [15], [16], or model predictive control [17], [18], 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. In P-V2B, the SoC at the next arrival depends on both the previous departure SoC and the off-site energy usage, so the end of a day is no longer a valid decomposition boundary. Consequently, the transient-fleet assumption used by prior V2B methods excludes feasible cross-day strategies, inducing a structural modeling error.

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 [19], [20]. 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") [22], [23].

The Research Gap: Beyond Daily Decomposition A clear gap emerges at the intersection of these fields. The state-of-the-art in V2B [3], [24] 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 [25]. 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 is enabled by the P-V2B structure: uncertainty lies in exogenous, action-independent state components, while control affects only deterministic SoC evolution, allowing multiple sampled futures to share a common first-stage action through Sample Average Approximation (SAA) [26].

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.

III. 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 which includes the use of persistence information, such as 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.

Persistent EV users and session information. Over the billing period $\mathcal{T}$, the set of EVs $v \in \mathcal{V}$ exhibit persistent behavior, where each session is indexed by $k \in \mathcal{K}_v = \{0, \dots, K_v\}$, and each session lasting $[t^{\text{arr}}_{v,k},\, t^{\text{dep}}_{v,k}] \subseteq \mathcal{T}$. The state of charge (SoC) of EV v at time t is $E_t(v) \in [0,1]$, expressed as a fraction of its battery capacity $E^{\text{cap}}(v)$ (in kWh). Each session has a required departure SoC $E^{\text{req}}(v,k)$ that must be met by $t^{\text{dep}}_{v,k}$. Between sessions, the EV consumes $\Delta E^{\text{ext}}(v,k)$ energy off-site (in kWh), which determines the SoC at the next arrival. We assume $\Delta E^{\text{ext}}(v,k) \le E^{\text{dep}}(v,k) E^{\text{cap}}(v)$ so that off-site usage does not exceed the energy available at departure. Off-site usage may include driving-related SoC depletion and external charging sessions that are not modeled in our system. The first session begins with an initial SoC $E^{\text{ini}}(v)$. In practice, all session-related quantities ($t^{\text{arr}}_{v,k}$, $t^{\text{dep}}_{v,k}$, $E^{\text{arr}}(v,k)$, $E^{\text{req}}(v,k)$, $E^{\text{cap}}(v)$, $\Delta E^{\text{ext}}(v,k)$) must be estimated from historical data or user-reported commuting patterns.

Chargers and charging control. The building can have N_c heterogeneous chargers, which can be bidirectional, supporting controlled charging and discharging, or unidirectional, only supporting controlled charging. Each EV v during session k is mapped to a charger using $\zeta(v,k)$ and the assignment remains throughout its duration $[t^{\text{arr}}_{v,k},\, t^{\text{dep}}_{v,k}]$. The charging rate (in kW) at each time step is $r^{v,k}_t \in [\min r^{\zeta(v,k)}, \max r^{\zeta(v,k)}]$ 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 ease of real-world implementation, we adopt a fixed first-come, first-served (FCFS) assignment rule. We define an EV-session availability indicator $a^{v,k}_t = 1$ if $t \in [t^{\text{arr}}_{v,k}, t^{\text{dep}}_{v,k}]$ and EV v is assigned to a charger in session k, and $a^{v,k}_t = 0$ otherwise.

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

$$P_t \;=\; b_t \;+\; \textstyle\sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}_v} a^{v,k}_t r^{v,k}_t, \quad P_t \ge 0 \qquad \forall t \in \mathcal{T} \quad (1)$$

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

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

$$E_{t+1}(v,k) \;=\; E_t(v,k) \;+\; r^{v,k}_t \;\Delta t / E^{\text{cap}}(v) \qquad (2)$$

The EV user provides a SoC requirement $E^{\text{req}}(v,k)$ to be fulfilled till $t^{\text{dep}}_{v,k}$ for each session, and if feasible, we guarantee to reach the required SoC by departure:

$$E^{\text{dep}}(v,k) \geq E^{\text{req}}(v,k) \tag{3}$$

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

Off-site usage between sessions (persistence coupling). Between session $k-1$ and session k, the EV departs the site and uses energy off-site. The next session's arrival SoC is

$$E^{\text{arr}}(v,k) = E^{\text{dep}}(v,k-1) - \Delta E^{\text{ext}}(v,k-1), \forall k \in \mathcal{K}_v \backslash 0. \tag{4}$$

We assume that $\Delta E^{\text{ext}}(v,k) \leq E^{\text{req}}(v,k-1)$.

