POLITECNICO DI TORINO
Repository ISTITUZIONALE

Tackling Heterogeneity in Federated Learning via Variance-Reduced Boltzmann Sampling within Homogeneous Social Coalitions

Original
Tackling Heterogeneity in Federated Learning via Variance-Reduced Boltzmann Sampling within Homogeneous Social Coalitions / Licciardi, A., Raineri, R., Proskurnikov, A., Rondoni, L., Zino, L.. - In: IEEE INTERNET OF THINGS JOURNAL. - ISSN 2327-4662. - (In corso di stampa). [10.1109/JIOT.2026.3746140]

Availability:
This version is available at: 11583/3016853 since: 2026-10-21T11:14:22Z

Publisher:
IEEE

Published
DOI:10.1109/JIOT.2026.3746140

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©9999 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

26 November 2026

Tackling Heterogeneity in Federated Learning via Variance-Reduced Boltzmann Sampling within Homogeneous Social Coalitions

Alessandro Licciardi, *Member, IEEE,* Roberta Raineri, *Member, IEEE,* Anton Proskurnikov, *Senior Member, IEEE,* Lamberto Rondoni, Lorenzo Zino, *Senior Member, IEEE*

***Abstract*—Server-based federated learning (FL) enables privacy-preserving collaborative model training, but its effectiveness is often constrained by heterogeneity in client data distributions. In this paper, we address this limitation by introducing a client selection algorithm that i) dynamically forms non-overlapping coalitions of clients based on asymptotic agreement and ii) selects one representative from each coalition using a variance-reduction criterion. Our approach is inspired by social network modeling, leveraging homophily-based proximity matrices for spectral clustering and informative node selection for estimating a group's aggregate opinion. We provide theoretical insights into the asymptotic behavior of the algorithm under explicit bounded-moment and gradient-alignment assumptions. Empirical validation against a broad set of heterogeneity-aware baselines demonstrates that our method achieves superior or highly competitive accuracy under non-IID data distributions.**

***Index Terms*—Federated Learning, Variance Reduction, Biased Sampling, Social Coalitions**

I. Introduction

FEDERATED learning (FL) has emerged as a powerful paradigm for enabling collaborative model training across multiple agents without directly sharing sensitive local data [1]. By decoupling model optimization from data aggregation, FL offers a principled framework for privacy-preserving and regulation-compliant machine learning, which is particularly appealing in data-sensitive domains such as healthcare [2], [3], finance [4], and IoT applications [5]. As a result, FL attracts significant and growing interest from the scientific community, spurring research on communication efficiency, robustness, personalization, and security in distributed learning systems. In the server-based FL setting considered in this work, a central aggregator coordinates training across distributed clients, enabling efficient model aggregation and scalable knowledge sharing while preserving the client data privacy. IoT ecosystems constitute a particularly compelling application domain for the framework developed in this work. Federations of smart devices and edge gateways operate under tight uplink-bandwidth and energy budgets, communicate over unreliable wireless links, and collect heterogeneous data, since each device senses a specific user, location, or physical process [5], [6]. In such systems, partial participation is a structural constraint rather than a design choice: only a limited number of devices can upload their model updates in each communication round [7]. Moreover, in view of the per-round uplink traffic and the energy consumption [8], the effectiveness of an FL deployment ultimately hinges on how much information the server extracts from each transmitted update. Henceforth, by grouping statistically similar clients into coalitions and sampling from each coalition its most informative representative, the proposed method aims to extract more statistical value from a fixed communication budget, with the goal of reducing the number of rounds required to attain a target accuracy and, with them, the aggregate bandwidth and energy consumption.

A. Licciardi and L. Rondoni are with the Department of Mathematical Sciences "G.L. Lagrange," Politecnico di Torino, Turin, Italy, and Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Turin, Italy ({alessandro.licciardi,lamberto.rondoni}@ polito.it). R. Raineri, A. Proskurnikov, and L. Zino are with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy ({roberta.raineri,anton.proskurnikov,lorenzo.zino}@ polito.it). R. Raineri is also with Modelway S.r.l., Turin, Italy. A.Licciardi and L.Rondoni worked under the auspices of Italian National Group of Mathematical Physics (GNFM) of INdAM. A.Licciardi was supported by the Project Piano Nazionale di Ripresa e Resilienza — Next Generation EU (PNRR-NGEU) from Italian Ministry of University and Research (MUR) under Grant DM 117/2023. L. Rondoni gratefully acknowledges support from the Italian Ministry of University and Research (MUR) through the grant PRIN2022-PNRR project (No. P2022Z7ZAJ) "A Unitary Mathematical Framework for Modelling Muscular Dystrophies" (CUP: E53D2301807 0001). This manuscript reflects only the authors' views and opinions, neither the European Union, the European Commission, nor the Ministry can be considered responsible for them.

A. Literature review

While FL reduces the need to centralize data, improving privacy and regulatory compliance, its effectiveness is often constrained by statistical heterogeneity among clients. In practice, client data are often non-IID, with each dataset following a distinct distribution [9], [10]. This heterogeneity can significantly hinder the performance and convergence of the global model, leading to inconsistent local updates, slower convergence, and reduced model generalization, thus undermining the potential benefits of collaborative learning. To address this challenge, several approaches have been proposed. Regularization methods, such as `FedProx` [9], `FedMD` [11], `Scaffold` [12], and `Mime` [13], constrain local client updates or modify the global aggregation process to improve robustness and convergence in heterogeneous settings. Personalization techniques adapt the global model or learn client-specific models tailored to individual data distributions [14], [15], with notable examples including `Per-FedAvg` [16] and `pFedMe` [17]. Collaboration and clustering strategies identify and leverage similarities among clients by grouping them into clusters, enabling more effective federated training within these clusters. Prominent approaches in this domain include `IFCA` [18], `CFL` [19], `FeSEM` [20]. More recent studies on clustered FL have explored forming groups via consensus-based optimization [21] or by training pairwise discriminators to estimate client similarities [22]. Concurrently,

client-sampling strategies have gained prominence. Practical issues such as intermittent client availability, energy constraints, and limited bandwidth necessitate selecting only a small subset of clients in each round [7]. While uniform sampling serves as a common baseline, growing evidence suggests that a biased selection, taking data heterogeneity into account, can accelerate convergence and improve model quality [23], [24]. Various methods address data heterogeneity by prioritizing clients with higher local loss on the current global model [23] or by promoting diverse client participation [25]. Other approaches focus on variability in client training dynamics [26], data and/or update quality [27], or energy efficiency [8]. Methods such as `FedCBS` [28] prioritize clients to achieve more class-balanced sampling in each training round. More advanced policies, such as `Oort` [29] and `Harmony` [30], employ analytical rules to address heterogeneity in device capabilities and resources. A distinct research thread develops selection policies with reinforcement learning, typically framing client selection as a Markov decision process [31]. Notable examples include `FedRank` [32], `AutoFL` [33], `Favor` [34], and `FedMarl` [35]. In recent years, interest in FL has extended beyond the machine learning community to other disciplines. Game theory provides a rigorous framework for modeling strategic incentives and collaboration among federated learning participants. Early work [36] applied coalition formation games to energy-aware resource management in distributed systems, laying the groundwork for strategic collaboration, though not specifically addressing FL's statistical challenges. Donehue and Kleinberg used hedonic games to analyze coalition stability in federated linear regression, deriving MSE-optimal aggregation schemes, but limited their analysis to linear models [37]. This was extended in [38], which framed `FedAvg` as a coalition game, introducing the price of anarchy to quantify inefficiencies in naive collaboration. The optimal coalition formation algorithm has limited scalability, though submodularity of the cost function suggests potential for efficient approximations. Blum et al. [39] studied equilibria in different FL collaboration structures (*one-for-one* vs. *all-for-all*), characterizing when each is optimal (under stylized utility models). To address dynamic interactions, Ota et al. [40] proposed graphical coalitional games, forming coalitions via synergy measures such as cosine similarity or Improvement Classification Accuracy and analyzing robustness under adversarial settings. Nagalapatti et al. [41] proposed `S-FedAvg` for incentive-aligned client selection, using Shapley values to quantify client utility and exclude unhelpful participants. It is worth noticing that the client selection problem is related to a broad class of *subset selection* problems that have been extensively studied in various contexts and spanning a wide range of domains. Examples include feature selection in machine learning [42]–[44], sensor placement for environmental monitoring [45], smart testing problem to contain the spread of epidemics [46]. A greedy algorithm for selecting the most informative agents in an opinion dynamics framework was proposed in [47], [48].

