Learning in Orbit: A Physics-Aware Graph-Decentralized Federated Learning for Multi-Satellite Space Situational Awareness

Gagandeep Kaur[*1], **Ranjitha Prasad**[2]

[1]IntelliCom Lab, IIIT Delhi, India
gagandeepk@iiitd.ac.in
[2]IntelliCom Lab, IIIT Delhi, India
ranjitha@iiitd.ac.in

Abstract

The orbital environment is increasingly congested, heightening collision risk and demanding robust Space Situational Awareness (SSA). Ground-based tracking and centralized learning face latency, fragmented datasets, and strict privacy limits on telemetry sharing. While ML aids orbital prediction, purely data-driven models fail under sparse or irregular observations; Physics-Informed Neural Networks (PINNs) embed dynamics to ensure physical consistency. However, locally trained PINNs lack shared context, fragmenting awareness. Collaboration is often constrained by privacy policy, export controls, or mission secrecy—sometimes forcing purely local learning and leaving blind spots. This motivates Federated Learning (FL), where satellites share model updates (not raw data) to refine physics-consistent predictors while preserving data autonomy. However, single-server FL is ill-suited for orbital networks, as it creates a single point of failure, over-smooths data, and exposes the system to vulnerability from link outages. We therefore propose Graph-Decentralized Federated Learning (Graph-DFL) for multi-satellite SSA: a serverless framework where satellites exchange quantized incremental updates only with neighbors and reach consensus via topology-aware diffusion. Each Low Earth Orbit (LEO) client trains a GRU-based deinterleaver and a local PINN, while Medium Earth Orbit (MEO) relays apply confidence-weighted Cayley–Menger × Light-Cone (CM × LC) fusion. Experiments on real Two-Line Element (TLE)–derived SGP4 trajectories show that Graph-DFL achieves high deinterleaver accuracy and low trajectory RMSE, indicating a resilient, physics-consistent, privacy-preserving solution for SSA without centralization.

Introduction

The orbital environment is becoming densely populated as thousands of active satellites, retired payloads, and debris share narrow shells with expanding mega-constellations (European Space Agency 2025; McDowell 2020). Each object evolves under gravity, drag, and solar perturbations, tightening collision margins and making reliable *Space Situational Awareness (SSA)*, i.e., the ability to observe, identify, and predict object trajectories vital to prevent cascading conjunctions (Kessler and Cour-Palais 1978; European Space Agency 2025).

Traditional ground-based tracking and centralized learning pipelines face latency, fragmented data, and strict privacy constraints. While machine learning (ML) aids orbit prediction, purely data-driven models degrade under sparse observations (Caldas and Soares 2024). *Physics-Informed Neural Networks (PINNs)* integrate known dynamics into training, improving generalization and consistency with fewer samples (Raissi, Perdikaris, and Karniadakis 2019; Karniadakis et al. 2021).

Physics-Informed Neural Networks (PINNs)

PINNs embed governing equations into neural training:

$$\mathcal{J}\big(\mathbf{u}(t)\big) = \dot{\mathbf{u}}(t) - j\big(\mathbf{u}(t)\big) = 0, \tag{1}$$

where $\mathbf{u}(t) \in \mathbb{R}^d$ is the system state (e.g., orbital position–velocity) and $j(\cdot)$ is the physics-based vector field. A neural surrogate $u_{\mathbf{w}}(t)$, parameterized by weights $\mathbf{w}$, minimizes

$$\mathcal{L}_{\text{PINN}} = \lambda_{\text{data}}\big\|\mathbf{u}_{\mathbf{w}}(t) - \mathbf{u}_{\text{obs}}(t)\big\|_2^2 + \lambda_{\text{phys}}\big\|\mathcal{J}\big(\mathbf{u}_{\mathbf{w}}(t)\big)\big\|_2^2, \tag{2}$$

balancing data fidelity and physical consistency.

In SSA, each LEO node trains a local PINN on partially deinterleaved telemetry, producing dynamics-consistent trajectories. However, isolated PINNs lack shared situational context, as privacy and policy restrictions prevent data exchange, leaving blind spots that elevate collision risk. *Federated Learning (FL)* addresses this by enabling collaborative model updates without sharing raw telemetry (McMahan et al. 2017; Kairouz et al. 2021).

Federated Learning

Federated Learning (FL) enables distributed model training across edge devices (*clients*) coordinated by a central *server*. Each client c trains on private data $\mathcal{S}_c$ of size D_c, and the server aggregates updates to form a global model (Kairouz et al. 2021). The global objective is

$$\min_{\mathbf{w}(t)\in\mathbb{R}^N} F(\mathbf{w}(t)) \triangleq \frac{1}{C}\sum_{c=1}^{C} f_c(\mathbf{w}(t)), \tag{3}$$

[*]Accepted at the 1st Workshop on Federated Learning for Critical Applications (FLCA), in conjunction with AAAI 2026.

where $f_c(\cdot)$ is the local loss and $\mathbf{w}(t)$ the global parameters at round $t \in [T]$. A common aggregation rule is Federated Averaging (FedAvg) (McMahan et al. 2017):

$$\mathbf{w}(t) = \frac{1}{C}\sum_{c=1}^{C} \mathbf{w}_c(t), \tag{4}$$

where $\mathbf{w}_c(t)$ are client-local parameters.