Electricity Bill. The electricity bill comprises of the energy cost and demand cost. The energy cost is:

$$\phi^{\text{energy}}_t = w^e_t \cdot P_t \cdot \Delta t, \tag{5}$$

where w^e_t denotes the time-of-use price. The demand charge applies on the maximum power used till time τ:

$$\Phi^{\text{demand}}_\tau = w^d \cdot \max_{j \in [0,\tau]} P_j \tag{6}$$

where w^d is the demand charge rate. Thus, the total electricity bill for building combines energy costs and demand charges over the billing cycle, till t_{end}:

$$\Phi^{\text{total}} = \sum_{t \in \mathcal{T}} \phi^{\text{energy}}_t + \Phi^{\text{demand}}_{t_{\text{end}}}. \tag{7}$$

IV. 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 $\mathbb{S}_t$ is

$$\Big(P^{\max,\text{hist}}_t,\, b_t,\, \{\,\zeta(v,k),\, E_t(v,k),\, E^{\text{req}}(v,k),\, \hat{U}_t(v)\,\}_{v \in \mathcal{V}}\Big),$$

where $P^{\max,\text{hist}}_t = \max_{\tau<t} P_\tau$ is the historical peak load up to time t, b_t is the building load at time t, $\zeta(v,k)$ is the charger currently serving EV v in session k, $E_t(v,k)$ is its state of charge, $E^{\text{req}}(v,k)$ is its required SoC at departure, 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 $A_t = \{\, r^{v,k}_t \,\}_{v \in \mathcal{V}, k \in \mathcal{K}_v}$, and the rate satisfies $r^{v,k}_t \in \big[\, r^{\zeta(v,k)}_{\min},\ r^{\zeta(v,k)}_{\max} \big]$, with $r^{\zeta(v,k)}_{\min}$ and $r^{\zeta(v,k)}_{\max}$ denoting the allowable power range of the charger currently assigned to EV v in session k.

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^{total} is computed according to Eq. (7). Thus the episode reward is $R = -\Phi^{\text{total}}$, reflecting both energy and demand costs for chosen actions.

A. 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_f sampled future trajectories, forming a scenario set $\mathcal{F}$. A single MILP is then solved to minimize the expected cost Φ (Eq. (7)) over the predictive horizon $[t,\, t_{\text{eod}}]$, where t_{eod} denotes the end of the current day. This procedure generates a first-stage action A_t, performing 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. We address this gap with a symbolic controller: 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}. We use $\Phi^{\text{demand}}_{t_{\text{eod}}}$ as a short-horizon surrogate for daily peak shaping, and J_{VF} to capture the residual long-term monthly demand-charge cost-to-go beyond t_{eod}. The controller minimizes the expected short-term energy cost, demand charge until the end of the day t_{eod}, and proactive charging, plus the predicted long-term value:

$$\min_{A_t} \mathbb{E}_{f \in \mathcal{F}} \Bigg[\sum_{j=t}^{t_{\text{eod}}} \left(\phi^{\text{energy}}_j - w_{\text{pc}} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} r^{v,k}_j \Delta t \right) + \Phi^{\text{demand}}_{t_{\text{eod}}} + J_{VF}(\mathbb{S}_{t_{\text{eod}}}) \Bigg] \tag{8}$$

subject to all physical and user constraints (Eqs. 1–4). This objective integrates three elements:

(1) **Short-horizon cost.** The expectation over $f \in \mathcal{F}$ spans N_f sampled future trajectories, capturing short-term energy cost ϕ_j^{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 $-w_{\text{pc}} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} r_j^{v,k} \Delta t$, encouraging slight overcharging beyond immediate SoC needs. The small coefficient w_{pc} is tuned to avoid new peaks, enabling cross-day persistence 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 residual monthly demand charge from t_{eod} to t_{end}, conditioned on $\mathbb{S}_{t_{\text{eod}}}$, providing the long-term foresight that short-horizon MPC alone cannot supply. The structure and training of this value function are detailed in Section IV-B.

B. Value Function Definition and Integration

The value function estimates the remaining monthly demand cost from the end-of-day state.