B. Contribution

In this paper, we introduce a novel client sampling algorithm termed `FedCVR-Bolt` (Federated Coalition Variance Reduction with Boltzmann Exploration) that addresses statistical heterogeneity across client distributions via a two-step procedure, as illustrated in Figure 1. The first step is an adaptive coalition detection phase that dynamically groups clients with similar model states. The second step is a within-coalition selection process that uses a Boltzmann-like probability measure to favor clients with larger intra-coalition variance-reduction scores [49], thereby directly mitigating the effect of heterogeneity. Our main focus is conceptual and methodological: we position client selection in FL through an analogy with homophily-based social selection mechanisms [50], and we integrate established tools, i.e., spectral clustering, similarity-based grouping, Boltzmann exploration, and variance-reduction sampling, into a unified selection strategy. We leverage results from estimation theory [51] to justify the variance-reduction principle underlying our algorithm, building on related insights developed for opinion dynamics in [47] and adapting them to the FL setting. Thus, the novelty of the framework does not lie in its individual methodological components, but in their principled integration through a social-network perspective, which provides a unified interpretation of client similarity, coalition formation and informative client selection. In summary:

i) We propose `FedCVR-Bolt`, a client-sampling algorithm for heterogeneous FL that combines homophily-inspired client similarity, spectral clustering and variance-reduction sampling within a unified framework. Drawing on social-interaction models, we interpret groups of statistically similar clients as coalitions and select in each coalition the most informative client for the global model update.
ii) We provide theoretical insights into the convergence properties of the proposed algorithm under an explicit gradient-alignment assumption, and we characterize its computational complexity.
iii) We extensively evaluate `FedCVR-Bolt` across different heterogeneous settings and standard FL benchmarks [52], showing that the proposed strategy can outperform existing client-sampling baselines under non-IID data distributions.

It is worth clarifying the operating regime for which `FedCVR-Bolt` is designed. Following the standard taxonomy of FL [6], our method primarily targets the *cross-silo* and moderate-scale IoT setting, in which the federation comprises from a few tens up to a few hundreds of clients, such as institutions, organizations, or edge gateways, rather than the extreme *cross-device* regime involving up to millions of participants. In this regime, the cost of forming and clustering a client-similarity matrix (which is common practice in clustered FL [19], [53]) is negligible and, crucially, is incurred *entirely on the server*, which is assumed to be computationally unconstrained: the clients only perform local training and transmit their updates, so that no additional computation, energy, or communication burden is placed on the resource-constrained edge devices. For larger federations, the quadratic cost can be further amortized by keeping the coalitions fixed over several communication rounds before re-clustering; the stochasticity of the Boltzmann sampling preserves within-coalition exploration and prevents the repeated selection of the same client. The same stochasticity makes the policy robust to intermittent

device availability, since restricting the sampling measure to the currently reachable clients of each coalition preserves its structure; while approximate spectral methods such as the Nyström approximation [54] extend applicability to large-scale cross-device deployments. The rest of the paper is organized as follows. Section II introduces the FL framework and the problem statement. A detailed presentation of the algorithm and its pseudocode is given in Section III. The core theoretical insights into our algorithm are presented in Section IV. In Section V, we present the numerical simulations. Section VI concludes the paper and outlines future research directions.

II. Federated Learning

The standard FL framework [1] involves K clients (agents), each holding its own training data. Over T communication rounds, the clients jointly estimate a global vector of D parameters $\theta_{gl} \in \mathbb{R}^D$, often called the global model. Ideally, the estimate should minimize the global loss:

$$\theta_{gl} \in \arg\min_{\theta \in \mathbb{R}^D} \mathcal{L}(\theta), \quad \mathcal{L}(\theta) := \sum_{k=1}^{K} \frac{n_k}{n} \mathcal{L}_k(\theta), \quad (1)$$

where n_k denotes the number of training samples held by the k-th client, $n := \sum_{k=1}^{K} n_k$, and $\mathcal{L}_k(\theta)$ denotes the average empirical loss over the local dataset of client k.

Weighted-average global model: We adopt the standard Federated Averaging (FedAvg) framework [1]. Unlike ensemble methods that simply average final local solutions – which would fail to minimize the global objective $\mathcal{L}(\theta)$ – FedAvg minimizes the global loss iteratively.

Starting from a random initialization of the global model, at each round t, the server sends the current global model $\theta_{gl}(t)$ to the clients. Then, each client performs $S \geq 1$ local training iterations using a stochastic optimizer (e.g., stochastic gradient descent) to minimize their local loss $\mathcal{L}_k$, thus resulting in a locally updated model $\theta_k(t+1)$. Each client communicates the updated model back to the server, which aggregates the updates as a weighted average, namely

$$\theta_{gl}(t+1) = \sum_{k=1}^{K} \alpha_k(t) \theta_k(t+1). \quad (2)$$

where the weights $\alpha_k(t)$ are such that $\sum_{k=1}^{K} \alpha_k(t) = 1$. In general, as in [1], they are set $\alpha_k(t) = \frac{n_k}{n}$ for every $t = 1, \ldots, T$. However, other algorithms propose different weighting schemes, e.g., [55]. By periodically resetting clients to the shared global state, (2) effectively averages the accumulated gradients across clients. Theoretical analysis confirms that this iterative process converges to a stationary point of the global loss function $\mathcal{L}(\theta)$, rather than a suboptimal average of local minima [55]. This setting is natural when all clients aim to estimate the same D-dimensional parameter vector (e.g., they share the same model architecture) using only their local datasets.

Partial participation: Since the number of clients K can be very large, the server typically interacts with only a fraction of them in each round. This partial-participation strategy keeps both the throughput and energy consumption manageable, mitigates stragglers, and thereby shortens each communication round without compromising convergence [1], [56]. Specifically, the global model is communicated to the randomly selected subset of clients $\mathcal{P}_t \subseteq \{1, \ldots, K\}$ whose size $P = |\mathcal{P}_t|$ is kept constant across rounds and coincides with the maximum number of clients that can participate in each communication round, thus reflecting the system's available communication and computational resources. We remark that P also determines the number of coalitions into which clients are grouped (see Section III), so that P/K coincides with the participation rate and K/P with the average coalition size. The method is therefore naturally suited to the limited-participation regime typical of realistic FL deployments: excessively large participation rates would induce over-clustering, with coalitions shrinking to one or two clients, which both undermines the variance-reduction rationale and exposes individual clients; a sensitivity analysis on P is reported in Section V-B. Only these selected clients compute their local updates $\theta_k(t+1)$ and upload them to the server for aggregation. The server updates the global model by aggregating the received estimates:

$$\theta_{gl}(t+1) = \sum\nolimits_{k \in \mathcal{P}_t} \tilde{\alpha}_k(t)\, \theta_k(t+1). \quad (3)$$

where the weights $\tilde{\alpha}_k(t)$ satisfy $\tilde{\alpha}_k(t) \geq 0$ and $\sum_{k \in \mathcal{P}_t} \tilde{\alpha}_k(t) = 1$. These weights differ from the original α_k as they are re-normalized to the active subset $\mathcal{P}_t$.

Statistical learning setup: We now assume that each client's training set is a sample drawn from an underlying data-generating distribution. The parameter vector $\theta_k \in \mathbb{R}^D$ learned by client k (typically, obtained by minimizing its local empirical risk) thus becomes a random variable. Consequently, the vector $\theta_k(t)$ produced by client k in communication round t is a realization of that random variable, which the server then incorporates into the global aggregation. We define the random vector $\theta^d = (\theta_1^d, \ldots, \theta_K^d)^\top \in \mathbb{R}^K$, whose entries are the d-th components of the clients' parameter vectors, with the mean $\bar{\theta}^d = \mathbb{E}(\theta^d)$ and the symmetric covariance matrix $C^d = \mathbb{C}ov(\theta^d) \in \mathbb{R}^{K \times K}$. The entries of C^d quantify client heterogeneity for the d-th model parameter: its diagonal elements capture the variance across clients, while the off-diagonal entries capture pairwise linear statistical dependence between clients' values. Consistent with the aggregation logic of the estimators at communication round t in (2), it is natural to define each global model components as a scalar random variable corresponding convex combination of the local model random variables d-components, i.e., $\theta_{gl}^d = \alpha^\top \theta^d$. We make a standard assumption on the covariance matrix that will be useful in the subsequent analysis and is not restrictive since the diagonal elements refers to clients' variance.

Assumption 1. *The matrix C^d has positive diagonal elements, i.e., $C_{kk}^d > 0$ for every k.*

In our implementation, this holds by construction for the online estimates introduced below (see Appendix **??**).

Heterogeneity in FL: Recent studies show that, under data heterogeneity, the choice of the participating-client set $\mathcal{P}_t$ is critical for convergence as biased sampling policies can significantly accelerate and stabilize training [18], [23]. In the rest of this paper, we provide a theoretically-principled approach to address heterogeneity. We introduce a novel FL

TABLE I
NOTATION SUMMARY.

