Intelligent Engineering and Informatics, Vol.1, 2026

Lightweight Federated Multi Agent Learning for Behind the Meter Resource Coordination: A Survey

Selvam Sambasivam, Edriss Eisa Babikir Adam*

Mainefhi College of Engineering and Technology, Department of Electrical and Electronics Engineering, Eritrea.

E-Mail: dr.edrisseisa@mashreq.edu.sd

Abstract— Small devices that do not reveal private data are required to coordinate behind-the-meter solar systems, batteries, electric cars, and flexible loads; yet, these systems might be advantageous to distribution grids and homeowners alike. This work is an investigation of lightweight federated multi-agent learning for this purpose. From recent literature survey, and categorizes the field based on learning topology, model-efficiency mechanism, coordinated resource, and evaluation realism. Thanks to the evidence, it is clear that personalized and sparse federated reinforcement learning can be deployed more easily than the approach of full-model synchronous training, and that the majority of the multi-agent literature continues to focus on simulations. Safe coordination among non-independent and identically distributed (non-IID) households under intermittent communication, with limited memory and latency budgets, is the central open problem. Thus, the deployment-oriented benchmark and research agenda are suggested.

Keywords—Edge intelligence, federated reinforcement learning, energy management, multi-agent reinforcement learning, privacy control

1. Introduction

Residential energy systems are becoming a part of the power system. The technologies that can shift demand, store excess energy, and give an aggregator and/or local market flexibility include rooftop photovoltaic (PV) generation, home batteries, electric vehicle (EV), heat pumps, air conditioners, water heaters, and smart appliances. The assets are located behind the customer meter and are distributed across homes and are strongly tied to the customer's occupancy, comfort, travel, and appliance habits. Coordination is thus good but hard to accomplish. The key attributes of a successful controller are the ability to handle imperfect price signals and weather data, obey device and network constraints, leverage other homes, and run on a cheap home gateway. Reinforcement learning (RL) is very appealing because it can make sequential decisions without having a physical model, and multi-agent RL (MARL) can model the household or device interactions [1], [2]. The subject is at present particularly active because of three reasons. The first is the inability to cost-minimize adequately a number of peaks at the distribution transformers, given high penetrations of household solar and electric vehicles. Second, data from smart meters and other smart devices are growing more sensitive—central collection can provide details about people's occupancy and mobility.

Third, the system requiring a big cloud model, continuous communication or a powerful graphics processor is not compatible in many homes. Federated learning (FL) enables clients to share information about models instead of raw records, while edge intelligence enables inference to take place near the assets [3], [4]. Recent federated reinforcement learning (FRL) research manages several houses, batteries, and electric-vehicle fleets, as well as peer-to-peer (P2P) markets [5]–[18]. This intersection keeps developing and drives an intent survey of deployable

Journal of Intelligent Engineering and Informatics, Volume 1, Issue 1, Pages 28-39
Corresponding author. E-mail address: dr.edrisseisa@mashreq.edu.sd
Received 10 July 2026; **revised** 17 August 2026; **accepted** 21 August 2026; **published** 01 September 2026.

and distributed learning. This intersection is split across existing reviews. The organization of RL algorithms has largely been addressed by the so-called demand-response (DR) surveys [1] and is related to the concept of personalization and sparsification in federation. To date, energy-system RL reviews have been quite general in nature and do not consider communication and memory specifications as first-order control constraints [2]. Edge intelligence surveys encompass not only control but also forecasting and security operations [3].

The study and analysis of FL in renewable energy focus on collaboration prediction and data sharing instead of a closed-loop multi-resource coordination [4]. The smart-home RL reviews do not necessarily apply any taxonomy to the federated topology, personalization, update sparsity, deployment evidence, etc., but focus on comparing batteries and appliances [19]. Through our three-month, targeted research up to now, we did not identify any survey that connects all of these four axes in a hard behind-the-meter (BTM) limit. This difference is stated in Table 1. Behind-the-meter, home energy management, residential microgrid, electric vehicle, shared storage, peer-to-peer energy, federated learning, reinforcement learning, multi-agent, edge, sparse, personalized, compressed and safe are all search strings. Pure load forecasting, non-invasive load monitoring, front-of-the-meter dispatch, and centralized optimization without learning have not been considered. There are four views linked to the main taxonomy. The learning-topology view distinguishes between local deep reinforcement learning (DRL), centralized training and decentralized execution (CTDE) MARL as well as server-based FRL, personalized or clustered FRL and decentralized peer federation. The efficiency view considers partial or sparse parameter sharing, update scheduling, packing of policy, and personalization to avoid harmful global averaging. Use cases include intra home scheduling, energy trading and shared storage coordination, multi-home aggregation, and peer-to-peer energy and carbon trading. The evidence view categorizes the studies as synthetic simulation, real-trace simulation, hardware-in-the-loop (HIL), testbed, and field operation. Figure 1 correlates these views and makes it impossible to separate privacy claims from their communication and safety costs. Safe, heterogeneity-aware and budgeted coordination is the survey's key research issue.