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

$$\text{VFInput}(\mathbb{S}_t) = \left(t, \widehat{E}_t^{(1)}, \widehat{E}_t^{(2)}, \widehat{E}_t^{(3)}\right), \tag{9}$$

where t is the current day in the billing cycle. 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 during their session k $\widehat{E}_t^{(c)} = \sum_{v \in \mathcal{C}^{(c)}} \max(0, E_t(v,k) - E^{\text{req}}(v,k))$, aggregating SoC above required levels for users in cluster $\mathcal{C}^{(c)}$. These three values capture the cross-day charging buffer 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 residual demand charge from end-of-day to month-end. Thus, $J_{VF}(\text{VFInput}(\mathbb{S}_t)) = \hat{\Phi}_{t_{\text{end}}}^{\text{demand}} - \Phi_{t_{\text{eod}}}^{\text{demand}}$, with the reward signal defined as the residual 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. (8)). 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 $\left(\hat{E}^{(1)}, \hat{E}^{(2)}, \hat{E}^{(3)}\right)$ 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} [27], [28]. This interpolation is linearized using *McCormick envelopes* [29], 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. In our implementation, the best-performing VF is a two-hidden-layer ReLU MLP with architecture [64,64]. Embedding this network directly into the MILP with Big-M constraints led to solve times exceeding 5 minutes per decision epoch, which is incompatible with the 15-minute control interval. The PWL surrogate uses 20 grid points per dimension and reduces solve time to approximately 2 seconds per epoch while maintaining low approximation error (MAE $\approx$ 0.030–0.034; mean relative error $\approx$ 1.2%–1.7%). Appendix **??** reports additional fidelity/runtime details, including a grid-resolution tradeoff table.

C. 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_f future trajectories is generated from the stochastic environment model (detailed in Section VI). 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. (8) 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 boundary $\hat{P}^{\max}$, obtained from oracle solutions of historical data. If the current building power is below this boundary, $P^{\text{diff}} = \hat{P}^{\max} - P_t > 0$, the controller can safely increase charging rates without affecting the monthly demand cost. The refinement step increases charging for eligible EVs v during their session k as follows:

$$\overline{r}_t^{v,k} = \min\left(r_t^{v,k} + \frac{P^{\text{diff}}}{|\mathcal{V}_t|}, \max r^{\zeta(v,k)}\right), \quad \text{if } E_t(v,k) + \frac{\overline{r}_t^{v,k} \Delta t}{E_v^{\text{cap}}} \leq E^{\text{req}}(v,k_t) \tag{10}$$

where $|\mathcal{V}_t|$ is the count of connected EVs.

Algorithm 1: P-V2B Charging Optimization

Input: Initial state S_0, number of samples N_f, 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_f **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. (8)
  Refine action to $\tilde{A}_t = f_{\text{refine}}(A_t, \hat{P}^{\max})$ by Eq. (10)
  Apply $\tilde{A}_t$ and update system state to $\mathbb{S}_{t+1}$;

D. 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. 7) subject to all physical and user constraints (Eq. 1–4). At the end of each day t, we extract the value-function state $\mathrm{VFInput}(\mathbb{S}_t)$ (Eq. 9) and pair it with the oracle residual demand charge $\Phi^{\text{demand}}_{t_{\text{end}}} - \Phi^{\text{demand}}_{t_{\text{eod}}}$. These $(\mathrm{VFInput}(\mathbb{S}_t),\ \Phi^{\text{demand}}_{t_{\text{end}}} - \Phi^{\text{demand}}_{t_{\text{eod}}})$ pairs form a high-quality supervised dataset that trains J_{VF} to approximate the mapping $\mathrm{VFInput}(\mathbb{S}_t) \mapsto \hat{\Phi}^{\text{demand}}_{t_{\text{end}}} - \Phi^{\text{demand}}_{t_{\text{eod}}}$.

(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 $\mathrm{VFInput}(\mathbb{S}_t)$; (ii) continue the simulation to month-end to obtain the realized residual demand charge $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}} - \Phi^{\text{demand}}_{t_{\text{eod}}}$; (iii) store $(\mathrm{VFInput}(\mathbb{S}_t),\ \hat{\Phi}^{\text{demand}}_{t_{\text{end}}} - \Phi^{\text{demand}}_{t_{\text{eod}}})$ in a replay buffer; and (iv) update J_{VF} via supervised regression toward $\hat{\Phi}^{\text{demand}}_{t_{\text{end}}} - \Phi^{\text{demand}}_{t_{\text{eod}}}$, gradually refining its estimate of the long-term cost. This value-iteration refinement aligns the VF with MPC behavior without requiring complete training on MC-MPC.