Symbol	Description
K	number of clients
$\mathcal{K}$	set of clients
$A(t)$	RBF affinity matrix among clients at round t
$W(t)$	row normalization of $A(t)$
T	number of communication rounds
D	number of model parameters
S	number of training iterations
P	number of participating clients
η	learning rate
$\theta_{gl} \in \mathbb{R}^D$	global model
$\theta_{gl}(t) \in \mathbb{R}^D$	global model estimate at round t
$\theta_k(t) \in \mathbb{R}^D$	local model of client k at round t
$\tilde{\theta}_k(t) \in \mathbb{R}^D$	normalized local model of client k at round t
$\mathcal{L}(\theta)$	global loss
$\mathcal{L}_k(\theta)$	client loss
$\mathcal{P}_t$	subset of participating clients at round t
α_k	aggregation weight of client k in the full federation
$\tilde{\alpha_k}(t)$	aggregation weight of k at round t, re-normalized over $\mathcal{P}_t$
$\theta^d_{gl} \in \mathbb{R}$	the d-th component of the global model
$\theta^d \in \mathbb{R}^K$	random vector of d-th components of the model
$\bar{\theta}^d$	mean of random vector θ^d
$\bar{\theta}^d_k(t+1)$	best linear prediction of the d-th model component of client k at round $t+1$, used in the covariance update
C^d	covariance of random vector θ^d
ρ_{jk}	Pearson correlation between random variables j and k
$v^d_k \in \mathbb{R}$	the component-wise variance-reduction score for estimating θ^d_{gl} when sampling θ^d_k
v_k	component-wise aggregate variance-reduction score associated with client k
$\Pi(K)$	set of all partitions of $\{1,\dots,K\}$
$\mathcal{C}_i(t)$	i-th coalition at round t, subset of $\{1,\dots,K\}$
$\pi_p(k;t)$	Boltzmann-like probability that client k is selected in $\mathcal{C}_i(t)$
n_k	number of training samples of k-th client

algorithm with a client selection mechanism inspired by social dynamics models, which performs biased sampling of clients. We demonstrate the algorithm's performance both theoretically and empirically through numerical studies against established benchmarks. We summarize the notation in Table I.

III. Algorithm Design

As discussed in the previous sections, a key challenge in FL is to design an effective partial-participation strategy taking into consideration data heterogeneity. Based on the state-of-the-art literature and building on the statistical learning setup previously introduced, we introduce a novel sampling strategy based on variance reduction techniques. Precisely, we will partition the federation into P groups of cooperating clients whose data are drawn from the same (or closely related) distribution, and then, at each round, we will select the most informative representative from each group using the strategy detailed below. In the following, we present the steps performed at every communication round $t \in \{1,\dots,T\}$, illustrated in Fig. 1. Then, we summarize the procedure in Algorithm 1.

Analogy with Social Networks: We formalize the underlying network structure aimed at studying the problem through the lens of social agents. Here, the set of clients of the federations can be seen as the set of agents $\mathcal{K} = \{1,\dots,K\}$ of an arbitrary network and their individual client model θ_k as the agents' social opinions [57]. Let $\theta(t)$ indicate the vector of current models at round t, such that $\theta_k(t)$ is the model associated to

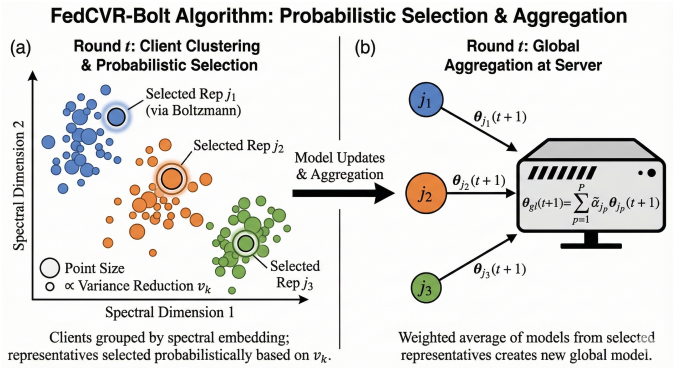


Fig. 1. Representation of `FedCVR-Bolt` sampling and aggregation strategy at round t. *Left.* Clients are grouped according to the affinity of the models distribution into P clusters. Colors denote the clusters $(\mathcal{C}_1,\dots,\mathcal{C}_P)$ identified via spectral clustering. The size of each node is proportional to its variance reduction v_k, representing the informative value of the client's model. One representative is sampled from each cluster using the Boltzmann-like probability $\pi_p(k;t) \propto e^{v_k}$. *Right.* The selected updates $\theta_{j_p}(t+1)$ are aggregated by the server to compute the next global model $\theta_{gl}(t+1)$.

client k at communication round t. Let $W(t)$ be the influence matrix that encodes the pairwise influence between clients at time t. In the context of opinion dynamics, this matrix captures phenomena such as homophily [50], i.e., agents' tendency to be mainly influenced by individuals who hold similar opinions. Following [50], a natural way to implement homophily is to define the affinity between agents k and j as

$$A_{kj}(t) = e^{-\gamma\|\tilde{\theta}_k(t)-\tilde{\theta}_j(t)\|_2^2}, \tag{4}$$

where $\tilde{\theta}_k(t) = \theta_k(t)/\|\theta_k(t)\|$ is the normalized model of client k, and $\gamma > 0$ is a parameter that amplifies the role of models similarities, i.e., the larger γ, the more pronounced is the homophily effect. The influence of agent j on agent k is then obtained by row normalization, $W(t) = D(t)^{-1}A(t)$, with $D(t) = \text{diag}(A(t)\mathbf{1})$, i.e., $W_{kj}(t) = A_{kj}(t)/\sum_{m\in\mathcal{K}} A_{km}(t)$, so that each agent weighs the opinions of the others in proportion to their affinity. The resulting matrix admits a clear social-systems interpretation: agents (clients) with similar opinions (model parameters) exert stronger mutual influence and naturally form clusters. Importantly, computing this matrix, which is standard practice in clustered and personalized FL [19], [22], here reduces to evaluating Euclidean distances between server-available parameter vectors. It is worth noting that we expect clients belonging to the same cluster to converge toward similar local models. Consequently, A, and hence W, will asymptotically exhibit a block-like structure, with each block corresponding to a distinct cluster. This observation supports the specific formulation of these matrices, as it aligns with our objective of accurately identifying and distinguishing network clusters.

Coalition Formation: We move the focus on the coalitions formed which may vary with respect to the previous round. Formally, the goal of this step is to identify the P coalitions formed at round t that will be indicated by $\{\mathcal{C}_1(t),\dots,\mathcal{C}_P(t)\} \in \Pi(K)$, where $\Pi(K)$ denotes the set of all possible partitions of the set $\{1,\dots,K\}$. Notice that the number of communities P is

neither assumed to be the true number of clusters in the data, nor is an output of the clustering step, but it is fixed a priori and it coincides with the number of participating clients per round. This choice is consistent with common practice in clustered FL methods (such as IFCA [18] and FeSEM [20]).

As proposed in [58], a natural way to detect communities in social networks is via spectral clustering, as illustrated in the left panel of Fig. 1. Following this approach, we apply normalized spectral clustering [59] to the symmetric affinity matrix A defined in (4), with $\gamma = 1$. Specifically, clients are embedded into the eigenspace associated with the normalized Laplacian of A, and k-means is then applied in this spectral embedding to identify the P coalitions. The projection onto the eigenvectors space is key since it allows to significantly distinguish similar nodes into more distanced positions in feature space, assigning similar values to members of the same community and capturing soft boundaries between communities, both convex and non-convex.

Sampling Strategy: Once the P coalitions of clients have been identified, the key step of our algorithm consists in determining the most informative client to sample for each coalition. To this end, we use variance reduction [47] as a measure of the informativeness of each client. We adapt the variance-reduction criterion of [47, Proposition 2] to our federated-learning setting. Given θ^d_{gl} the d-th component of the global model to estimate given θ^d_k the sampled model, the associated variance reduction is computed as

$$\mathbb{V}ar(\theta^d_{gl}) - \mathbb{E}[\mathbb{V}ar(\theta^d_{gl}|\theta^d_k)] = \mathbb{V}ar(\mathbb{E}[\theta^d_{gl}|\theta^d_k]), \tag{5}$$

where the second equality comes from law of total variance.

Remark 1. *For the sake of clarity, in the following variance-reduction and linear-prediction arguments we work in centered coordinates, assuming* $\mathbb{E}[\theta^d] = 0$*. This does not affect the variance-reduction criterion in* (5)*, which is translation invariant.*

While computing the conditional expectation can be difficult for non-Gaussian random variables, we can instead approximate it by the L^2-optimal linear projection of θ^d_{gl} onto θ^d_k, that is,

$$\mathrm{Proj}(\theta^d_{gl}|\theta^d_k) = \frac{\mathbb{E}[\theta^d_{gl}\theta^d_k]}{\mathbb{E}[(\theta^d_k)^2]}\theta^d_k \tag{6}$$

This projection is equivalent to the standard least-squares linear regression of θ^d_{gl} on θ^d_k (with zero intercept under the assumption of centered variables). The approximation[1] of the variance reduction (5) can then be defined as

$$v^d_k := \mathbb{V}ar(\mathrm{Proj}(\theta^d_{gl}|\theta^d_k)) = \frac{\left(\mathbb{E}[\theta^d_{gl}\theta^d_k]\right)^2}{\mathbb{E}[(\theta^d_k)^2]}. \tag{7}$$

Applying this to the global model $\theta^d_{gl} = \alpha^\top \theta^d$ and by noticing that $\theta^d_k = \mathbf{e}_k^\top \theta^d$, where $\mathbf{e}_k$ is the k-th canonical basis vector, the (approximate) variance reduction v^d_k is thus found as follows:

$$v^d_k = \frac{(\mathbf{e}_k^\top C^d \alpha)^2}{\mathbf{e}_k^\top C^d \mathbf{e}_k} = \frac{(C^d\alpha)^2_k}{C^d_{kk}}. \tag{8}$$

[1]This approximation is exact when θ^d is Gaussian, since for jointly Gaussian random variables the conditional expectation $\mathbb{E}[\theta^d_{gl} \mid \theta^d_k]$ is affine (and reduces to a linear function when the variables are centered).