Modern NNs, such as GPT-3 with 175B parameters (Brown et al. 2020), are highly over-parameterized, i.e., $N \gg |\mathcal{S}_c|$. In FL, each client's limited data motivates the use of over-parameterized networks (ONNs), whose slow weight evolution under lazy training (Chizat and Bach 2018) results in transmitted updates that are small and compressible. This property supports efficient communication via sparsification or quantization (Alistarh et al. 2017; Reisizadeh et al. 2020).

Communication efficiency via quantization To preserve communication efficiency in over-parameterized FL, we adopt *quantization*, which is a precision-reduction operator that preserves vector dimensionality while mapping continuous updates to a finite codebook. We will be quantizing the client (LEO satellite) model incremental updates for the c-th client in the FL communication round t given by

$$\Delta\mathbf{w}_c(t) = \mathbf{w}_c(t) - \mathbf{w}_c(t-1). \tag{5}$$

Further, let us define a b-bit quantizer, $\mathcal{Q}_b : \mathbb{R}^d \to \mathbb{R}^d$ applied to the input $\Delta\mathbf{w}_c(t)$ as follows:

$$\mathcal{Q}_b(\Delta\mathbf{w}_c(t)) \;=\; \Delta\widehat{\mathbf{w}}_c(t) + \boldsymbol{\varepsilon}_c(t), \tag{6}$$

$$\mathbb{E}\big[\|\boldsymbol{\varepsilon}_c(t)\|_2^2\big] \le \delta^q \|\Delta\mathbf{w}_c(t)\|_2^2, \tag{7}$$

where $\Delta\widehat{\mathbf{w}}_c(t)$ is the quantized incremental update (gradient), $\boldsymbol{\varepsilon}_c(t)$ is quantization noise, and each entry of the vector $\delta^q \in [0, 1)$. δ^q bounds mean-square distortion relative to signal energy.

Uniform Quantizer A uniform b-bit quantizer divides the dynamic range $[-X_{\max}, X_{\max}]$ into 2^b equal intervals. The quantization step size is

$$\Delta = \frac{2X_{\max}}{2^b}. \tag{8}$$

For an input $x \in [-X_{\max}, X_{\max}]$, the quantization error, ε^q is bounded by

$$|\varepsilon_q| \le \frac{\Delta}{2}. \tag{9}$$

Under the standard high-resolution assumption, where ε_q is uniformly distributed over $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$, the variance of the quantization error satisfies (Gersho and Gray 1992; Gray and Neuhoff 1998)

$$\mathrm{Var}(\varepsilon_q) \le \frac{\Delta^2}{12}. \tag{10}$$

Thus, quantization noise is both bounded and has mean-square magnitude scaling as $\mathcal{O}(\Delta^2)$, or equivalently $\mathcal{O}(2^{-2b})$ with respect to bit-width b.

Graph-Decentralized Federated Learning (Graph-DFL) Setup

Limitations of Centralized FL in SSA

Conventional FL (e.g., FedAvg, FedProx) relies on a *central server* to maintain a single global model $\mathbf{w}(t)$ (McMahan et al. 2017; Li et al. 2020; Kairouz et al. 2021). In multi-satellite SSA, heterogeneous orbital regimes and sensing distributions result in a single $\mathbf{w}(t)$ *over-smoothing* client diversity, thereby degrading local fidelity (Kairouz et al. 2021). Intermittent inter-satellite links and latency induce bottlenecks, making the server a *single point of failure*, and central aggregation risks the leakage of model statistics, which conflicts with inter-agency privacy policies (Zhu, Liu, and Han 2019; Melis et al. 2019).

From Central to Graph-Decentralized FL. We adopt *Graph-DFL*, where nodes exchange *quantized incremental* updates only with neighbors and reach agreement through weighted consensus (Olfati-Saber, Fax, and Murray 2007; Lian et al. 2017; Koloskova, Stich, and Jaggi 2019; Alistarh et al. 2017). This topology and physics-aware process mirrors the constellation's communication graph, improving robustness, scalability, and privacy without a central coordinator.

Graph-DFL Formulation. Let the inter-satellite network be a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node $c \in \mathcal{V}$ maintains a local model $\mathbf{w}_c(t) \in \mathbb{R}^d$ and transmits gradient increments $\Delta\mathbf{w}_c(t)$. Local learning and decentralized consensus evolve as

$$\begin{aligned}\Delta\mathbf{w}_c(t+1) = \Delta\mathbf{w}_c(t) - \eta\nabla F_c\big(\Delta\mathbf{w}_c(t)\big) \\ + \alpha\sum_{j\in\mathcal{N}_c} \mathbf{P}_{cj}\big(\Delta\mathbf{w}_j(t) - \Delta\mathbf{w}_c(t)\big),\end{aligned} \tag{11}$$

where $\mathcal{N}_c$ is the neighbor set of node c, $\mathbf{P}$ is a row-stochastic mixing matrix ($\sum_j \mathbf{P}_{cj} = 1$), η is the learning rate, and α the consensus strength (Nedić and Ozdaglar 2009; Xiao and Boyd 2004). This diffusion mechanism yields approximate agreement while preserving local data privacy.

Cayley Menger and Light Cone Fusion Weighted DFL

Limitations of Basic DFL and need for CM×LC Fusion In multi-satellite SSA, (11) neglects orbital physics and asynchronous communication. The key issues are:

(a) Spatial mismatch. Neighbors in $\mathcal{G}$ may be misaligned; averaging across distinct orbital planes mixes incompatible dynamics and disrupts phase alignment.

(b) Temporal misalignment. LEO–MEO links are intermittent; updates arrive at uneven times t_i with delays Δt_{ij}, desynchronizing learning (Olfati-Saber and Murray 2004).