V. Theoretical Guarantees of the Persistent Model Formulation

A structural optimality theorem showing $\Phi_P \le \Phi_D$ and its proof are provided in Appendix **??**. Here, Φ^*_P and Φ^*_D denote the optimal expected total costs of the persistent P-V2B problem and the daily-decomposed D-V2B problem, respectively.

A. 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 $\mathbb{S}_0$:

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

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 targets the persistent optimum Φ^*_P and avoids the structural error ϵ_{model} introduced by daily decomposition. Let $J^*_P(S)$ denote the optimal persistent cost-to-go from state S, so that $\Phi^*_P = J^*_P(\mathbb{S}_0)$ at the initial state. Its performance is limited by $\epsilon_{\text{forecast}}$ and the VF approximation error, $\epsilon_{VF} = \max_{S\in\mathbb{S}} |J_{\text{VF}}(S) - J^*_P(S)|$ 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 yields the key condition

$$\epsilon_{VF} \le \epsilon_{\text{model}}.$$

under which MC-MPC+VF can outperform the daily baselines.

VI. 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 [30]) to support the persistent user model. All the MILPs are solved using IBM ILOG CPLEX Optimization Studio [31].

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

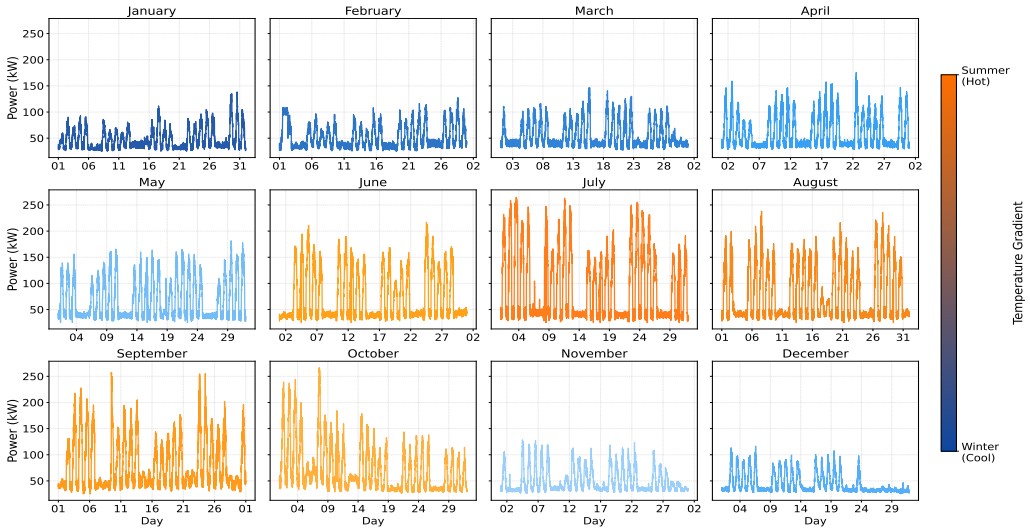

Fig. 2: Building load profiles for all 12 months of 2024, showing seasonal & temporal variability to evaluate charging control.

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 E$\leq$ 83%, then decreases linearly: $-\frac{4}{3}\cdot \mathrm{E} + 130$ for 83%–90%, and $-\mathrm{E} + 100$ 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 via generative models fit to anonymized individual mobility data. For each user, the time between workplace visits, arrival time, stay duration, and off-site energy usage are sampled from fitted probabilistic models conditioned on historical behavior. Each generated sample contains 15 persistent users, creating the cross-session coupling in Eq. 4. Summary statistics 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 [1] 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_t^{base}), 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 $\pm\mathbf{5}\%$ margin is added to the baseline load at each time step. This fluctuating load serves as the ground truth for the controller. Accordingly, our evaluation is a closed-loop, data-driven simulation using real tariffs, a real building load profile, and fleet-behavior generative models derived from industrial data. This is a standard validation setting for V2B/V2G control algorithms prior to field deployment.

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 - 4. 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% 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.