Table 1. Existing Reviews and Research Gaps

Review	Primary scope	Missing for this topic	Role here
Vázquez-Canteli and Nagy [1]	RL for demand response	Modern FL, edge budgets, MARL deployment	RL baseline
Perera and Kamalaruban [2]	RL across energy systems	Exact BTM boundary and communication accounting	Broad algorithm context
Gooi et al. [3]	Edge intelligence for smart grids	Focused sequential multi-resource coordination	Edge-compute context
Grataloup et al. [4]	FL in renewable-energy applications	Closed-loop FRL/MARL and device safety	Federation context
Pinthurat et al. [19]	RL for smart-home storage	Federated topology, sparsity, inter-home coordination	home energy management system (HEMS) context
This survey	Lightweight federated MARL for BTM resources	Bounded to control, coordination, and deployment evidence	Integrated taxonomy

Table 1 shows that the closest reviews provide essential foundations but do not jointly treat federated topology, multi-agent coordination, edge budgets, and deployment evidence within a behind-the-meter control boundary.

The policy that a household receives needs to be appropriate for the household devices and activities but also needs to contribute to a feeder or community goal. Updates should only be exchanged if there is a reason to coordinate, the number of bytes, energy, and latency are expected to deliver value, and operation remains safe in the event of a client dropping away or a data distribution change. None of the studies reviewed simultaneously show personalization, sparseness, explicit operating requirements and closed hardware assessment. Thus, in the future, we believe that the size of the model as well as the local training energy, the communication volume, the inference latency, the number of constraint violations, and the out-of-distribution performance ought to be reported along with cost and carbon. This deployment test is not just another isolated reward improvement but is the motivation of the paper and the main contribution.

2. Review Scope and Method

2.1. Research Questions and Unit of Analysis

Four research questions (RQs) are answered by this survey. RQ1 focuses on learning topologies to coordinate multiple behind-the-meter actors. RQ2 asks: what is the memory, computation or communication loss associated with studies? RQ3 concerns the usage of assets, objectives, constraints, datasets and experimental settings. RQ4 asks what evidence still needs to be included for practical deployment. The updating unit is a learning-based controller/coordination mechanism, not a forecasting model. For any paper, at least one decision variable influences charging/discharging, appliance scheduling, customer premises market bidding or flexibility provision, directly or indirectly.

2.2 Inclusion and Exclusion Boundary

The paper's three exclusions ensure that it is focused. Forecasting-only FL is not included since predicting the demand does not coordinate action. However, unless customer-side resources are clear participants, large-scale generation and dispatch at the front-of-meter are not included. Finally, papers on blockchain, pricing and optimization without a learning controller are not part of the 30-paper evidence set, but only provide background context. The boundaries differentiate the survey from general smart-grid artificial intelligence (AI) reviews [3], [4] and from home-energy reviews consuming all optimization families [19].

3. Technical Foundations

3.1 Behind-the-Meter Coordination Model

A local state containing the net demand, photovoltaic output, electricity price, indoor conditions, storage state of charge, and availability of electric vehicles is observed by a household agent. The action involves continuous power set-points in addition to, in some systems, discrete market or appliance decisions. These decisions are connected to the grid exchange, as the behind-the-meter power balance of (1) illustrates. Import (positive) grid power is possibly allowed or clipped by the tariff. In addition to the physical model, device limitations, battery state changes, comfort ranges and transformer or feeder limitations are provided.

$$P^{grid}_t = P^{load}_t + P^{EV}_t + P^{ch}_t - P^{dis}_t - P^{PV}_t \tag{1}$$

In (2), the "reward" should not simply be interpreted as relating to the electricity price. The realistic goal integrates energy cost, marginal emissions, contribution to coincident peak, battery degradation, loss of comfort and the cost of exchange of the model. Only a fraction of the studies reviewed report on this, so high levels of reward across papers must be compared with caution. The explicit bytes term is a crucial requirement for a lightweight survey as it transforms implementation details into a part of the control problem.

$$J_i = E[\Sigma\gamma^t(c_t P^{grid}_t + \lambda_c e_t P^{grid}_t + \lambda_p peak_t + \lambda_b deg_t + \lambda_m bytes_t)\ ] \tag{2}$$

3.2. Federated Policy Sharing

Federation brings the separation of local experience collection and shared learning. Selected clients update local policies and send parameters/gradients at round r. Masked aggregation: only selected parameters are stored in each binary/structural mask in Equation (3) and the client contribution is weighted by α. When all masks are one, federated averaging (FedAvg) becomes full-model FedAvg. These include, but are not limited to, partial sharing for FRL [6], personalized layers for personalized federated deep reinforcement learning (PFDRL) [10], adaptive personalized aggregation [12] and sparse EdgeHEM updates [17], which represent various positions in the design space.

$$\theta^{r+1}=\theta^{r}+\Sigma\alpha_i(M^{r}_{i}\odot\Delta\theta^{r}_{i}),i^{\in S_r} \tag{3}$$