So far, we have proposed an explicit formulation for the variance reduction function of the global model θ^d_{gl} for each d-th component, for $d = 1, \dots, D$. We define a component-wise surrogate for the total variance reduction by adding the variance reductions of the individual model components:

$$v_k = \sum\nolimits_{d=1}^{D} v^d_k. \tag{9}$$

Thus, v_k measures how informative client k is for estimating the global model: a larger v_k means that observing this client is expected to reduce more uncertainty about the global model. This component-wise construction, however, neglects correlations between different model components.

We then use the score v_k to select the most informative client in each coalition. Specifically, for each $p = 1, \dots, P$, we choose $j_p = j_p(t)$ such that

$$j_p \in \arg\max\nolimits_{k \in \mathcal{C}_p(t)} v_k \,. \tag{10}$$

Selecting one representative per cluster mitigates overfitting to any single sub-distribution while preserving global diversity. Throughout our work, α in (8) is the vector of full federation weights $\alpha_k = n_k/n$, fixed across rounds. The score v_k thus measures the informativeness of client k for the full federation aggregate $\theta^d_{gl} = \alpha^\top \theta^d$, and it serves as a proxy for the partial participation aggregate (12), whose weights are re-normalized at every round depending on the selected set.

Boltzmann Exploration: It is worth noticing that, in real-world scenarios, the selection rule (10) could become restrictive due to the unavailability of the true data distribution, which introduces an inherent estimation error in the covariance matrix. To mitigate the impact of this uncertainty, while exploiting the informative content of the estimated variance, we incorporate an additional exploration term into the objective.

In particular, instead of a simple greedy, deterministic selection based on the estimated variance, we employ a Boltzmann exploration policy [49], as illustrated in the left panel of Fig. 1. This introduces a controlled stochasticity that helps mitigate the risk of becoming trapped by early biased or noisy estimates of the covariance matrix [60]. Drawing an analogy with statistical mechanics [49], where the probability of a system occupying a specific energy state is proportional to the exponential function of the negative energy divided by temperature, we let the probability of selecting a client k within coalition $\mathcal{C}_p(t)$ during round t based on its associated value v_k be an indicator of the desirability of sampling client k. Higher values indicate greater potential to reduce the variance of the global model estimate. This Boltzmann-like probability measure is defined as:

$$\pi_p(k;t) = \frac{e^{\beta v_k}}{\sum_{j \in \mathcal{C}_p(t)} e^{\beta v_j}}, \tag{11}$$

where $\pi_p(k;t)$ is the probability that client k in coalition $\mathcal{C}_p(t)$ is selected at round t. Here, $\beta > 0$ is related to the inverse temperature of the policy, which interpolates between greedy and uniform selection as β decreases; we fix by design $\beta = 1$, as discussed in Appendix B3. With respect to the standard form of the Gibbs measure of statistical mechanics [49], i.e., $\propto e^{-\beta \times \textbf{Energy}}$, the exponent in (11) carries a positive sign.

as a common practice in statistical decision process theory [61] and reinforcement learning [31] However, interpreting v_k as an energy, Eq. (11) is still a Gibbs measure, but with formally negative temperature, concentrating more probability mass at high energy states. We remark, that such measure is well defined whenever the set of accessible states is bounded, as introduced in Onsager's statistical description of vortices [62], and in small systems with bounded state spaces [63], [64]. In our framework, the states are the finitely many clients of $\mathcal{C}_p(t)$ (see Appendix **??**).Furthermore, since only the product βv_k enters the softmax law, β may be assigned reciprocal units to v_k, making βv_k dimensionless. This design assigns clients with larger estimated variance-reduction v_k proportionally higher selection probabilities, aligning sampling with their expected contribution within each coalition. Moreover, this framework keeps nonzero probability for lower-scoring clients, preventing myopic exploitation and enabling gains under uncertainty. Since $\pi_p(k;t)$ is strictly positive, no client is excluded from selection at any round; together with the uniform-sampling warm-up phase (Appendix **??**), this promotes broader coverage of the client population.

Update Policy: Once the participating clients $\mathcal{P}_t = \{j_1, \dots, j_P\}$ have been selected, the server sends them the current model. Then, they locally update the central model according to their local dataset, obtaining the individual update $\theta_{j_p}(t+1)$ which is sent to the server. Finally, it coherently updates the global model as

$$\theta_{gl}(t+1) = \sum\nolimits_{p=1}^{P} \tilde{\alpha}_{j_p}(t)\theta_{j_p}(t+1), \tag{12}$$

as illustrated in the right panel of Fig. 1. Because the participating clients are sampled nonuniformly, the aggregate in (12) is generally a biased estimator of the corresponding full-participation update in (2).

Within each coalition $\mathcal{C}_p(t)$, only the selected client j_p performs a local update. Hence,

$$\theta_k(t+1) = \begin{cases} \theta_{j_p}(t+1), & k = j_p, \\ \theta_k(t), & k \in \mathcal{C}_p(t) \setminus \{j_p\}. \end{cases}$$

To update the estimate of the covariance matrix used in the next round, an update policy for the expected value of θ_k must be formalized.

Proposition 1. *Under the centering convention of Remark 1, let $\theta_{j_p}(t+1)$ be the sampled model at round $t+1$. The best linear predictor (in the L_2 sense) of the d-th component θ_k^d for any client $k \in \mathcal{C}_p(t)$, given the observation $\theta_{j_p}(t+1)$, is*

$$\bar{\theta}_k^d(t+1) = \tilde{\rho}_{kj_p}^d(t)\theta_{j_p}^d(t+1), \tag{13}$$

for all components d, where the linear regression coefficient $\tilde{\rho}_{kj_p}(t)$ is obtained via normalization of the Pearson correlation coefficient between the models of clients k and j_p, that is,

$$\tilde{\rho}_{kj_p}^d(t) = \rho_{kj_p}^d(t)\frac{\sqrt{C_{kk}^d(t)}}{\sqrt{C_{j_pj_p}^d(t)}}, \quad \rho_{kj_p}^d(t) = \frac{C_{kj_p}^d(t)}{\sqrt{C_{kk}^d(t)}\sqrt{C_{j_pj_p}^d(t)}}.$$

Proof. Omitting the time dependence for brevity, the linear projection of θ_k^d onto $\theta_{j_p}^d$ in the mean-square (L_2) sense is

$$\bar{\theta}_k^d = \mathrm{Proj}(\theta_k^d|\theta_{j_p}^d) = \frac{\mathbb{E}[\theta_{j_p}^d\,\theta_k^d]}{\mathbb{E}[(\theta_{j_p}^d)^2]}\theta_{j_p}^d = \frac{C_{kj_p}^d}{C_{j_pj_p}^d}\theta_{j_p}^d. \tag{14}$$

It remains to notice that from Pearson correlation coefficient definition it holds $\rho_{kj_p} = \frac{C_{kj_p}^d}{\sqrt{C_{kk}^d}\sqrt{C_{j_pj_p}^d}}$. □

Online estimation of covariance: Final step is now the online estimation of the covariance matrix C^d of the individual model's components which represent a central core of our method. The entries of this matrix are given by

$$C_{kj}^d = \mathbb{C}ov(\theta_k^d, \theta_j^d) = \mathbb{E}[(\theta_k^d - \bar{\theta_k^d})(\theta_j^d - \bar{\theta_j^d})]\,. \tag{15}$$

Motivated by the structure of the entries of the covariance matrix in (15), we recursively average the available residual outer products, obtaining the *online covariance estimate* $C^d(t)$ as $C^d(t+1) = (1-\gamma_t)C^d(t) + \gamma_t[\theta^d(t+1) - \bar{\theta^d}(t+1)][\theta^d(t+1) - \bar{\theta^d}(t+1)]^\top$, with $\gamma_t = 1/(t+1)$ and $\bar{\theta^d}(t+1)$ the mean vector updated according to (13). With this choice, $C^d(t)$ is the running average of the initialization $C^d(1)$ and of the residual outer products (see Appendix **??**). Observe that $C^d(t)$ is intended as an online estimate of the covariance C^d of the evolving client parameters defined in Section II, and not a posterior uncertainty after observing a representative; Proposition 1 is used only to fill in the unobserved clients before the estimate is updated.