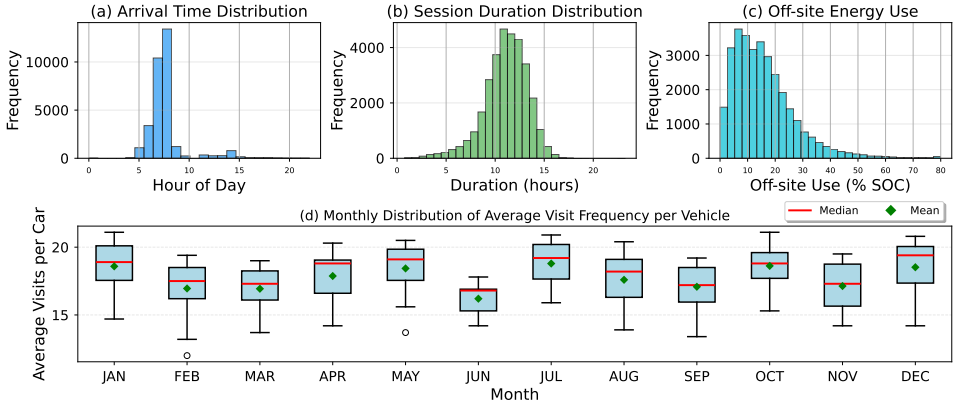

Fig. 3: Aggregate EV charging statistics across all scenarios: arrival time, session duration, off-site energy use, and monthly visit frequency.

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 $[16, 16]$, $[32, 32]$, $[64, 64]$, and learning rates $10^{-3}-10^{-5}$.

A. Models for Comparison

We compare the following seven controllers:

Proposed Model: Neuro-symbolic MC-MPC+VF (MPC+VF). Our proposed framework from Section IV, combining a daily-horizon MC-MPC with the strategically-trained Value Function J_{VF}.

Oracle 1: P-V2B MILP (P-MILP). A full-horizon oracle (MILP) with perfect foresight of all user behavior for the entire month, representing the theoretical optimum(lower bound).

Oracle 2: Daily V2B MILP (D-MILP). P-MILP with user persistence removed, i.e., unable to model inter-day charging.

Baseline 1: Daily V2B RL (D-RL). An end-to-end, monolithic DDPG-based reinforcement learning agent trained on daily episodes to directly optimize the daily-decomposed V2B problem [4]. Its reward matches Eq. (11), i.e., minimizing the daily energy and demand costs.

Baseline 2: Daily V2B MCTS (D-MCTS). A domain-guided Monte Carlo Tree Search (MCTS) technique solving the daily-decomposed V2B problem [9]. Uses $\mathcal{F}$ Monte Carlo sampled scenarios for action generation, similar to our MC-MPC.

Baseline 3: Daily V2B Least Laxity First (D-LLF): A smart, non-persistent heuristic with prioritized EV charging based on the EV's departure time, the current building load, and a load threshold [32].

Baseline 4: Daily V2B Fast Charging (D-FC): Charges all EVs, $v \in \mathcal{V}$ at the maximum rate $\max r^{\zeta(v,k)}$ to reach $E^{\text{req}}(v,k)$ for session $k \in \mathcal{K}_v$ irrespective of the current building load, representing the real-world charging standard.

In addition to these seven controllers, we report a **Building-Only** reference profile (no EV charging or discharging actions) in Tables I and II to quantify the baseline building demand and cost without EV flexibility.

We do not include a persistence-aware RL or MCTS baseline because, under the same P-V2B constraints, such a controller would require a single month-long constrained episode ($30 \times 96 = 2880$ control intervals) with continuous actions and a dominant demand-charge signal revealed only at the terminal step. This creates severe credit-assignment and exploration difficulties and would require substantial algorithmic redesign beyond the state of the art considered here. Instead, we compare against the strongest tractable daily-decomposed baselines, the persistence-aware oracle, and persistence-aware ablations of our own controller.

B. Evaluation Metrics

We compare all models across the full dataset (170 train and 30 test samples for each of the 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.

C. Experimental Plan

Experiment 1: Baseline Comparison. We run all models, including our proposed MC-MPC+VF framework, with the primary goal of assessing 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 significantly outperforms standard 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 contributions of the VF and proactive charging components in our MC-MPC+VF model, we perform targeted ablations.