3.3 Definition of Lightweight and Safe

Lightweight operation must be measured at inference and training time. A small actor at run time may need a large replay buffer, frequent federation, or a central critic that observes the world. The joint deployment condition of Equation (4) is that network constraints must be satisfied with high probability, the local latency limit for inference must be satisfied and the cumulative communication must not exceed a limit governed by a device budget. For this survey, a method would be considered lightweight if it reduces, or is able to report a reduction in, at least one of the following: the number of model memory entries, multiply-accumulate operations (MACs), update bytes, communication rounds, local training time, or energy consumed without shifting the burden to the server.

$$Pr\{g(x_t,u_t,\xi_t)\leq 0,\ \forall t\}\geq 1-\varepsilon;L^{infer}_{i}\leq\tau_i;\quad \Sigma bytes^{r}_{i}\leq B_i \tag{4}$$

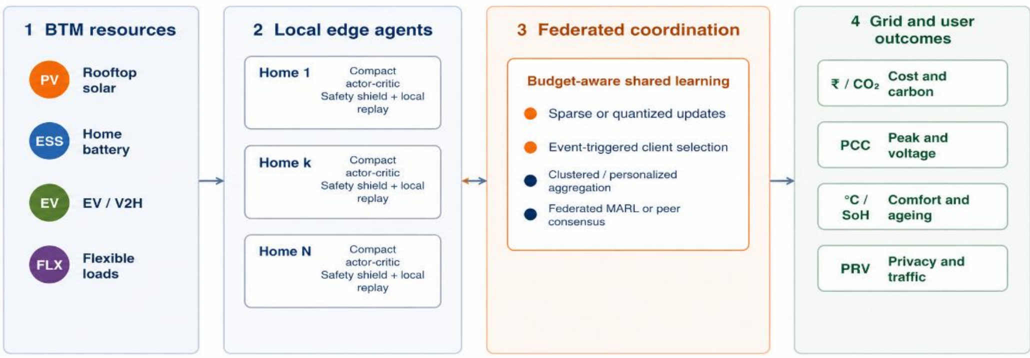

Figure 1. Deployment Framework for Federated Multi-Agent Systems

Figure 1 shows the paper's main synthesis: resource states stay local, compact agents perform closed-loop inference, and only budgeted learning information crosses the communication boundary. The right side makes cost, grid, user, and privacy outcomes co-equal, while the evidence ladder distinguishes simulation claims from operational proof.

4. Classification of Existing Methods

4.1 Local and Centralized Edge Learning Baselines

Single-home DRL sets its lower communication bound because no model exchange is required. In addition, deep deterministic policy-gradient control has been added to harness the heating, ventilation, and air-conditioning (HVAC) and storage capabilities under uncertainty regarding prices and weather conditions [20] and mixed discrete-continuous action models were introduced to accommodate appliances and battery power more easily [21]. The deep reinforcement learning home energy management system (DRL-HEMS) also considers the customer and operator side of the equation [22], while photovoltaic-aware EV charging illustrates how mobility choices and decisions can be linked to local generation [23]. While these methods can be a helpful local baseline, a reward optimization method could be employed independently of these to synchronize charging, shift peaks to different timeframes, or disregard transformer limits. They also have a slow learning rate when one home doesn't have much data. The benefit is that, if trained, the small actor could operate locally with known delays.

The downside is that the benefit to the community has to be applied via an outside price or constraint, not through negotiated cooperation. The typical architecture for MARL is CTDE. A critic gives feedback on joint or aggregated information when training, and each house has a local actor at run time. Entropy-based collective MARL was able to support the home scheduling beyond independent agents [24]. Multi-cluster and mean-field approximations only include a smaller number of explicit pairwise interactions in the trading communities [25], [26]. Attention has been used to help focus a critic on the relevant buildings, and it enhances the DR. Although these have lowered communication during execution, their training stack is not necessarily light: critics grow with the number of agents, replay buffers store joint transitions and retraining may be costly after adding/removing a household.

4.2. Server-Based Federated Reinforcement Learning

Server-based FRL trains local policies and collects averages of model information at an aggregator. Convergence of an early multiple-smart-home framework was faster for heterogeneous homes, without uploading raw consumption data [5]. In turn, shared-storage FRL was enhanced by sharing an arbitrarily-chosen portion of local networks, allowing for a linkage between privacy and communication efficiency [6]. Cooperative multi-residential scheduling learned a policy under time-of-use and demand charges [7] and a federated soft actor-critic (SAC) architecture coordinated electric-vehicle charging, considering user anxiety and community penalties [8]. Common structure learning among homes is actually demonstrated by these studies, with the downside being the assumption of reliable links and compatible models with synchronous rounds. Full averaging is also vulnerable if homes are equipped with different devices and tariffs, or have different objectives.

Local building agents can operate under an energy-management federation layer using hierarchical federated deep reinforcement learning (FDRL) [9]. It is scalable in simulation and has one server that is the point of coordination and trust. Server-based FRL makes sense in cases where there is already a utility or aggregator that interacts with home gateways, but may not be an option where there is a peer community without a central owning entity. Model exchange alone does not offer security. Gradients may contain confidences about training; toxic updates may corrupt the shared policy; secure aggregation introduces compute and traffic that many energy studies do not account for.