(c) Staleness and drift. Outages yield stale updates, degrading consensus (Nedić and Ozdaglar 2009):

$$\Delta\mathbf{w}_c(t+1) = \Delta\mathbf{w}_c(t) + \alpha\sum_{j\in\mathcal{N}_c} \mathbf{P}_{cj}\big(\Delta\mathbf{w}_j(t-\tau_{cj}) - \Delta\mathbf{w}_c(t)\big), \tag{12}$$

where $\tau_{cj} \geq 0$ denotes delay between nodes c and j. Large or uneven τ_{cj} inflates the spectral radius, slowing convergence (Xiao and Boyd 2004; Olfati-Saber, Fax, and Murray 2007).

To address these effects, fusion must enforce (i) geometric consistency, (ii) temporal causality, and (iii) staleness attenuation via physics-aware weighting.

Cayley–Menger (CM): Spatial Coherence For LEO positions $r_c \in \mathbb{R}^3$, define

$$D_{cj} = \|r_c - r_j\|_2^2. \tag{13}$$

The Cayley–Menger determinant,

$$\mathrm{CM}(r_1, \ldots, r_n) = \frac{(-1)^{n+1}}{2^n (n!)^2} \begin{vmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & D_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & D_{n1} & \cdots & 0 \end{vmatrix}, \tag{14}$$

is proportional to the squared simplex volume (Menger 1930; Blumenthal 1953). Large CM implies spatial coherence; near-zero implies degeneracy. The geometric trust for client c in cluster C_l is

$$\nu_{c,l}^{(\mathrm{CM})} \propto \mathrm{CM}\left(\{r_j\}_{j \in C_l}\right). \tag{15}$$

Light-Cone (LC): Temporal Co-Visibility Clients c and j are co-visible if

$$|t_c - t_j| \leq \frac{\|r_c - r_j\|_2}{c_0}, \tag{16}$$

with propagation speed c_0 (Rindler 2006). A smooth LC kernel is

$$\kappa_{\mathrm{LC}}(c, j) = \exp\left(-\frac{(|t_c - t_j| - \|r_c - r_j\|_2 / c_0)^2}{2\sigma_t^2}\right), \tag{17}$$

$$\nu_{c,l}^{(\mathrm{LC})} \propto \sum_{j \in C_l} \kappa_{\mathrm{LC}}(c, j), \tag{18}$$

which rewards causal synchrony and down-weights stale updates.

Combined CM×LC Fusion At relay l, the final trust for client c is

$$\nu_{c,l} = \frac{\exp\left(\beta_1 w_{c,l}^{(\mathrm{CM})} + \beta_2 w_{c,l}^{(\mathrm{LC})}\right)}{\sum_{j \in C_l} \exp\left(\beta_1 w_{j,l}^{(\mathrm{CM})} + \beta_2 w_{j,l}^{(\mathrm{LC})}\right)}, \tag{19}$$

and the relay update is

$$\Delta \mathbf{w}_{M,l}(t+1) = (1 - \alpha_{\mathrm{intra}}) \Delta \mathbf{w}_{M,l}(t) + \alpha_{\mathrm{intra}} \sum_{c \in C_l} \nu_{c,l} \Delta \mathbf{w}_c(t+1). \tag{20}$$

CM×LC fusion restores convergence under asynchronous, quantized communication—achieving physics-consistent, staleness-resilient decentralized learning across the constellation (Koloskova, Stich, and Jaggi 2019; Alistarh et al. 2017; Vaswani et al. 2017).

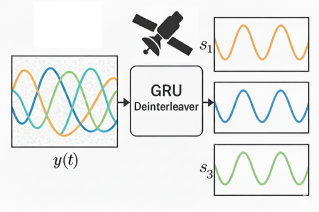

Figure 1: GRU-based deinterleaving: the received mixture $y(t)$ (left) combines multiple sources; the GRU deinterleaver (center) reconstructs individual signals s_1, s_2, s_3 (right) (Chung et al. 2014).

Need for Deinterleaving

Why Interleaving of Satellite Models is Inevitable

In multi-satellite SSA, each satellite perceives only a partial, time-shifted view of global orbital dynamics. These fragments, modulated by shared gravity, drag, and illumination, naturally entangle across satellites, leading to *interleaved latent representations* (Rindler 2006).

ELINT analogy. Following Electronic Intelligence (ELINT) convention, let I emitters produce pulse trains

$$s_i(t) = \sum_{n=1}^{N_i} \delta\big(t - \tau_{i,n}\big), \qquad i = 1, \ldots, I,$$

where $\{\tau_{i,n}\}_{n=1}^{N_i}$ are the emission instants for emitter i. A receiver collects their superposition

$$y(t) = \sum_{i=1}^{I} s_i(t) = \sum_{k=1}^{K} \delta\big(t - t_k\big),$$

with observed pulse times $\{t_k\}_{k=1}^{K}$ interleaved across sources. Deinterleaving seeks a partition

$$\bigcup_{i=1}^{I} \{\tau_{i,n}\} = \{t_k\}_{k=1}^{K}, \qquad \tau_{i,1} < \tau_{i,2} < \cdots < \tau_{i,N_i} \ \ \forall i,$$

so that each subsequence corresponds to a single latent emitter (RSO).

Fig. 1 gives a visual intuition into the interleaving and deinterleaving processes. Interleaving thus arises from fundamental orbital geometry and light-cone delays, not from system design.