1. MC-MPC+VF\VF. The daily-decomposed MC-MPC approach described in section IV, without value function integra-

TABLE I: 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
Building-Only	115 ± 2	119 ± 2	131 ± 2	135 ± 2	154 ± 2	194 ± 2	254 ± 3	195 ± 2	191 ± 2	229 ± 2	104 ± 2	87 ± 2	159 ± 2
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

TABLE II: 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
Building-Only	7735	7735	8825	9310	10872	11565	14960	12791	12438	12221	7678	6920	10254
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

tion. We test the controller performance, focusing on proactive charging, and solving for the modified objective:

$$\min_{A_t} \mathbb{E}_{f\in\mathcal{F}} \left[\sum_{j=t}^{t_{\text{eod}}} \left(\phi_j^{\text{energy}} - w_{\text{pc}} \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}_v} r_j^{v,k} \Delta t \right) + \Phi_{t_{\text{eod}}}^{\text{demand}} \right]$$

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:

$$\min_{A_t} \mathbb{E}_{f\in\mathcal{F}} \left[\sum_{j=t}^{t_{\text{eod}}} \left(\phi_j^{\text{energy}} \right) + \Phi_{t_{\text{eod}}}^{\text{demand}} \right] \tag{11}$$

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 $\mathcal{F}$ and the final testing building load. We analyzed increasing uniform noise levels of $\pm 10\%$ and $\pm 20\%$, moving beyond the nominal $\pm 5\%$ environment. This perturbation provides a comprehensive measure of the system's operational performance under realistic, high-volatility conditions.

VII. Results

We evaluated the MC-MPC+VF framework against a broad set of controllers (ranging from simple heuristics to perfect-foresight oracles), using all 12 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 $\sim$2 s for MC-MPC+VF, $\sim$1 s for base MC-MPC, milliseconds for D-RL/D-LLF/D-FC, and $\sim$15 s for D-MCTS. At larger scales, the MC-MPC+VF implementation remains comfortably within the 15-minute decision window, with runtimes of $\sim$6 seconds for 30 chargers (with twice as many users) and $\sim$22 seconds for 45 chargers (with three times as many users).

A. Experiment 1: Baseline Comparison

We compare our persistent MC-MPC+VF against non-persistent RL, MCTS, and heuristic baselines.

Peak Shaving Performance. As shown in Table I, our method achieves the lowest mean peak power among online controllers. The proposed model reduced the mean monthly peak to 156.20 kW ($\pm$3.37), representing a significant reduction compared to the most widely used real-world policy, Fast Charging (FC) heuristic ($\approx$ 199 kW), and consistently surpassed complex policy-search methods like the Daily V2B MCTS baseline ($\approx$ 173 kW). Relative to the Building-Only reference (mean peak $\approx$ 159 kW), MC-MPC+VF remains close to the intrinsic building peak while still serving EV charging demand. 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 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. For context, the Building-Only baseline (i.e., no EV charging/discharging flexibility) yields a mean monthly bill of $10,254 (Table II), indicating that controlled EV participation creates measurable savings over the underlying building load alone. The MC-MPC+VF policy achieves the lowest mean bill among valid online solutions at $10,238.21 ($\pm$43.89), demonstrating operational savings compared to other state-of-the-art Daily V2B baselines and significantly undercutting high-cost heuristics. Critically, MC-MPC+VF consistently met all user SoC requirements, ensuring off-site usability while delivering these savings through better timing of energy delivery rather than reduced charging.

B. Experiment 2: Ablation Analysis

TABLE III: 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** (%)
MC-MPC+VF	**156.20 $\pm$ 3.37**	**10238.21 $\pm$ 43.89**	**5.26 $\pm$ 1.91**
Ours$\setminus$ VF	158.13 $\pm$ 4.50	10265.53 $\pm$ 46.47	3.80 $\pm$ 1.87
Ours$\setminus$ {VF, PC}	163.66 $\pm$ 46.58	10334.60 $\pm$ 43.04	0.0

Table III shows a clear progression from myopic daily MC-MPC to a long term value-guided persistent control. Ours$\setminus\{$VF, PC$\}$ maintains no reserve and yields the highest cost. Adding proactive charging alone provides limited benefit, while adding the VF achieves the lowest bill by maintaining the largest excess-SoC buffer. This supports the claim that the VF captures the long-term value of early charging. Disabling Action Refinement (AR) raises the mean bill to $10,315, indicating AR improves real-time efficiency but is not the primary source of long-horizon gains.