Remark 2. *The recursion is of Robbins–Monro type [65], [66], and more general step sizes $\gamma_t \in (0,1)$ with $\sum_t \gamma_t = \infty$ and $\sum_t \gamma_t^2 < \infty$ can be used. These conditions alone, however, do not ensure that $C^d(t)$ converges to C^d, since stochastic-approximation guarantees also require conditions on the driving process [67]. If the residuals are stationary and ergodic, $C^d(t)$ converges to their second-moment matrix, which coincides with C^d if, additionally, the residual outer products are unbiased estimators of C^d.*

All the steps of the algorithm, extensively discussed in the above, are summarized in Algorithm 1.

Remark 3 (Privacy Guarantees)**.** *Let us notice that the privacy of client data in `FedCVR-Bolt` algorithm is preserved by adhering to the fundamental principles of FL: raw data remains on client devices and is never transmitted to the server. Clients selected from the set $\mathcal{P}_t$ perform local updates using the global model $\theta_{gl}(t)$ to produce their updated local models $\theta_{j_p}(t+1)$. Moreover, clients subsequently communicate to the server only these model parameters, not the underlying private data. To further protect such parameters during aggregation steps (e.g., for the computation of $\theta_{gl}(t+1)$), it is possible to employ secure aggregation protocols [68]. While the distinctive server-side operations of `FedCVR-Bolt`, such as clustering based on individual models $\{\theta_k(t)\}$ and deriving $\bar{\theta}_k(t+1)$ from specific client model updates $\theta_{j_p}(t+1)$, require the server to access these individual parameters for its core functionality, this architectural choice is consistent with federated architectures where the central server orchestrates the learning process using the model parameters received from the clients [6] for advanced tasks like personalization or clustering [14].*

Algorithm 1 `FedCVR-Bolt` Algorithm

Inputs: K, D, T, $\{\theta_k(1)\}_{k=1}^K$, $\{C^d(1)\}_{d=1}^D$, $\theta_{gl}(1)$, γ_t, P
for $t = 1, \dots, T-1$ **do**
  Normalize: $\tilde{\theta}_k(t) = \theta_k(t)/\|\theta_k(t)\|$
  Compute affinity matrix: $A_{kj}(t) = e^{-\gamma\|\tilde{\theta}_k(t)-\tilde{\theta}_j(t)\|_2^2}$
  Set: $\{\mathcal{C}_i(t)\}_{i=1}^P \leftarrow \text{SpClustering}(A(t), P)$
  Compute v_k^d with $C^d(t)$ and $\alpha_k = n_k/n$; set $v_k = \sum_{d=1}^D v_k^d$
  Compute $\pi_p(k;t) = e^{v_k}/\sum_{j\in\mathcal{C}_p(t)} e^{v_j}$
  Sample $j_p \sim \pi_p(\cdot;t)$ in each cluster $\mathcal{C}_p(t)$
  Set $\mathcal{P}_t = \{j_1, \dots, j_P\}$
  for all $j_p \in \mathcal{P}_t$ **do**
    Receive $\theta_{gl}(t)$, perform local update $\rightarrow \theta_{j_p}(t+1)$
  end for
  for all $p = 1, \dots, P$, $k \in \mathcal{C}_p(t)$ and $d = 1, \dots, D$ **do**
    $\bar{\theta}_k^d(t+1) = \tilde{\rho}_{kj_p}^d(t)\theta_{j_p}^d(t+1)$
  end for
  for $d = 1, \dots, D$ **do**
    $C^d(t+1) = (1-\gamma_t)C^d(t) + \gamma_t\left(\theta^d(t+1) - \bar{\theta}^d(t+1)\right)\left(\theta^d(t+1) - \bar{\theta}^d(t+1)\right)^\top$
  end for
  $\theta_{gl}(t+1) = \sum_p \tilde{\alpha}_{j_p}(t)\theta_{j_p}(t+1)$, $\tilde{\alpha}_{j_p}(t) = \frac{n_{j_p}(t)}{\sum_p n_{j_p}(t)}$
end for

IV. Theoretical Analysis

A. Asymptotic Properties

We derive insights into the asymptotic properties of `FedCVR-Bolt`. Our analysis builds on assumptions that are standard in the theoretical analysis of stochastic optimization and Federated Learning [55], [69], [70]: smoothness of the global loss and bounded variance of the stochastic federated gradient. We extend these with a mild alignment condition on the client selection policy.

The latter ensures that, in expectation, the update direction preserves a positive correlation with the true gradient, thereby maintaining descent. Under these assumptions, Theorem 1 establishes that the time-averaged expected squared gradient norm is asymptotically bounded by a quantity of order $O(\eta)$, where η is the learning rate, which can be chosen sufficiently small. This differs from stronger notions of convergence of the estimates $\theta_{gl}(t)$ to a stationary point, which is precluded by the intrinsic variance of stochastic updates.

For simplicity, and as common in the analysis of biased stochastic-gradient methods [69], [71], we present the result for a single local step, $S = 1$. In this case, the global update is $\theta_{gl}(t+1) = \theta_{gl}(t) - \eta\, g(t)$, where

$$g(t) := \sum\nolimits_{k\in\mathcal{P}_t} \tilde{\alpha}_k(t)\nabla\mathcal{L}_k(\theta_{gl}(t)) \tag{16}$$

is the federated gradient generated by the clients selected at round t. The extension to multiple local steps, $S > 1$, is discussed in Appendix **??**.

In the following, $\mathbb{E}_t[\cdot]$ denotes the conditional expectation given all information available at the beginning of round t, including the current global model, covariance estimates, coalitions, and sampling probabilities. Thus, $\mathbb{E}_t$ averages over the randomness generated during round t.

Assumption 2. *The global loss function $\mathcal{L}(\theta)$ is L-smooth, i.e., it is differentiable with L-Lipschitz gradient. Moreover, the federated gradient* (16) *has uniformly bounded conditional second moment, i.e., there exists $M > 0$ such that $\mathbb{E}_t[\|g(t)\|^2] \le M$ for every t. Moreover, $\mathcal{L}$ is uniformly bounded from below, i.e., $\mathcal{L}^* := \inf_{\theta\in\mathbb{R}^D}\mathcal{L}(\theta) > -\infty$.*

Assumption 3. *There exists $c > 0$ such that $\nabla\mathcal{L}(\theta_{gl}(t))^\top\mathbb{E}_t[g(t)] \ge c\|\nabla\mathcal{L}(\theta_{gl}(t))\|^2$ for every t.*

Theorem 1. *Let $\{\theta_{gl}(t)\}$ be the sequence of global models update produced by `FedCVR-Bolt` algorithm with a single local step, $S = 1$ and constant learning rate $\eta > 0$. Under Assumptions 2–3, the asymptotic relation holds*

$$\limsup_{T\to\infty} \frac{1}{T}\sum\nolimits_{t=0}^{T-1} \mathbb{E}[\|\nabla\mathcal{L}(\theta_{gl}(t))\|^2] \le \frac{LM\eta}{2c}. \tag{17}$$

Proof. The proof is reported in Appendix **??**. □

Assumption 3 is a substantive alignment condition on the expected biased update, and Theorem 1 is conditional on this property. Here, $c > 0$ quantifies the degree of alignment and is not a parameter of `FedCVR-Bolt`. Alignment conditions such as Assumption 3 are standard in the analysis of stochastic-gradient methods with biased updates [69], [71]. It is worth noticing that Theorem 1 is a general asymptotic stationarity result for biased stochastic gradients and does not provide an explicit characterization of the impact of the parameters of `FedCVR-Bolt` on the convergence rate, as they are indirectly in the parameters M, L and c from Assumptions 2–3.

B. Computational Cost

The main overhead lies in performing spectral clustering on the server at each communication round, whose computational complexity can be computed as follows.

Proposition 2. *The computational cost is $\mathcal{O}(K^2D + K^2P + IKP^2)$.*

Proof. The computational cost at each round can be broken down as follows [59]. First, for the similarity matrix construction, pairwise similarities between the K clients are computed using a RBF kernel. This requires evaluating squared Euclidean distances between K vectors in $\mathbb{R}^D$, resulting in a complexity of $\mathcal{O}(K^2D)$. Second, constructing the normalized Laplacian from the similarity matrix incurs a cost of $\mathcal{O}(K^2)$. Third, extracting the P eigenvectors associated with the smallest eigenvalues of the $K \times K$ Laplacian requires $\mathcal{O}(K^2P)$ operations in standard implementations. Fourth, the final step clusters the K clients into P clusters in the P-dimensional eigenspace. For I K-means iterations, this has complexity $O(IKP^2)$. Fifth, D covariance matrices $C^d \in \mathbb{R}^{K\times K}$ occupy $O(DK^2)$ memory, while their updates and the computation of the variance-reduction scores require $O(DK^2)$ operations per round; this term is of the same order as the $O(K^2D)$ similarity-matrix computation and therefore does not alter the overall bound. □

In practice, since $P \ll D$ (e.g., in our numerical setting, $P = 10$, $D = 300$), the complexity is effectively driven by the similarity matrix computation, yielding a leading cost of $\mathcal{O}(K^2D)$ per round. This overhead is modest and entirely server-side, since server resources are not a bottleneck.