Contributions and Significance

This work introduces the first physics-aware decentralized learning framework for Space Situational Awareness (SSA), addressing the long-standing trade-off between physical fidelity, privacy, and communication efficiency. Its key contributions and significance are:

- **Graph-DFL for Orbital Learning.** A fully serverless learning framework where LEO clients and MEO relays jointly train GRU–PINN models via quantized weight deltas. This removes the single point of failure in FedAvg-style systems and enables resilient, privacy-preserving collaboration across orbital regimes.

- **CM×LC-Weighted Consensus.** A novel Cayley–Menger × Light-Cone fusion rule ensures spatial–temporal coherence in decentralized learning, aligning updates to orbital geometry and causal visibility, which is vital for real multi-orbit coordination.
- **Quantized NTK Convergence.** We extend NTK theory to 4-bit quantized CM×LC-weighted updates, providing the first convergence bound for decentralized physics-informed networks under communication constraints.
- **Comprehensive Real-Orbit Experiments.** Using real TLE/SGP4 trajectories of nine LEO satellites, we demonstrate a significant reduction in RMSE, high deinterleaver accuracy, and good inter-relay consensus, establishing practical on-orbit scalability and robustness.

Problem Statement

Space Situational Awareness (SSA) is the collective process of tracking, predicting, and understanding the motion of Resident Space Objects (RSOs) like LEO satellites in the Earth's orbit. Now let us discuss the problem that we are addressing in this work.

Overview: From Physics Solvers to PINNs and Graph-DFL

Orbital motion follows nonlinear dynamics with perturbations. Under sparse, noisy, or asynchronous sensing conditions, physics-only solvers can be brittle, while purely data-driven models may violate invariants and exhibit poor extrapolation. *PINNs* bridge this gap by learning from data while penalizing violations of known dynamics (Raissi, Perdikaris, and Karniadakis 2019).

In SSA, telemetry is siloed and cannot be freely shared. *Federated Learning (FL)* enables collaborative training without raw data exchange, but central schemes (e.g., FedAvg/FedProx) face privacy, scalability, and heterogeneity limits (Kairouz et al. 2021).

We therefore adopt *Graph-Decentralized FL (Graph-DFL)*, where satellites exchange quantized updates only with neighbors and reach agreement via iterative, topology-aligned consensus (Lian et al. 2017; Koloskova, Stich, and Jaggi 2019). Each LEO acts as an autonomous learner; cooperation occurs through physics-aware, communication-efficient consensus rather than centralized aggregation. We next recall the orbital dynamics governing the RSOs to be learned.

Physical Problem Definition

The physical system underlying our framework is the orbital motion of resident space objects (RSOs) around the Earth, governed by the laws of celestial mechanics. Each LEO node observes only partial, noisy measurements of one or more RSOs and must learn a local surrogate model consistent with both its data and the governing differential equations of motion.

Perturbed two–body dynamics. The acceleration of an RSO in the Earth–centered inertial (ECI) frame satisfies a second–order nonlinear ordinary differential equation (ODE):

$$\ddot{\mathbf{r}}(t) = -\mu \frac{\mathbf{r}(t)}{\|\mathbf{r}(t)\|^3} + a_{J_2}\big(\mathbf{r}(t)\big) + a_D\big(\mathbf{r}(t), \mathbf{v}(t)\big) + a_{\text{other}}(t), \tag{21}$$

where

- $\mathbf{r}(t) \in \mathbb{R}^3$ is the position vector of the object,
- $\mu = G_E$ is Earth's gravitational parameter,
- a_{J_2} represents the acceleration due to Earth's oblateness (J_2 perturbation),
- a_D denotes atmospheric drag, and
- a_{other} captures small unmodeled perturbations such as solar radiation pressure or thruster impulses.

First–order form. For numerical integration and learning, the second–order ODE is expressed as a first–order system:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{r}(t)\\ \mathbf{v}(t)\end{bmatrix} = \begin{bmatrix}\mathbf{v}(t)\\ -\mu \dfrac{\mathbf{r}(t)}{\|\mathbf{r}(t)\|^3} + a_{J_2}(\mathbf{r}) + a_D(\mathbf{r}, \mathbf{v}) + a_{\text{other}}(t)\end{bmatrix} = f_{\text{2body+pert}}\big(x(t)\big), \tag{22}$$

where $\mathbf{x}(t) = [\mathbf{r}(t)^\top, \mathbf{v}(t)^\top]^\top$ is the six–dimensional state vector (Vallado 2013; Montenbruck and Gill 2000).

Physics residual for learning. Each LEO's PINN learns a surrogate trajectory $\hat{\mathbf{x}}_{\mathbf{w}}(t)$ whose dynamics satisfy the governing equation through the residual constraint

$$\mathcal{F}\big(\hat{\mathbf{x}}_{\mathbf{w}}(t)\big) = \dot{\hat{\mathbf{x}}}_{\mathbf{w}}(t) - f_{\text{2body+pert}}\big(\hat{\mathbf{x}}_\theta(t)\big) \approx 0,$$

as given in PINN formulations (Raissi, Perdikaris, and Karniadakis 2019).

Why PINNs Are Over-Parameterized and Must Be Compressed

PINNs for orbital prediction are deliberately over-parameterized to fit residuals reliably, yielding high-dimensional weights that must be synchronized across satellites. With low data rate, intermittent links, and long round-trip delays, transmitting full-precision parameters is infeasible in an SSA scenario. We therefore compress on-orbit updates by quantizing LEO gradients/parameter deltas before transmission—to meet bandwidth and latency constraints (Alistarh et al. 2017; Reisizadeh et al. 2020; Gersho and Gray 1992; Gray and Neuhoff 1998).