4.3. Federated Learning Approaches

Personalization addresses non-IID household data. PFDRL divides the network into shared base layers and residence-specific layers and removes dependence on a cloud service [10]. Adaptive personalized FRL for multiple storage systems introduces, into multiple electricity markets, a novel way of combining global and local policies [12]. A 2025 smart-home approach is another approach used to create personalization of FDRL for heterogeneous buildings [16]. A range of designs across different devices and comfort preferences is more likely than a one-size-fits-all approach. The costs of this are extra local storage, client-specific model selection, and more complicated validation; each client may have been issued a different policy; therefore, an aggregator will need to certify a family of controllers, not one specific model. Decentralized FRL takes the place of or diminishes the importance of the central server. Distributed federation in a local area [10] is adopted in PFDRL and parameter-sharing FRL can be used for peer-to-peer energy trading [18]. Consensus traffic can increase in a neighborhood with higher levels of connectivity and decentralization can raise community governance and lower central trust. It also poses challenges related to membership, malicious peers, and network-partition recovery. There is a practical intermediate approach, clustered aggregation, where clustered homes use a common model, and the clustered home is responsible only for low-dimensional signals, such as clustering houses with similar devices, tariffs, or behavior. This notion is implicit in the energy literature examined, but currently there are no studies that compare the cost of clustering, the cluster drift and the level of safety of any feeder on a common basis.

4.4. Communication-Efficient Model Updates

The best way toward lightweight federation is selective sharing. Shared-storage FRL only exchanges a randomly selected portion of the network [6]. Everyone can improve the broadcasting when they are not too frequent [10] by PFDRL. EdgeHEM specifically aims to support memory-limited edge devices running with sparse FRL and reveals a hardware platform [17]. These methods decrease payload/stored parameters; however, definitions of sparsity vary in the literature, and the number of bytes transmitted per household-day is not often reported. With no denominator, messages can be masked by a decrease in rounds, and local training can be masked by a decrease in model size. For closed-loop behind-the-meter coordination, these tested methods for general edge learning—event-triggered communication, quantization, pruning, knowledge distillation, and parameter-efficient adapters—have only been weakly tested.

A useful trigger will be a combination of gradient norm and expected system value—the change may be significant due to a home's proximity to a voltage limit, or because of the addition of a new EV arrival, which alters the flexibility of the feeder. Co-design is the core research opportunity as a result. The energy policy should not only be selected to be optimal under given constraints on the transformer and comfort; it is important to select this simultaneously with the client selection and compression of the updates under the same constraints of transformer, comfort and communication.

4.5. Federated Multi-Agent Trading and Community Services

The multi-agent problem is coupled with trading as the bid of one household affects the other households' clearing price and flexibility available when they bid. Multi-cluster DRL is used to increase scalability through coordination of groups of peers [25]. Under MARL, an aggregate market signal in lieu of explicit opponent states [26] can be used to model energy conversion and exchange across interconnected microgrids [27]. Local electricity markets are capable of learning the coordinated trading and flexibility services in conjunction [28]. FRL is used in an energy and carbon trading protocol [11] with a privacy-preserving parameter-sharing layer and in a P2P negotiation protocol [18] with a privacy-preserving parameter-sharing layer.

The main constraint is lack of physical abstraction. There are many trading environments which result in an energy-economic clearing and only become a feeder constraint as a penalty. So, a profitable policy might not be possible on an unbalanced low-voltage network. Market fairness is also a function of data volume: houses with more data history or bigger batteries may outperform shared training. For lightweight community learning, it is best to utilize low-dimensional public signals; nonetheless, the network state needs to be exposed enough to prevent unsafe trading. The trade-off between market privacy and grid observability and the cost of communication is not solved by any of the reviewed architectures. The core FRL studies are summarized in Table 2.

Table 2. Overview of Federated Learning Studies

Study	Resources / task	Learning topology	Lightweight mechanism	Evidence	Main limitation
Lee and Choi [5]	PV, energy storage system (ESS), appliances; multi-home scheduling	Server FRL	Shared global policy	Simulation	Full synchronous model exchange
Lee et al. [6]	PV, HVAC, shared ESS	FRL + shared storage agent	Selective parameter sharing	Simulation	Three-building scale
Lee et al. [7]	ESS, renewables, demand charges	Cooperative server FRL	One reusable policy	Simulation	Homogeneous architecture assumption
Chu et al. [8]	plug-in electric vehicle (PEV) fleet, PV, battery energy storage system (battery energy storage system (BESS))	Federated SAC	Attention-weighted sharing	Trace-based case	No hardware network