C. Experiment 3: Sensitivity Analysis

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

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

Table IV evaluates the MC-MPC+VF and D-MCTS controllers against degraded forecasts, testing their robustness by applying higher noise ($\pm$10% and $\pm$20%) to July 2024 in both the Monte Carlo sampled future scenarios and the realized building load, moving beyond the nominal $\pm$5% setting. MC-MPC+VF degrades much less than D-MCTS under increasing load uncertainty: at $\pm$20% noise, its peak load rises by only 6.2%, versus 20.1% for D-MCTS. In Table V, we also tested robustness to unexpected user dropout, which constitutes a distribution shift for the learned VF. Here the MPC layer provides immediate feasibility correction: if a user fails to arrive, the controller observes zero availability and re-optimizes over the remaining fleet. Under 5% random dropout over July–September, MC-MPC+VF achieves $13,382 $\pm 1,054$ versus $13,479 $\pm 1,044$ for D-RL over 30 test samples. Under 10% dropout, MC-MPC+VF achieves $13,363 $\pm 1,052$ versus $13,428 $\pm 1,044$ for D-RL, remaining robust to unexpected user behavior.

TABLE V: Robustness under unexpected user dropout over July–September (30 test samples). Lower is better.

Dropout Rate	**MC-MPC+VF ($)**	**D-RL ($)**
5% random dropout	**13,382 $\pm$ 1,054**	13,479 $\pm$ 1,044
10% random dropout	**13,363 $\pm$ 1,052**	13,428 $\pm$ 1,042

VIII. Conclusion

We introduced the Persistent Vehicle-to-Building (P-V2B) problem and showed that neglecting user persistence limits peak-shaving opportunities. To solve this inter-day coupled control problem, we proposed a neuro-symbolic MC-MPC+VF framework that combines real-time constraint satisfaction with long-term value guidance. On real tariffs, real building load, and fleet-derived mobility data, the method outperformed all practical baselines in monthly cost and peak reduction while maintaining zero SoC violations. Ablations showed that value-guided buffering is the main source of improvement, and sensitivity analysis showed robustness to load uncertainty. These results indicate that persistence-aware control is both necessary and practically effective for workplace EV charging. While real-world deployment will require integration with building management systems, communication layers, and operational fail-safes, this paper focuses on the algorithmic contribution and validates it in a closed-loop data-driven setting grounded in real tariffs, building load, and fleet-derived mobility data.

Acknowledgment

This work was supported by Nissan Advanced Technology Center–Silicon Valley, the National Science Foundation (Grants 1952011, 2238815), and the Vanderbilt Center for Sustainability, Energy and Climate. Results were obtained using the NSF-supported Chameleon Testbed. Any opinions, findings, and conclusions are those of the authors and do not reflect the views of the sponsors.

REFERENCES

[1] Silicon Valley Power, "Commercial rates and fees," 21-07-2025. [Online]. Available: https://www.siliconvalleypower.com/businesses/rates-and-fees

[2] W. Kempton, J. Tomic, S. Letendre, A. Brooks, and T. Lipman, "Vehicle-to-grid power: battery, hybrid, and fuel cell vehicles as resources for distributed electric power in california," 2001.

[3] Z. He, J. Khazaei, and J. D. Freihaut, "Optimal integration of vehicle to building (v2b) and building to vehicle (b2v) technologies for commercial buildings," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100921, 2022.

[4] F. Liu, R. Sen, J. P. Talusan, A. Pettet, A. Kandel, Y. Suzue, A. Mukhopadhyay, and A. Dubey, "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*, ser. AAMAS '25. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2025, p. 1345–1353.

[5] K. Tanguy, M. R. Dubois, K. L. Lopez, and C. Gagné, "Optimization model and economic assessment of collaborative charging using vehicle-to-building," *Sustainable Cities and Society*, vol. 26, pp. 496–506, 2016.

[6] G. McClone, A. Ghosh, A. Khurram, B. Washom, and J. Kleissl, "Hybrid machine learning forecasting for online mpc of work place electric vehicle charging," *IEEE Transactions on Smart Grid*, vol. 15, no. 2, pp. 1891–1901, 2023.

[7] Y. Zheng, Y. Song, D. J. Hill, and K. Meng, "Online distributed mpc-based optimal scheduling for ev charging stations in distribution systems," *IEEE transactions on industrial informatics*, vol. 15, no. 2, pp. 638–649, 2018.