Research Question

How can physics-informed orbital dynamics be collaboratively learned over a decentralized, quantized, and privacy-preserving communication graph, while maintaining geometric and temporal coherence across heterogeneous multi-orbit constellations?

Physics-Aware Graph-DFL Optimization Formulation

Each satellite node $c \in \mathcal{V}$ jointly learns a physics-consistent local model subject to decentralized consensus, quantized communication, and privacy limits. Formally, the objective is given by

$$\min_{\{\mathbf{w}_c\}_{c\in\mathcal{V}}} \sum_{i\in\mathcal{V}} \left[\mathcal{L}_{\text{data},c}(\mathbf{w}_c) + \lambda_{\text{phys}} \left\| \ddot{\mathbf{r}}_c + \tfrac{\mu}{\|\mathbf{r}_c\|^3}\mathbf{r}_c - \mathbf{a}_{\text{drag},i} \right\|^2\right] \tag{23}$$

s.t.

$$(\mathbf{I}-\mathbf{P})^T[\mathbf{w}_1,\ldots,\mathbf{w}_N]^T = \mathbf{0}, \text{ (decentralized consensus)} \tag{24}$$

$$\Delta\widehat{\mathbf{w}}_c(t) = Q_b\big(\Delta\mathbf{w}_i(t)\big). \tag{25}$$

Here, $\mathcal{L}_{\text{data},c}$ denotes the data-fitting loss for client c, and the second term enforces adherence to the perturbed two-body dynamics (Vallado 2013; Montenbruck and Gill 2000; Curtis 2013). The consensus constraint aligns local models via the row-stochastic mixing matrix $\mathbf{P}$ (Nedić and Ozdaglar 2009; Xiao and Boyd 2004); The quantization constraints bound communication distortion under a b-bit quantizer Q_b (Alistarh et al. 2017; Reisizadeh et al. 2020). This formulation unifies physics-constrained learning, quantized communication, and decentralized consensus within a cohesive Graph-DFL optimization framework (Koloskova, Stich, and Jaggi 2019; Lian et al. 2017).

Proposed Solution Framework

System setup

Let the satellite network be a graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}), \quad \mathcal{V} = \mathcal{V}_L \cup \mathcal{V}_M,$$

with LEO clients $\mathcal{V}_c$ and MEO relays $\mathcal{V}_m$. Client $c \in \mathcal{V}_c$ observes an interleaved stream $\mathcal{D}_c = \{(t_i, \mathbf{y}_i)\}_{i=1}^{N_c}$, where each $\mathbf{y}_i$ is a noisy mix of multiple satellites.

A GRU deinterleaver $f_c(\cdot;\mathbf{u}_c)$ assigns labels $\hat{z}_i \in \{1,\ldots,M\}$ and is trained in the $s+1$-th training epoch as follows:

$$\mathbf{u}_c(s+1) = \mathbf{u}_c(s) - \eta_g \nabla_{\mathbf{u}_c} \mathcal{L}_{\text{deint}}(f_c(\mathcal{D}_c;\mathbf{u}_c(s))), \tag{26}$$

with a temporal-smoothness regularizer (omitted for brevity). Fig. 2 gives a high-level block diagram of our system model.

Our Graph DFL Setup

Decentralized Learning with CM×LC Trust Modulation

We consider a constellation of M satellites collaboratively solving a physics-constrained learning problem under strict privacy and communication limits. Each LEO node c maintains a local PINN governed by two-body orbital dynamics with perturbations as given earlier:

$$\ddot{\mathbf{r}}(t) = -\mu\frac{\mathbf{r}(t)}{\|\mathbf{r}(t)\|^3} + a_{J_2}\big(\mathbf{r}(t)\big) + a_D\big(\mathbf{r}(t),\mathbf{v}(t)\big) + a_{\text{other}}(t), \tag{27}$$

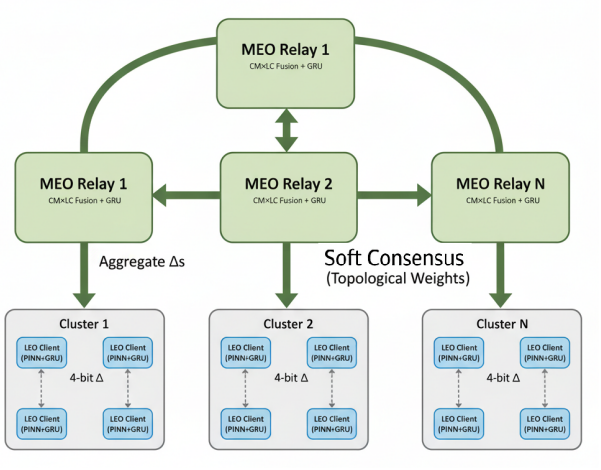

Figure 2: Two-tier LEO–MEO decentralized graph. LEO clients (PINN+GRU) exchange 4-bit quantized deltas within a cluster (dashed) and send aggregated updates to the cluster MEO relay (solid). Each MEO performs CM×LC-weighted fusion and runs a higher-level GRU; MEOs maintain soft consensus via bidirectional links weighted by Cayley–Menger (geometry) and Light-Cone (causality) factors. The serverless design yields global coherence through soft interpolation of model weights (Koloskova, Stich, and Jaggi 2019; Lian et al. 2017).