Gao et al. [10]	Residential standby energy	Decentralized personalized FRL	Shared base + local layers	Real traces	No grid constraint model
Qiu et al. [11]	Energy/carbon P2P trading	Federated MARL	Parameter sharing	Simulation	Market-centric network model
Wang and Dong [12]	Multiple ESS, EV, PV markets	Adaptive personalized FRL	Adaptive global/local mixing	Simulation	Central market assumptions
Tan et al. [13]	Residential microgrid clusters	FRL	Privacy-preserving weights	Simulation	Resource cost not reported
Sievers et al. [14]	Building battery management	FRL	Cross-client generalization	Real-trace simulation	No field communication
Chen et al. [16]	Heterogeneous smart homes	Personalized FDRL	Client personalization	Simulation	Small published evaluation
Li et al. [17]	PV, ESS, appliances	Sparse FRL	Sparse model + memory focus	Hardware testbed	Limited field duration
Ye et al. [18]	Decentralized P2P trading	Federated MARL	Peer parameter sharing	Simulation	Consensus and attack costs open

Table 2 consists of the core FRL studies inside the exact survey boundary. It shows that personalization and selective sharing are increasingly common, but hardware evidence remains exceptional.

5. Evaluation and Application Taxonomy

5.1 Metrics and Benchmark Practice

The practice of evaluation is not standardized. Household cost, peak reduction, carbon and comfort, and battery wear are the primary goals, in that order. Episodic reward or convergence rounds are typically reported as learning metrics. In a small number of cases where the method is described as edge or federated, model size, memory, traffic, latency, or training energy are still not reported. Table 3 therefore partitions operational quality, learning efficiency, and usage evidence. An implementation in a paper must demonstrate the resources consumed to acquire and execute the policy as well as faster convergence in order to establish lightweight operation.

Table 3. Minimum evaluation set for a deployment-oriented comparison.

Evaluation dimension	Minimum metric	Why it matters	Common weakness in the corpus
Economic	Bill, market revenue, worst-household regret	Separates average gain from unfair outcomes	Average cost dominates
Grid	Transformer peak, voltage/thermal violations	Tests whether coordination helps the feeder	Often replaced by a soft penalty
User/device	Comfort violations, EV readiness, battery degradation	Prevents cost savings from shifting harm	Degradation and override rarely combined
Learning	Return across seeds, convergence, out-of-distribution regret	Checks robustness rather than one best run	Single seed or unclear split
Communication	Bytes/client-day, rounds, dropout recovery	Makes lightweight claims measurable	Usually absent
Computation	Model size in megabytes (MB), random-access memory (RAM) peak, MACs, latency, training energy	Determines gateway feasibility	Only EdgeHEM reports hardware-oriented constraints [17]

Privacy/security	Attack success, poisoning tolerance, secure-aggregation overhead	Data locality is not a full privacy test	Mostly qualitative
Evidence	Simulation → traces → HIL/testbed → field	Shows transfer from algorithm to operation	No long-duration field study

Table 3 lists the minimum metrics needed to compare a new method with the reviewed literature. Operational, learning, privacy, and resource measurements must be reported together.

5.2. Evaluation Platforms and Empirical Evidence

Simulation remains dominant. The core FRL studies of multiple homes [5], shared storage [6], and multi-residential tariffs [7] have mainly been studied using software, along with smart microgrids [9], the trading study [11], personalized storage dispatch [12], home residential clusters [13], varying-scale microgrids [15], and decentralized trading [18]. Trace-driven studies provide enhanced realism through measured household or renewable profiles [10] or data-based building studies [14], but a trace does not simulate any communication loss or device timing or inverter behavior or occupant intervention. The notable exception is EdgeHEM, which offers an experimental hardware platform and uses explicit memory constraints [17]. There is no evidence of long-term field trials in live homes in the last set of studies.

5.3. Standardized Benchmarking Framework

Percentage gains are difficult to determine when the benchmark is broken up. Different tariffs, climate files, device sizes, comfort bands, train-test split and baseline controllers are used in the studies. Comparisons include rule-based control, or centralized optimization, or local DRL, or another federation method. A publication-ready benchmark should have the same household partitioning, non-IID data, as well as feeder constraints, communication trace and safety events. Results should cover the mean and dispersion of results across seeds, worst-household cost, constraint violations, number of communications per day, memory, latency of inference, and performance after client dropout or a change in the tariff. The method-family comparison is provided in Table 4.

Table 4. Method-family comparison and implementation judgment.

Method family	Coordination quality	Edge burden	Traffic burden	Heterogeneity	Implementation judgment
Local DRL [20], [21]	Low across homes; strong within one home	Low inference; local replay/training	None	Naturally personalized	Realistic local controller, incomplete community solution
CTDE MARL [24]–[29]	Strong for coupled objectives	High training critic and replay burden	Low at inference; high in training	Handles agents explicitly	Mostly simulation-based
Full-model server FRL [5], [7], [9]	Good shared learning	Moderate local training	High and synchronous	Weak under non-IID homes	Plausible with trusted aggregator, but bandwidth-sensitive
Personalized FRL [10], [12], [16]	Good local/community balance	Extra adapters or local layers	Moderate	Strong	Practical when model validation is managed
Sparse/selective FRL [6], [17]	Good if important parameters retained	Lowest reported memory/payload	Low to moderate	Compatible with personalization	Most deployment-ready; testbed evidence still limited
Decentralized FRL [10], [18]	Strong community autonomy	Peer consensus burden	Topology-dependent	Strong if clustered	Governance-friendly, but security and partitions unresolved

Table 4 compares complete method families rather than isolated papers. Sparse and personalized FRL have the strongest near-term implementation case; CTDE MARL and fully decentralized federation still depend heavily on simulation and controlled communication.