TABLE II
COMPARISON WITH FL BASELINES ON HETEROGENEOUS CLASSIFICATION BENCHMARKS.

Method	MNIST	CIFAR-10	CIFAR-100
FedAvg	86.30 ± 1.12	52.67 ± 1.25	22.92 ± 0.95
Power-of-Choice	81.00 ± 2.10	49.85 ± 1.71	23.46 ± 1.05
Active FL	88.89 ± 0.98	55.19 ± 1.34	23.35 ± 1.12
FedProx	89.69 ± 0.91	46.44 ± 1.88	22.13 ± 1.34
IFCA	89.46 ± 1.04	52.94 ± 1.09	22.55 ± 1.28
FeSEM	86.77 ± 1.85	44.66 ± 2.06	11.87 ± 0.77
FedCBS	90.05 ± 3.07	56.53 ± 2.36	21.57 ± 0.46
Oort	89.43 ± 1.49	55.99 ± 1.11	19.91 ± 1.00
FedCVR-Bolt	**90.23 ± 0.82**	**57.06 ± 0.77**	**24.82 ± 0.88**

V. NUMERICAL SIMULATIONS

We present experimental results evaluating FedCVR-Bolt in heterogeneous FL scenarios. We employ established FL benchmark datasets [9], [52]; additional experiments on a synthetic dataset specifically designed for a controlled federated linear regression analysis, made publicly available to support result validation and reproducibility, are reported in Appendix **??**. A detailed description of dataset settings, parameters used, and model architectures is provided in Appendix **??**. In our analysis, we compare FedCVR-Bolt against a diverse set of FL baselines. From client selection methods, we include the uniform random policy of FedAvg [1], the exploitative Power-of-Choice strategy [23] that favors clients with higher local losses, and ActiveFL [24], which balances exploration and exploitation through probabilistic selection. From the broader FL literature, we also consider FedProx [9], a regularization-based approach designed for heterogeneous settings, and two clustering-based personalized methods, IFCA [18] and FeSEM [20]. We additionally implement two recent client-selection policies: FedCBS [28], which promotes class-balanced participation, and Oort [29], which selects clients according to a utility score combining statistical informativeness and system efficiency; both are evaluated under the same experimental protocol as all other baselines. Section V-A presents the comparison on classification tasks on real-world datasets; Section V-B analyzes the sensitivity to the participation rate; finally, in Section V-C some extra simulations are presented to further validate our choice for the proximity matrix. Deployment-oriented analyses under intermittent client availability, covering cohort representativeness, device energy and server overhead, are reported in Appendix **??**.

A. FL Classification Benchmarks

We evaluate FedCVR-Bolt on standard FL classification benchmarks. We consider MNIST [72], CIFAR-10, and CIFAR-100 [73], each partitioned across clients using a Dirichlet distribution with concentration parameter $\alpha = 0.1$ [52], which introduces both label and quantity heterogeneity. Each dataset is trained with $P = 10$ clients sampled per round and $S = 10$ local epochs. We set $T = 100$ communication rounds for CIFAR-10 and CIFAR-100, and $T = 200$ for MNIST. The training horizon is chosen per dataset to correspond to a comparable computational budget: MNIST is trained with a lightweight multilayer perceptron on low-dimensional inputs, so that its per-round cost is much smaller than that of the convolutional networks used for CIFAR-10/100, and 200 MNIST rounds require approximately the same wall-clock time as 100 CIFAR rounds. This choice does not affect the fairness of the comparison, since within each dataset all methods are evaluated under identical settings (T, P, S, model architecture, and data partition). Model architectures and parameters are detailed in Appendix **??**. Table II reports average test accuracy [74], averaged across clients (the larger the better). Our results suggest that FedCVR-Bolt achieves the best or highly competitive performance across all benchmarks. In particular, it delivers notable gains over random client sampling (FedAvg) and over the loss-aware Power-of-Choice, which can fail to generalize well in heterogeneous regimes. Regularization with FedProx and clustering-based methods such as IFCA and FeSEM yield competitive results, but remain below FedCVR-Bolt. This confirms the effectiveness of our selection strategy in both simple and highly heterogeneous datasets. Results in Table II are reported as mean ± twice the standard deviation across independent runs; the bands thus describe the run-to-run dispersion of each method. We observe that the margins naturally shrink on MNIST, an almost saturated benchmark on which all heterogeneity-aware baselines perform similarly, and the advantage of FedCVR-Bolt over the strongest baseline is small in absolute terms; the gains instead widen on the harder CIFAR-10 and CIFAR-100 tasks, where statistical heterogeneity has a stronger impact. Furthermore, beyond the average accuracy, FedCVR-Bolt consistently exhibits the smallest dispersion on MNIST and CIFAR-10. Concerning the newly added selection baselines, Oort remains below FedCVR-Bolt on all benchmarks and exhibits markedly larger run-to-run dispersion, reflecting its greedy loss-driven selection. FedCBS attains performance comparable to FedCVR-Bolt on the 10-class benchmarks, but falls clearly behind on CIFAR-100: its class-rebalancing objective structurally degrades whenever the number of classes exceeds the participation budget ($P < B$), a regime excluded by its own design guidelines [28] yet common in realistic deployments. Moreover, FedCBS requires the server to know the clients' label distributions – or trusted-hardware/homomorphic-encryption infrastructure to compute their pairwise products privately – and is defined only for classification tasks. FedCVR-Bolt requires neither: it operates solely on the model updates that the server already receives in standard FL, applies to arbitrary learning tasks (cf. the regression experiments in Appendix **??**), and adapts its selection to the evolving training dynamics rather than to a static dataset property, making it more readily deployable in privacy-sensitive IoT scenarios.

B. Sensitivity to the Participation Rate

Recall that in FedCVR-Bolt the number of participating clients P coincides with the number of coalitions, so that P/K is the participation rate and K/P the average coalition size. In the main experiments we fix $P = 10$, i.e., a 10% participation rate, which is the common setup in FL benchmarks [52]. Figure

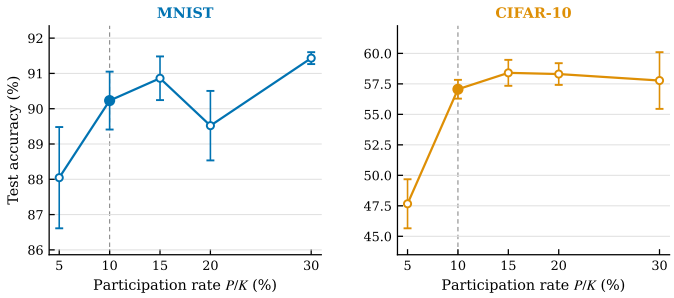


Fig. 2. Final test accuracy of `FedCVR-Bolt` on MNIST (left) and CIFAR-10 (right) as a function of the participation rate P/K, varied from 5% to 30% ($K = 100$). Points report mean ± 2 standard deviations across independent runs; the filled marker denotes the $P = 10$ operating point used in the main experiments (Table II). Accuracy saturates beyond the 10% rate, which attains near-peak performance at the lowest communication cost.

TABLE III
TEST ACCURACY ACROSS PROXY METRICS. BEST PER DATASET IN BOLD.

	RBF (ours)	Cosine Sim.	Laplacian Ker.	Sigmoidal Ker.
MNIST	**90.23 ± 0.82**	88.72 ± 0.97	88.28 ± 0.65	88.92 ± 0.85
CIFAR-10	**57.06 ± 0.77**	55.32 ± 0.78	49.85 ± 0.71	50.25 ± 1.12

2 reports the final test accuracy of `FedCVR-Bolt` on MNIST and CIFAR-10 as the participation rate varies from 5% to 30%. Accuracy degrades markedly below the 10% rate, where too few coalitions are available to cover the heterogeneous client population, and saturates above it: raising the rate to 30% yields no appreciable gain while tripling the per-round communication cost. Larger rates also shrink the average coalition size, making the within-coalition selection progressively closer to uniform sampling and exposing individual clients as quasi-deterministic representatives. The 10% operating point thus achieves near-peak accuracy at the lowest communication budget, consistent with the limited-participation regime that `FedCVR-Bolt` is designed for.

C. Proximity matrix validation

We validate our choice for the proximity matrix through an ablation study to analyze our framework's robustness. We implement our algorithm with other similarity functions used in coalition formation [75]:

- **Cosine Similarity**: $A_{kj} = \theta_k(t)^\top \theta_j(t) / (||\theta_k(t)||_2 ||\theta_j(t)||_2)$
- **Laplacian Kernel**: $A_{kj} = \exp(-\gamma ||\theta_k(t) - \theta_j(t)||_1)$
- **Sigmoid Kernel**: $A_{kj} = \tanh(\gamma \theta_k(t)^\top \theta_j(t) + c)$

Unlike the RBF and Laplacian kernels, the cosine and sigmoid kernels may take negative values, in which case the guarantees of normalized spectral clustering, which require nonnegative affinities [59], no longer hold, and for these two kernels the spectral step should thus be regarded as heuristic. While our theoretically-motivated choice yields the best performance (90.23% on MNIST, 57.06% on CIFAR-10), our framework continues to substantially outperform the baselines even with the alternative similarity metrics as shown in Table III.