[8] J. Mrkos and R. Basmadjian, "Dynamic pricing for charging of evs with monte carlo tree search," *Smart Cities*, vol. 5, no. 1, pp. 223–240, 2022.

[9] R. Sen, Y. Zhang, F. Liu, J. P. Talusan, A. Pettet, Y. Suzue, A. Mukhopadhyay, and A. Dubey, "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)*, ser. ICCPS '25. New York, NY, USA: Association for Computing Machinery, 2025.

[10] H. M. Abdullah, A. Gastli, and L. Ben-Brahim, "Reinforcement learning based ev charging management systems–a review," *IEEE Access*, vol. 9, pp. 41 506–41 531, 2021.

[11] W. Findeisen, F. N. Bailey, M. Brdys, K. Malinowski, P. Tatjewski, and A. Wozniak, *Control and coordination in hierarchical systems*. John Wiley & Sons, 1980.

[12] E. Castillo, R. Mínguez, A. Conejo, and R. Garcia-Bertrand, "Decomposition techniques in mathematical programming," *ed: Springer Heidelberg*, 2006.

[13] F. Ye, Y. Qian, and R. Q. Hu, "A real-time information based demand-side management system in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 329–339, 2015.

[14] M. Salman, M. Arslan, S. A. Khan, S. Fahad, M. Imran, and S. Ullah, "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, 2025, pp. 375–403.

[15] H. Kazmi and J. Driesen, "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, 2020, pp. 45–76.

[16] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A review of deep reinforcement learning for smart building energy management," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12 046–12 063, 2021.

[17] D. Thomas, O. Deblecker, and C. S. Ioakimidis, "Optimal operation of an energy management system for a grid-connected smart building considering photovoltaics' uncertainty and stochastic electric vehicles' driving schedule," *Applied Energy*, vol. 210, pp. 1188–1206, 2018.

[18] O. Ouramdane, E. Elbouchikhi, Y. Amirat, and E. Sedgh Gooya, "Optimal sizing and energy management of microgrids with vehicle-to-grid technology: A critical review and future trends," *Energies*, vol. 14, no. 14, p. 4166, 2021.

[19] D. Q. Mayne, "Model predictive control: Recent developments and future promise," *Automatica*, vol. 50, no. 12, pp. 2967–2986, 2014.

[20] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[21] A. Parisio, E. Rikos, and L. Glielmo, "Stochastic model predictive control for economic/environmental operation management of microgrids: An experimental case study," *Journal of Process Control*, vol. 43, pp. 24–37, 2016.

[22] D. Bertsekas, *Dynamic programming and optimal control: Volume I*. Athena scientific, 2012, vol. 4.

[23] W. B. Powell and H. Topaloglu, "Approximate dynamic programming for large-scale resource allocation problems," in *Models, Methods, and Applications for Innovative Decision Making*. Informs, 2006, pp. 123–147.

[24] S. M. Ahsan, H. A. Khan, S. Sohaib, and A. M. Hashmi, "Optimized power dispatch for smart building and electric vehicles with v2v, v2b and v2g operations," *Energies*, vol. 16, no. 13, p. 4884, 2023.

[25] S. Gupta, A. Maulik, D. Das, and A. Singh, "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*, vol. 155, p. 111861, 2022.

[26] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

[27] J. P. Vielma, S. Ahmed, and G. Nemhauser, "Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions," *Operations research*, vol. 58, no. 2, pp. 303–315, 2010.

[28] A. Gupte, S. Ahmed, M. S. Cheon, and S. Dey, "Solving mixed integer bilinear problems using milp formulations," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 721–744, 2013.

[29] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems," *Mathematical programming*, vol. 10, no. 1, pp. 147–175, 1976.

[30] J. P. Talusan, R. Sen, A. Pettet, A. Kandel, Y. Suzue, L. Pedersen, A. Mukhopadhyay, and A. Dubey, "Optimus: Discrete event simulator for vehicle-to-building charging optimization," in *2024 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2024, pp. 223–230.

[31] I. I. Cplex, "V12. 1: User's manual for cplex," *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009.

[32] Y. Nakahira, N. Chen, L. Chen, and S. H. Low, "Smoothed least-laxity-first algorithm for ev charging," in *Proceedings of the Eighth International Conference on Future Energy Systems*, 2017, pp. 242–251.