6. Innovative Research Challenge

6.1 Problem Definition

Safe Heterogeneity-Aware Budgeted Federated Coordination (SHB-FC) is the proposed research focus. This is a challenge definition, not necessarily a claim of an accomplished algorithm. It poses the question as to whether it is possible to coordinate various complementary behind-the-meter resources while at the same time meeting four conditions: household personalization, feeder and device safety, bounded edge resources, and robustness to changing participation. Typically, current research addresses two or three aspects. Personalization is covered in [10] and [12] and [16]; sparse edge operation is covered in [17]; and community coordination is covered in [11] and in [18] and [25]–[29]. The two have intersections which are still not closed.

6.2 Candidate Architecture

A good architecture consists of three layers. A small actor and a deterministic safety shield that throws out any action deemed unsafe by device limits are present in each home. There is a small common backbone for transferable knowledge, and a local adapter/final policy layer captures household-specific behavior. The coordination layer allocates clients and parameters using a value-of-information score based on policy novelty, grid urgency, staleness and message size. At the feeder/market layer, one or more signals with low dimensionality provide information on transformer loading, voltage margin, clearing price or flexibility request. This flow is represented in Figure 1, and (3) and (4) represent aggregation and the deployment constraint.

6.3. Evaluation Protocol

Five sources of heterogeneity—asset portfolio, occupant preference, climate, tariff, and communication quality—should be crossed by the benchmark. Client arrival/departure, missing updates, delayed prices, photovoltaic forecast error, and rare safety events should be part of training. Personalization, sparsity, safety shielding and urgency-aware client selection should be carried out one by one in the ablations. The principal comparison does not just concern rewards: it concerns the Pareto frontier between community cost, worst-client regret, violations, bytes, latency and memory. It is only when a method meets the criteria of improving this frontier and performs on unseen households that it can be called deployable. The resulting research roadmap is summarized in Table 5.

Table 5. Research roadmap for the proposed SHB-FC challenge.

Research problem	Required method	Decisive experiment	Success criterion
Changing non-IID homes	Adapters + online clustering	Tariff, occupancy, and device drift	Low worst-client regret after drift
Budgeted federation	Urgency-aware sparse/event updates	Packet loss and byte caps	Better cost-safety frontier per MB
Safe coordination	Local shield + constrained shared policy	Rare feeder and device-limit events	Zero or bounded severe violations
Private and robust updates	Secure aggregation + anomaly filtering	Gradient inversion and poisoning	Low attack success with reported overhead
Deployment transfer	Digital twin, HIL, then occupied-home pilot	Gateway timing and human override	Stable closed-loop operation and rollback

Table 5 translates the open challenge into falsifiable experiments. Each success criterion is measurable and prevents a cost-only result from being presented as deployment readiness.

7. Critical Analysis

The image is obviously set in comparison with the real world. Although implementing standard local DRL does not require communication, it does not learn feeder-wide coordination and can even move peaks from one home to another [20], [21]. In this way, centralized-training/decentralized-execution MARL is quite powerful for coupled markets/building cases [24]–[29] but is primarily a research tool during training time, due to the joint critics, the global replay data, and a meticulously synchronized simulator. Where a home-energy platform is already in use at an aggregator, server-based FRL is more realistic. However, there are several scalability points for full-model synchronous averaging that make it a weak choice for deployment: stable links and similar client models.

Personalized FRL [10], [12], [16] is more convenient for various device portfolios and preferences but stores or validates the shared as well as the local components. The least "ready" family for implementation is the sparse or selective sharer. Partial sharing can be used to reduce exposure and payload [6] and EdgeHEM can be used to treat memory limitations directly and provide a hardware testbed [17]. Even on this scale and for these durations, testbed scale and duration are still far removed from occupied-community operation. Trace-driven studies, such as PFDRL [10] and the sustainable FRL evaluation [14] offer the advantage of greater demand realism than synthetic simulations but they cannot simulate packet loss, gateway failure and occupant override.

Most privacy claims consider only the locality of raw data; they are not testing for gradient inversion or poisoning or membership inference or the cost of secure aggregation. There is a similar separation of safety. Although an energy hub is protected and there is a provision in mixed-action HEMS to reduce unsafe exploration [21], these protections are not integrated with personalized sparse federation. So, the most feasible design that can be implemented in the near future would be for this to be a small local actor with a rule or optimized safety layer, local adapters and event-driven sparse updating to some trusted aggregator. Although fully decentralized MARL is appealing for community governance, it is presently supported primarily by simulation. This is a conservative and operationally viable recommendation.