VI. CONCLUSION

We introduced `FedCVR-Bolt`, a novel framework addressing heterogeneous FL by drawing foundational insights from opinion dynamics models. The core of `FedCVR-Bolt` lies in a client selection strategy that biases sampling towards clients with larger variance-reduction scores. Distinct from conventional sampling methods, `FedCVR-Bolt` exploits concepts from coalition formation within opinion dynamics to identify clusters of clients exhibiting similar model characteristics, i.e., the "opinions." A representative client is then sampled from each identified cluster. This approach is supported by theoretical results that provide a foundation for our algorithm. Empirical evaluations on heterogeneous datasets, including a synthetic linear regression task specifically designed to highlight heterogeneity challenges and more complex benchmarks, demonstrate that our algorithm outperforms or matches state-of-the-art FL client selection algorithms.

The analysis done opens several directions for further methodological and theoretical developments. First, a further extension would be to jointly optimize the number of clusters and clients, rather than fixing it a priori, by incorporating more advanced clustering strategies and explicitly accounting for the system's communication and computational constraints, as well as the composition with device-level constraints, such as hardware heterogeneity, by restricting each coalition to its resource-feasible members before sampling. Moreover, while our deployment-oriented analysis shows that the coalition structure keeps the participating cohorts representative under moderate client dropout, characterizing regimes where the reachable clients become scarce relative to the participation budget, in which coalitions may remain unrepresented, is an important direction for IoT deployments. Although our method effectively leverages pairwise individual model correlation, extending this framework to hypergraphs would model more complex and hierarchical dependencies between coalitions. Furthermore, our current asymptotic analysis is conditional on bounded-moment and gradient-alignment assumptions; future work could derive these properties directly from the transient evolution of the client models, covariance estimates, and coalition formation. Finally, there is significant potential to bridge the gap between our heuristic Boltzmann exploration and a rigorous statistical mechanics analysis of system stability, including an analytical characterization of the gradient bias and variance induced by the proposed selection policy.

REFERENCES

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Stat.*, pp. 1273–1282, 2017.

[2] C. Yu, S. Shen, S. Wang, K. Zhang, and H. Zhao, "Communication-efficient hybrid federated learning for e-health with horizontal and vertical data partitioning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 3, pp. 5614–5628, 2025.

[3] Z. Yang, W. Xia, Z. Lu, Y. Chen, X. Li, and Y. Zhang, "Hypernetwork-based physics-driven personalized federated learning for ct imaging," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 2, pp. 3136–3150, 2025.

[4] H. Fang *et al.*, "Contrastive federated learning for graph anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–12, 2025.

[5] K.-S. Wong, D.-M. Nguyen, K. Le, L. Ho, C. Do, and D. Le-Phuoc, "An empirical study of federated learning on iot–edge devices: Resource allocation and heterogeneity," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–13, 2025.

[6] P. Kairouz *et al.*, "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021.

[7] K. Bonawitz *et al.*, "Towards federated learning at scale: System design," in *Proc. Mach. Learn. Syst.*, vol. 1, pp. 374–388, 2019.

[8] L. Li, H. Xiong, Z. Guo, J. Wang, and C.-Z. Xu, "SmartPC: Hierarchical pace control in real-time federated learning system," in *2019 IEEE Real-Time Syst. Symp.*, pp. 406–418, 2019.

[9] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. 2nd Mach. Learn. Syst.*, vol. 2, pp. 429–450, 2020.

[10] Z. Lu, H. Pan, Y. Dai, X. Si, and Y. Zhang, "Federated learning with non-IID data: A survey," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19188–19209, 2024.

[11] D. Li and J. Wang, "Fedmd: Heterogenous federated learning via model distillation," *arXiv preprint arXiv:1910.03581*, 2019.

[12] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Mach. Learn.*, pp. 5132–5143, 2020.

[13] S. P. Karimireddy *et al.*, "Breaking the centralized barrier for cross-device federated learning," in *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 28663–28676, 2021.

[14] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," in *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 4427 – 4437, 2017.

[15] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, pp. 1126 – 1135, 2017.

[16] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 3557–3568, 2020.

[17] C. T Dinh, N. Tran, and J. Nguyen, "Personalized federated learning with moreau envelopes," in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 21394–21405, 2020.

[18] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 19586–19597, 2020.

[19] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3710–3722, 2020.

[20] G. Long, M. Xie, T. Shen, T. Zhou, X. Wang, and J. Jiang, "Multi-center federated learning: clients clustering for better personalization," *World Wide Web*, vol. 26, no. 1, pp. 481–500, 2023.

[21] J. A. Carrillo, N. G. Trillos, S. Li, and Y. Zhu, "Fedcbo: Reaching group consensus in clustered federated learning through consensus-based optimization," *J. Mach. Learn. Res.*, vol. 25, no. 214, pp. 1–51, 2024.

[22] W. Bao, H. Wang, J. Wu, and J. He, "Optimizing the collaboration structure in cross-silo federated learning," in *Proc. 40th Int. Conf. Mach. Learn.*, pp. 1718–1736, 2023.

[23] Y. J. Cho, J. Wang, and G. Joshi, "Towards understanding biased client selection in federated learning," in *Proc. 25th Int. Conf. Artif. Intell. Stat.*, pp. 10351–10375, 2022.

[24] J. Goetz, K. Malik, D. Bui, S. Moon, H. Liu, and A. Kumar, "Active federated learning," *arXiv preprint arXiv:1909.12641*, 2019.

[25] R. Balakrishnan, T. Li, T. Zhou, N. Himayat, V. Smith, and J. Bilmes, "Diverse client selection for federated learning via submodular maximization," in *Proc. 10th Int. Conf. Learn. Rep.*, 2022.

[26] E. Diao, J. Ding, and V. Tarokh, "Heterofl: Computation and communication efficient federated learning for heterogeneous clients," in *Proc. 9th Int. Conf. Learn. Rep.*, 2020.

[27] Y. Liao, J. Feng, Z. Zhou, B. Shang, L. Liu, and Q. Pei, "Quality-aware client selection and resource optimization for federated learning in computing networks," in *2024 IEEE Int. Conf. Commun.*, pp. 2628–2633, 2024.

[28] J. Zhang *et al.*, "Fed-cbs: A heterogeneity-aware client sampling mechanism for federated learning via class-imbalance reduction," in *Proc. 40th Int. Conf. Mach. Learn.*, pp. 41354–41381, 2023.

[29] F. Lai, X. Zhu, H. V. Madhyastha, and M. Chowdhury, "Oort: Efficient federated learning via guided participant selection," in *15th USENIX Symp. Oper. Syst. Des. Implement.*, pp. 19–35, 2021.

[30] C. Tian, L. Li, Z. Shi, J. Wang, and C. Xu, "Harmony: Heterogeneity-aware hierarchical management for federated learning system," in *55th IEEE/ACM Int. Symp. Microarchitecture*, pp. 631–645, 2022.

[31] W. B. Powell, "From reinforcement learning to optimal control: A unified framework for sequential decisions," in *Handbook of Reinforcement Learning and Control* (K. G. Vamvoudakis, Y. Wan, F. L. Lewis, and D. Cansever, eds.), pp. 29–74, Springer, 2021.

[32] C. Tian, Z. Shi, L. Li, C.-z. Xu, *et al.*, "Ranking-based client imitation selection for efficient federated learning," in *Proc. 41st Int. Conf. Mach. Learn.*, 2024.

[33] Y. G. Kim and C.-J. Wu, "Autofl: Enabling heterogeneity-aware energy efficient federated learning," in *54th IEEE/ACM Int. Symp. Microarchitecture*, pp. 183–198, 2021.

[34] H. Wang, Z. Kaplan, D. Niu, and B. Li, "Optimizing federated learning on non-iid data with reinforcement learning," in *2020 IEEE Conf. Comput. Commun.*, pp. 1698–1707, 2020.

[35] S. Q. Zhang, J. Lin, and Q. Zhang, "A multi-agent reinforcement learning approach for efficient client selection in federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, pp. 9091–9099, 2022.

[36] M. Guazzone, C. Anglano, and M. Sereno, "A game-theoretic approach to coalition formation in green cloud federations," in *14th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput.*, pp. 618–625, 2014.

[37] K. Donahue and J. Kleinberg, "Model-sharing games: Analyzing federated learning under voluntary participation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, pp. 5303–5311, 2021.

[38] K. Donahue and J. Kleinberg, "Optimality and stability in federated learning: A game-theoretic approach," in *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 1287–1298, 2021.