Each client locally optimizes its PINN parameters $\mathbf{w}_c(t)$ using only its own measurements and a GRU-based deinterleaver that separates mixed orbital streams into latent trajectories (Raissi, Perdikaris, and Karniadakis 2019; Chung et al. 2014).

Quantized Incremental Communication: Instead of transmitting raw trajectories or full parameters, each LEO client communicates compressed incremental model updates:

$$\Delta\widehat{\mathbf{w}}_c(t) = Q_4\big(\mathbf{w}_c(t) - \mathbf{w}_c(t-1)\big), \tag{28}$$

where $Q_4(\cdot)$ denotes a uniform 4-bit quantizer applied elementwise. The resulting quantization noise ε_q satisfies $\text{Var}(\varepsilon_q) \le \Delta^2/12 = \mathcal{O}(2^{-2b})$ for bit-width $b = 4$, ensuring bounded mean-square distortion (Alistarh et al. 2017).

CM×LC-Confidence Modulated Diffusion: Each node (client) c in the inter-satellite communication graph adapts its effective diffusion weights based on geometric and temporal trust scores obtained from its relay l:

$$\widetilde{\mathbf{P}}_{cj}(t) = \frac{\mathbf{P}_{cj}\,\nu_{j,c,l}(t)}{\sum_{k\in\mathcal{N}(c)}\mathbf{P}_{ck}\,\nu_{k,c,l}(t)}. \tag{29}$$

Here, $\nu_{j,\ell(c)}(t)$ combines Cayley–Menger (CM) and Light-Cone (LC) coherence between neighboring orbits, ensuring that spatially consistent and temporally co-visible neighbors exert greater influence.

Weight-Space Diffusion and Local Physics Update. Each client performs a trust-weighted update combining its

local physics correction and neighbor deltas:

$$\mathbf{w}_c(t{+}1) = \mathbf{w}_c(t) + \sum_{j\in\mathcal{N}(c)} \widetilde{\mathbf{P}}_{cj}(t)\,\Delta\widehat{\mathbf{w}}_j(t) - \eta\,\nabla_{\mathbf{w}}\mathcal{L}_c^{\mathrm{PINN}}\big(\mathbf{w}_c(t)\big), \tag{30}$$

where η is the local learning rate. The first term preserves the prior state, the second diffuses trusted quantized increments, and the third enforces physics-based self-consistency.

Delta-Space Consensus. An equivalent consensus formulation in the delta domain is

$$\Delta\mathbf{w}_c(t{+}1) = \Delta\mathbf{w}_c(t) - \eta\,\nabla_{\mathbf{w}}\mathcal{L}_c^{\mathrm{PINN}}\big(\mathbf{w}_c(t)\big) + \alpha\sum_{j\in\mathcal{N}_c}\widetilde{\mathbf{P}}_{cj}(t)\big(\Delta\widehat{\mathbf{w}}_j(t) - \Delta\widehat{\mathbf{w}}_c(t)\big), \tag{31}$$

where α modulates the diffusion strength. This term approximates a Laplacian flow on the graph, converging to a consensus subspace under bounded quantization noise.

CM×LC Soft Trust at MEO Relays. Each Medium Earth Orbit (MEO) relay l computes soft trust weights for its clients:

$$\nu_{c,l} = \frac{\exp\big(\beta_1\, w_{c,l}^{(\mathrm{CM})} + \beta_2\, w_{c,l}^{(\mathrm{LC})}\big)}{\sum_{j\in C_l}\exp\big(\beta_1\, w_{j,l}^{(\mathrm{CM})} + \beta_2\, w_{j,l}^{(\mathrm{LC})}\big)}, \quad \beta_1,\beta_2 \geq 0, \tag{32}$$

where β_1 and β_2 balance geometric (CM) and causal (LC) reliability. These weights act as a softmax-based trust distribution over clients in cluster C_l.

Intra-Cluster Fusion at MEO Relays. Index notation: l for relay, c for client, m, n for neighboring relays. Further, C_l: clients of Relay (l) and N_l: relay neighbors. Each relay l (MEO) fuses the updates from its constituent LEO clients into a regional model:

$$\Delta\mathbf{w}_{M,l}(t{+}1) = (1-\alpha_{\mathrm{intra}})\,\Delta\mathbf{w}_{M,l}(t) + \alpha_{\mathrm{intra}}\sum_{c\in\mathcal{C}_l}\nu_{c,l}\,\Delta\mathbf{w}_c(t{+}1), \tag{33}$$

where $\alpha_{\mathrm{intra}} \in (0,1)$ controls the balance between the relay's prior and aggregated states. This CM×LC-weighted fusion preserves geometric coherence and temporal alignment across correlated or co-planar orbits.

Trust-Aware Inter-Relay Consensus. MEO relays further synchronize through soft, topology-aware consensus:

$$\widetilde{\mathbf{P}}_{lm}^{(R)}(t) = \frac{\mathbf{P}_{lm}^{(R)}\,\nu_{m|l}^{(R)}(t)}{\sum_{n\in\mathcal{N}_l}\mathbf{P}_{ln}^{(R)}\,\nu_{n|l}^{(R)}(t)}, \tag{34}$$

Notation. Let $\mathbf{w}(t) \in \mathbb{R}^d$ denote the global model at round t, with local copies at the beginning of each round t as $\mathbf{w}_c(t) = \mathbf{w}(t)$ and incremental updates $\Delta\mathbf{w}_c(t) = \mathbf{w}_c(t) - \mathbf{w}(t)$. Each client transmits a quantized update given by

$$\Delta\widehat{\mathbf{w}}_c(t) = \Delta\mathbf{w}_c(t) + \boldsymbol{\epsilon}_c(t), \tag{35}$$

where $\boldsymbol{\epsilon}_c(t)$ is the quantization noise satisfying $\mathbb{E}[\boldsymbol{\epsilon}_c(t)] = 0$ and since the variance is bounded, assume $\mathbb{E}[\boldsymbol{\epsilon}_c(t)\boldsymbol{\epsilon}_c^T(t)] \preceq \sigma_q^2\mathbf{I}_d$ (Alistarh et al. 2017; Suresh et al. 2017).