8. Open Challenges and Future Directions

First, non-IID behavior should be modeled as time-varying, not as a fixed partition; occupancy, EV ownership, and tariffs change after deployment. Second, privacy assessment has to go beyond data locality and include attacks on updates, secure-aggregation overhead, poisoning resilience, etc. Thirdly, during a federation delay, or if the client drops out, safety guarantees are required so that a policy trained with up-to-date information around the globe cannot act wrongly. Fourthly, lightweight claims need a shared accounting of model memory, local compute energy, bytes, rounds and end-to-end latency. Fifth, benchmarks are required to use unbalanced feeder models and a set of traces with various occupied homes; they should not rely on stand-alone building simulators. Sixth, it is important to have fairness communicated via worst-household regret, comfort and the possibility of flexibility revenue as well as the average cost. Lastly, investigations in the field require rollback, human override and audit logs to enable operators to know why an update has been accepted. Local inference should be separated from federation cost in a shared reporting template and results per client should be made public for independent checks of fairness and robustness. Potential avenues are offline-to-online FRL, safe policy distillation, adapter-based personalization, grid urgency event triggers, digital-twin hardware-in-the-loop testing and certificates linking compressed policy updates with the feeder constraints. These directions translate lightweight learning (model compression) to reliable operational coordination.

9. Conclusion

Coordinating behind-the-meter solar, storage, electric vehicles, and flexible demand is possible via lightweight federated multi-agent learning, without centralizing household data. The literature has evolved from local DRL/simulation-based approaches to federated, personalized, decentralized, sparse policies. The most current and most well-balanced approach with respect to heterogeneity and edge feasibility is personalized and selective sharing, and EdgeHEM shows the most hardware-oriented evidence. However, the majority of reported gains come from software simulations, and privacy tends to be treated the same as data locality, and safety is rarely associated with communication constraints. This survey is a deployment-oriented taxonomy, an acuity comparison of evidence, and the challenge SHB-FC, which is personalization, sparse exchange, formal constraints, and experimental realism. Subsequent achievements should be evaluated on the basis of a joint cost-safety-fairness-resource frontier and reproducible hardware and field validation.

References

[1] J. R. Vázquez-Canteli and Z. Nagy, “Reinforcement learning for demand response: A review of algorithms and modeling techniques,” Applied Energy, vol. 235, pp. 1072-1089, 2019, doi: 10.1016/j.apenergy.2018.11.002.

[2] A. T. D. Perera and P. Kamalaruban, “Applications of reinforcement learning in energy systems,” Renewable and Sustainable Energy Reviews, vol. 137, Art. no. 110618, 2021, doi: 10.1016/j.rser.2020.110618.

[3] H. B. Gooi, T. Wang, and Y. Tang, “Edge intelligence for smart grid: A survey on application potentials,” CSEE Journal of Power and Energy Systems, vol. 9, no. 5, pp. 1623-1640, 2023, doi: 10.17775/CSEEJPES.2022.02210.

[4] A. Grataloup, S. Jonas, and A. Meyer, “A review of federated learning in renewable energy applications: Potential, challenges, and future directions,” Energy and AI, vol. 17, Art. no. 100375, 2024, doi: 10.1016/j.egyai.2024.100375.

[5] S. Lee and D.-H. Choi, “Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources,” IEEE Transactions on Industrial Informatics, vol. 18, no. 1, pp. 488-497, 2022, doi: 10.1109/TII.2020.3035451.

[6] S. Lee, L. Xie, and D.-H. Choi, “Privacy-preserving energy management of a shared energy storage system for smart buildings: A federated deep reinforcement learning approach,” Sensors, vol. 21, no. 14, Art. no. 4898, 2021, doi: 10.3390/s21144898.

[7] J.-H. Lee, J.-Y. Park, H.-S. Sim, and H.-S. Lee, “Multi-residential energy scheduling under time-of-use and demand charge tariffs with federated reinforcement learning,” IEEE Transactions on Smart Grid, vol. 14, no. 6, pp. 4360-4372, 2023, doi: 10.1109/TSG.2023.3251956.

[8] Y. Chu, Z. Wei, X. Fang, S. Chen, and Y. Zhou, “A multiagent federated reinforcement learning approach for plug-in electric vehicle fleet charging coordination in a residential community,” IEEE Access, vol. 10, pp. 98535-98548, 2022, doi: 10.1109/ACCESS.2022.3206020.

[9] F. Rezazadeh and N. Bartzoudis, “A federated DRL approach for smart micro-grid energy control with distributed energy resources,” Proc. IEEE 27th Int. Workshop Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pp. 108-114, 2022, doi: 10.1109/CAMAD55695.2022.9966919.

[10] J. Gao, W. Wang, F. Nikseresht, V. G. Rajan, and B. Campbell, “PFDRL: Personalized federated deep reinforcement learning for residential energy management,” Proc. 52nd Int. Conf. Parallel Processing (ICPP), pp. 402-411, 2023, doi: 10.1145/3605573.3605641.

[11] D. Qiu, J. Xue, T. Zhang, J. Wang, and M. Sun, “Federated reinforcement learning for smart building joint peer-to-peer energy and carbon allowance trading,” Applied Energy, vol. 333, Art. no. 120526, 2023, doi: 10.1016/j.apenergy.2022.120526.