[39] A. Blum, N. Haghtalab, R. L. Phillips, and H. Shao, "One for one, or all for all: Equilibria and optimality of collaboration in federated learning," in *Proc. 38th Int. Conf. Mach. Learn.*, pp. 1005–1014, PMLR, 2021.

[40] M. Ota, Y. Sakurai, and S. Oyama, "Coalitional game theoretic federated learning," in *2022 IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol.*, pp. 48–55, 2022.

[41] L. Nagalapatti and R. Narayanam, "Game of gradients: Mitigating irrelevant clients in federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, pp. 9046–9054, 2021.

[42] A. Das and D. Kempe, "Algorithms for subset selection in linear regression," in *Proc. 14th Annu. ACM Symp. Theory Comput.*., STOC '08, pp. 45–54, 2008.

[43] A. Das and D. Kempe, "Approximate submodularity and its applications: Subset selection, sparse approximation and dictionary selection," *J. Mach. Learn. Res.*, vol. 19, no. 3, pp. 1–34, 2018.

[44] Y. Han and L. Yu, "A Variance Reduction framework for stable feature selection," in *2010 IEEE Int. Conf. Data Min.*, pp. 206–215, 2010.

[45] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian Processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, jun 2008.

[46] P. Batlle, J. Bruna, C. Fernandez-Granda, and V. M. Preciado, "Adaptive test allocation for outbreak detection and tracking in social contact networks," *SIAM J. Control Optim.*, vol. 60, no. 2, pp. S274–S293, 2022.

[47] R. Raineri, G. Como, and F. Fagnani, "Optimal selection of the most informative nodes for a noisy degroot model with stubborn agents," in *2025 Eur. Control Conf.*, pp. 558–563, 2025.

[48] R. Raineri, G. Como, and F. Fagnani, "Optimal selection of the most informative nodes in opinion dynamics on networks," in *Proc. 22nd IFAC World Congr.*, vol. 56, pp. 4192–4197, 2023.

[49] G. Gallavotti, *Statistical mechanics: A short treatise*. Berlin, Germany: Springer, 1 ed., 2013.

[50] M. Mäs, A. Flache, and D. Helbing, "Individualization as driving force of clustering phenomena in humans," *PLOS Comput. Biol.*, vol. 6, pp. 1–8, 10 2010.

[51] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice Hall, 1 ed., 1993.

[52] S. Caldas *et al.*, "Leaf: A benchmark for federated settings," *arXiv preprint arXiv:1812.01097*, 2018.

[53] S. Vahidian, M. Morafah, W. Wang, V. Kungurtsev, C. Chen, M. Shah, and B. Lin, "Efficient distribution similarity identification in clustered federated learning via principal angles between client data subspaces," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, pp. 10043–10052, 2023.

[54] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the nyström method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 214–225, 2004.

[55] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 7611–7623, 2020.

[56] B. Luo, W. Xiao, S. Wang, J. Huang, and L. Tassiulas, "Adaptive heterogeneous client sampling for federated learning over wireless networks," *IEEE Trans. Mob. Comput.*, vol. 23, no. 10, pp. 9663–9677, 2024.

[57] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annu. Rev. Control*, vol. 43, pp. 65–79, 2017.
[58] U.-U. Narantsatsralt and S. Kang, "Social network community detection using agglomerative spectral clustering," *Complexity*, vol. 2017, pp. 1–10, 2017.
[59] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, p. 395–416, 2007.
[60] W. B. Powell, *Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions*. Wiley, 1 ed., 2022.
[61] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Fontiers in Econometrics* (P. Zarembka, ed.), pp. 105–142, New York: Academic press, 1974.
[62] L. Onsager, "Statistical hydrodynamics," *Il Nuovo Cimento (1943-1954)*, vol. 6, no. Suppl 2, pp. 279–287, 1949.
[63] S. Calabrese and A. Porporato, "Origin of negative temperatures in systems interacting with external fields," *Physics Letters A*, vol. 383, no. 18, pp. 2153–2158, 2019.
[64] A. Porporato, S. Calabrese, and L. Rondoni, "Thermodynamic langevin equations," *Physical Review E*, vol. 110, no. 5, p. 054136, 2024.
[65] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.
[66] A. Licciardi, D. Leo, E. Fanì, B. Caputo, and M. Ciccone, "Interaction-aware Gaussian weighting for clustered federated learning," in *Proc. 42nd Int. Conf. Mach. Learn.*, vol. 267, pp. 37642–37666, 2025.
[67] H. J. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*. Springer, 2003.
[68] K. Bonawitz *et al.*, "Practical secure aggregation for federated learning on user-held data," *arXiv preprint arXiv:1611.04482*, 2016.
[69] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, 2018.
[70] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on non-IID data," in *Proc. 8th Int. Conf. Learn. Rep.*, 2020.
[71] A. Ajalloeian and S. U. Stich, "On the convergence of SGD with biased gradients," *arXiv preprint arXiv:2008.00051*, 2020.
[72] Y. LeCun, "The MNIST database of handwritten digits," 1998. Available at yann.lecun.com/exdb/mnist.
[73] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009. Master's thesis, University of Toronto.
[74] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Min. Knowl. Manag. Process.*, vol. 5, no. 2, p. 1, 2015.
[75] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
[76] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 1 ed., 2014.
[77] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford University Press, 3 ed., 2001.
[78] E. J. Gumbel, *Statistics of Extremes*. New York: Columbia University Press, 1958.
[79] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained iot devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1–24, 2021.
[80] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *ICC 2019-2019 IEEE international conference on communications (ICC)*, pp. 1–7, IEEE, 2019.



Alessandro Licciardi received the B.Sc. degree in Mathematics for Engineering and the M.Sc. degree in Mathematical Engineering, both cum laude, from Politecnico di Torino, Turin, Italy, in 2021 and 2023, respectively. He is currently pursuing the Ph.D. degree in Mathematical Sciences with Politecnico di Torino, under an industrial scholarship co-funded by Eltek S.p.A.. In 2025–2026, he was a Visiting Student Researcher with Stanford University, Stanford, CA, USA. His research interests include the mathematical foundations of machine learning, federated learning, wavelet scattering representations, and statistical-mechanics approaches to generative modeling and rare-event sampling.



Roberta Raineri received the B.S. and the M.Sc. (cum laude) in Mathematical Engineering, and the Ph.D. in Pure and Applied Mathematics from Politecnico di Torino, Italy, in 2019, 2020, and 2025, respectively. After a Research Fellowship with the Department of Electronics and Telecommunications, Politecnico di Torino, she is now an R&D Control System Engineer at Modelway, Turin, since 2025. Her research interests include analysis and control of network systems, with main applications in opinion dynamics and social networks.



Anton V. Proskurnikov (SM'18) received the M.Sc. and Ph.D. degrees in applied mathematics from St. Petersburg State University, St. Petersburg, Russia, in 2003 and 2005, respectively. He was an Assistant Professor with St. Petersburg State University (2003–2010), a Senior Researcher with the Institute for Problems in Mechanical Engineering, Russian Academy of Sciences, St. Petersburg (2006–2022), and a Postdoctoral Researcher with the University of Groningen, The Netherlands (2014–2016), and Delft University of Technology, The Netherlands (2016–2018). Since [2018], he has been with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy, where he is currently an Associate Professor. Dr. Proskurnikov received the IFAC/Elsevier Paper Prize for the best paper published in Annual Reviews in Control in 2020 and the O. Hugo Schuck Best Paper Award (Theory) from the American Automatic Control Council in 2023. He served as an Associate Editor of IEEE Transactions on Automatic Control from 2019 to 2025. His research interests include control theory, complex systems, nonlinear dynamics, multiagent systems, and models in social and natural sciences.



Lamberto Rondoni received the Laurea degree in nuclear engineering from the University of Bologna, Italy, and the M.Sc. degrees in mathematics and physics, and the Ph.D. degree in mathematics from Virginia Polytechnic Institute, USA. After holding positions at Virginia Tech, Rutgers University, and the University of New South Wales, he joined Politecnico di Torino in 1995, where he is currently a Full Professor of Mathematical Physics. He holds secondary appointments with the Italian Institute of Nuclear Physics and the University of Torino. His research concerns statistical mechanics, its foundations and its applications to biomedical problems, nanotechnology, and artificial intelligence.



Lorenzo Zino is an Associate Professor with the Department of Electronics and Telecommunications, Politecnico di Torino, Italy, since 2026. He received B.S. (2012), M.Sc. (summa cum laude, 2014), and Ph.D. (with honors, 2018) in Applied Mathematics from Politecnico di Torino. He held Research Fellowships at Politecnico di Torino, University of Groningen, and New York University. In 2024, he was the recipient of the Best Young Author Journal Paper Award from the IEEE CSS Italy Chapter. He is a member of the Editorial Board of *Scientific Reports*, the *International Journal of Control*, the *IEEE Control Systems Letters*, and *Automatica*. His research interests include the modeling, analysis, and control of dynamical processes over complex networks.