1. Model and NTK linearization (lazy regime). Given data $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, labels $\mathbf{y}^* \in \mathbb{R}^n$, and predictions $f(\mathbf{X};\mathbf{w}) \in \mathbb{R}^n$, define the residual as follows:

$$\mathbf{e}(t) = f(\mathbf{x};\mathbf{w}(t)) - \mathbf{y}^*. \tag{36}$$

Applying Taylor series expansion and linearizing f around initialization ($\mathbf{w}_0$) gives

$$f(\mathbf{x};\mathbf{w}) \approx f(\mathbf{x};\mathbf{w}_0) + \Phi(\mathbf{w} - \mathbf{w}_0), \tag{37}$$

where Φ is the Jacobian of outputs w.r.t. parameters, and $H = \Phi\Phi^\top$ is the NTK Gram matrix (Jacot, Gabriel, and Hongler 2018). Assume $\kappa_0, \kappa_1 > 0$ and $\delta \in (0,1)$ such that

$$\lambda_{\min}\big(\mathbf{H}(0)\big) \geq \kappa_0, \quad \lambda_{\max}\big(\mathbf{H}(0)\big) \leq \kappa_1.$$

Assume NTK stability and good conditioning, i.e.,

$$(1-\delta)\kappa_0\mathbf{I} \preceq \mathbf{I}(t) \preceq \kappa_1 I,$$

and bounded proximity to its initialization, i.e.,

$$\|\mathbf{H}(t) - \mathbf{H}(0)\| \leq \delta\kappa_0.$$

2. CM×LC–weighted incremental updates (with quantization). Each client c owns index set $\mathcal{I}_c$ with selection matrix $S_c \in \{0,1\}^{n_c\times n}$ and projection $\mathbf{P}_c = S_c^\top S_c$. In the NTK (lazy) regime with feature map $\Phi \in \mathbb{R}^{m\times p}$ and residual $\mathbf{e}(t) \in \mathbb{R}^n$, one local step is

$$\Delta\mathbf{w}_c(t) = -\eta_c\,\Phi^\top S_c^\top\big(S_c\,\mathbf{e}(t)\big), \qquad \eta_c > 0, \tag{38}$$

and the global (serverless) aggregation uses CM×LC soft trusts $\nu_c(t) \geq 0$ with $\sum_c \nu_c(t) = 1$. Assume each client transmits a *quantized* increment

$$\Delta\widehat{\mathbf{w}}_c(t) = \mathcal{Q}_c\big(\Delta\mathbf{w}_c(t)\big) = \Delta\mathbf{w}_c(t) + \boldsymbol{\epsilon}_c(t), \tag{39}$$

Note that each client c maintains a dynamic/soft CM × LC soft trust weight, $\nu_c(t)$, given by

$$\nu_c(t) = \frac{\exp\big(\beta_1\, w_c^{(\mathrm{CM})}(t) + \beta_2\, w_c^{(\mathrm{LC})}(t)\big)}{\sum_{j=1}^C \exp\big(\beta_1\, w_j^{(\mathrm{CM})}(t) + \beta_2\, w_j^{(\mathrm{LC})}(t)\big)}, \tag{40}$$

6. Convergence Result

Theorem 1 Convergence under CM×LC–weighted NTK with unbiased quantization *Assume NTK stability* $(1-\delta)\kappa_0\mathbf{I} \preceq \mathbf{H} \preceq \kappa_1\mathbf{I}$*, coverage* $\sum_c \alpha_c\mathbf{P}_c \succeq \tau\eta\mathbf{I}$*, trust floor* $\nu_c \in [\nu_{\min}, \nu_{\max}]$*, and unbiased stochastic quantization. Let* $\mu_\nu = (1-\delta)\kappa_0\tau\eta$ *and* $L_\nu = \kappa_1\eta$*. If*

$$\widehat{\rho}^{\,2} := 1 - 2\mu_\nu + L_\nu^2 + \bar{s}_\nu\,\eta^2\,\kappa_1^2 \;<\; 1, \tag{41}$$

then for all $t \geq 0$*,*

$$\mathbb{E}\big[\|\mathbf{e}(t{+}1)\|_2^2\big] \leq \widehat{\rho}^{\,2}\,\mathbb{E}\big[\|\mathbf{e}(t)\|_2^2\big] + \kappa_1\,\bar{\sigma}_0^2, \tag{42}$$

$$\mathbb{E}\big[\|\mathbf{e}(t)\|_2^2\big] \leq \widehat{\rho}^{\,2t}\,\|\mathbf{e}(0)\|_2^2 + \frac{\kappa_1\,\bar{\sigma}_0^2}{1-\widehat{\rho}^{\,2}}. \tag{43}$$

Discussion

Comments on convergence result

In particular, the convergence rate and floor are controlled by η, b (via s_c), the CM×LC weights $\nu_c(t)$ (via $\bar{s}_\nu, \bar{\sigma}_0^2$), and the NTK spectrum (κ_0, κ_1) under coverage τ.