[12] T. Wang and Z. Y. Dong, “Adaptive personalized federated reinforcement learning for multiple-ESS optimal market dispatch strategy with electric vehicles and photovoltaic power generations,” Applied Energy, vol. 365, Art. no. 123107, 2024, doi: 10.1016/j.apenergy.2024.123107.

[13] M. Tan et al., “Federated reinforcement learning for smart and privacy-preserving energy management of residential microgrids clusters,” Engineering Applications of Artificial Intelligence, vol. 139, Art. no. 109579, 2025, doi: 10.1016/j.engappai.2024.109579.

[14] J. Sievers et al., “Federated reinforcement learning for sustainable and cost-efficient energy management,” Energy and AI, vol. 21, Art. no. 100521, 2025, doi: 10.1016/j.egyai.2025.100521.

[15] Y. Zhang et al., “Federated deep reinforcement learning for varying-scale multi-energy microgrids energy management considering comprehensive security,” Applied Energy, vol. 380, Art. no. 125072, 2025, doi: 10.1016/j.apenergy.2024.125072.

[16] Z. Chen, J. He, L. Yu, and Q. Zhao, “Personalized federated deep reinforcement learning for smart home energy management,” IFAC-PapersOnLine, vol. 59, no. 9, pp. 193-198, 2025, doi: 10.1016/j.ifacol.2025.08.135.

[17] Y. Li, X. Chen, and Y. Wang, “EdgeHEM: Sparse federated reinforcement learning for home energy management at the edge,” IEEE Transactions on Smart Grid, vol. 16, no. 6, pp. 5602-5614, 2025, doi: 10.1109/TSG.2025.3598438.

[18] Z. Ye, D. Qiu, S. Li, Z. Fan, and G. Strbac, “Federated reinforcement learning for decentralized peer-to-peer energy trading,” Energy and AI, vol. 20, Art. no. 100500, 2025, doi: 10.1016/j.egyai.2025.100500.

[19] W. Pinthurat, T. Surinkaew, and B. Hredzak, “An overview of reinforcement learning-based approaches for smart home energy management systems with energy storages,” Renewable and Sustainable Energy Reviews, vol. 202, Art. no. 114648, 2024, doi: 10.1016/j.rser.2024.114648.

[20] L. Yu et al., “Deep reinforcement learning for smart home energy management,” IEEE Internet of Things Journal, vol. 7, no. 4, pp. 2751-2762, 2020, doi: 10.1109/JIOT.2019.2957289.

[21] C. Huang, H. Zhang, L. Wang, X. Luo, and Y. Song, “Mixed deep reinforcement learning considering discrete-continuous hybrid action space for smart home energy management,” Journal of Modern Power Systems and Clean Energy, vol. 10, no. 3, pp. 743-754, 2022, doi: 10.35833/MPCE.2021.000394.

[22] A. A. Amer, K. Shaban, and A. M. Massoud, “DRL-HEMS: Deep reinforcement learning agent for demand response in home energy management systems considering customers and operators perspectives,” IEEE Transactions on Smart Grid, vol. 14, no. 1, pp. 239-250, 2023, doi: 10.1109/TSG.2022.3198401.

[23] M. Dorokhova, Y. Martinson, C. Ballif, and N. Wyrsch, “Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation,” Applied Energy, vol. 301, Art. no. 117504, 2021, doi: 10.1016/j.apenergy.2021.117504.

[24] Y. Yang, J. Hao, Y. Zheng, and C. Yu, “Large-scale home energy management using entropy-based collective multiagent deep reinforcement learning framework,” Proc. 28th Int. Joint Conf. Artificial Intelligence (IJCAI), pp. 630-636, 2019, doi: 10.24963/ijcai.2019/89.

[25] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, “Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach,” Applied Energy, vol. 292, Art. no. 116940, 2021, doi: 10.1016/j.apenergy.2021.116940.

[26] D. Qiu, J. Wang, Z. Dong, Y. Wang, and G. Strbac, “Mean-field multi-agent reinforcement learning for peer-to-peer multi-energy trading,” IEEE Transactions on Power Systems, vol. 38, no. 5, pp. 4853-4866, 2023, doi: 10.1109/TPWRS.2022.3217922.

[27] T. Chen, S. Bu, X. Liu, J. Kang, F. R. Yu, and Z. Han, “Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning,” IEEE Transactions on Smart Grid, vol. 13, no. 1, pp. 715-727, 2022, doi: 10.1109/TSG.2021.3124465.

[28] Y. Ye, D. Papadaskalopoulos, Q. Yuan, Y. Tang, and G. Strbac, “Multi-agent deep reinforcement learning for coordinated energy trading and flexibility services provision in local electricity markets,” IEEE Transactions on Smart Grid, vol. 14, no. 2, pp. 1541-1554, 2023, doi: 10.1109/TSG.2022.3149266.

[29] J. Xie, A. Ajagekar, and F. You, “Multi-agent attention-based deep reinforcement learning for demand response in grid-responsive buildings,” Applied Energy, vol. 342, Art. no. 121162, 2023, doi: 10.1016/j.apenergy.2023.121162.