Theorem 1 establishes an important result, i.e., geometric decay of the prediction error in the NTK (lazy) regime to a steady-state floor set by the quantization variance σ_0^2.

The factor $\widehat{\rho}^2$ couples curvature (κ_0), conditioning (κ_1), learning rate (η), and graph coverage (τ). The additive term reflects quantization noise; with a uniform b-bit quantizer, $\mathrm{Var}(\varepsilon_q) = O(2^{-2b})$, so increasing precision lowers the floor without altering the linear rate.

Empirical alignment. The observed three-orders-of-magnitude drop in both data-fit and physics residual losses, followed by stabilization, is consistent with the predicted linear decay to a small steady-state floor under 4-bit communication.

Related Works

Federated & Decentralized Learning. FL enables collaborative training without sharing raw data (McMahan et al. 2017; Kairouz et al. 2021); FedProx tackles client heterogeneity (Li et al. 2020). Serverless, graph-based methods use peer-to-peer consensus for robustness and scale (Lian et al. 2017; Koloskova, Stich, and Jaggi 2019), but typically ignore physical priors.

Physics-Informed Neural Networks. PINNs embed governing equations into the loss, improving generalization under sparse data (Raissi, Perdikaris, and Karniadakis 2019; Karniadakis et al. 2021). Existing work is largely centralized.

SSA and On-Orbit Learning. Growing constellations and perturbations stress traditional pipelines. Hybrid ML+physics approaches exist, but most rely on ground servers. Our serverless CM×LC-weighted Graph-DFL aligns learning with orbital geometry and causal visibility.

Communication Efficiency. Quantization/sparsification reduce update bandwidth in FL (Alistarh et al. 2017; Reisizadeh et al. 2020). We extend these ideas to decentralized, physics-informed training and provide NTK-based convergence under 4-bit updates.

Experiments

Setup and Dataset

Experiments were conducted on real Two-Line Element (TLE)–derived orbital trajectories propagated via the SGP4 model (Hoots and Roehrich 1980; Vallado 2013), covering nine LEO satellites grouped into three clusters with MEO relays. Each LEO client hosts a GRU-based deinterleaver and a PINN trained on partial, noisy, interleaved measurements of Resident Space Objects (RSOs). MEO relays perform Cayley–Menger × Light-Cone (CM×LC)–weighted fusion of 4-bit quantized weight updates to preserve communication efficiency under the decentralized topology.

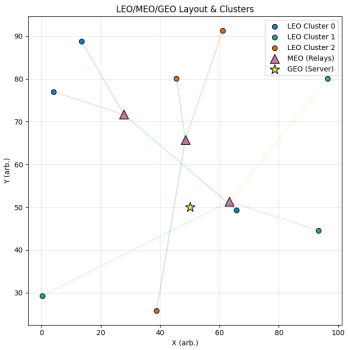

Figure 3: LEO–MEO–GEO layout and CM×LC clustering.

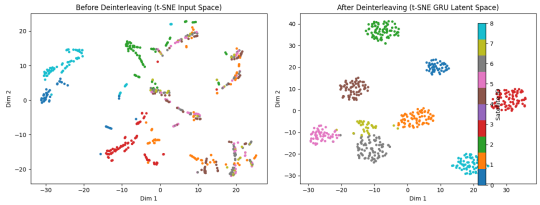

Figure 4: t-SNE embeddings of measurement sequences before and after GRU-based deinterleaving. *Left:* Raw (pre-training) input space shows mixed, overlapping manifolds due to interleaving. *Right:* GRU latent space (final round) forms nine well-separated clusters (color = satellite ID), indicating successful temporal disentangling and identity-consistent representations.

Performance Metrics

We report: (i) *Root Mean Square Error (RMSE)* between predicted and true orbital positions, (ii) *Deinterleaver accuracy* at LEO, and (iii) *Consensus metrics*—mean cosine similarity of GRU weights across peers.

Results and Analysis

Fig. 3 shows the layout of our satellite clusters.

Latent Representation Disentangling. Fig. 4 compares t-SNE embeddings of satellite telemetry before and after GRU-based deinterleaving, showing good separation after deinterleaving.

Physics Loss and Convergence. Figure 5 shows convergence of the data-fit and physics residual components of the PINN objective. Both terms decrease by over three orders of magnitude, with physics residuals stabilizing near zero, demonstrating well-balanced physical fidelity and data agreement.

Consensus and Accuracy. As shown in Fig. 6, MEO GRUs achieve rapid consensus (mean cosine similarity $\approx$ 0.6 within 10 rounds), while LEO GRUs align very gradually.

Trajectory Reconstruction. Orbit reconstruction for representative satellites (Fig. 7) shows that Graph-DFL's physics-consistent PINNs closely track true trajectories with negligible phase drift. LEO localization RMSE trends across rounds (Fig. 8) confirm stable convergence for all satellites.

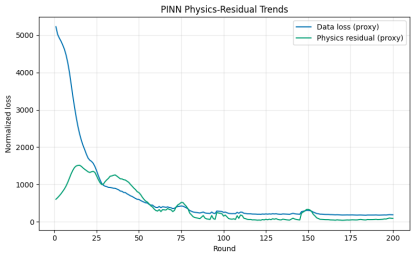

Figure 5: Convergence of PINN data and physics residual losses.

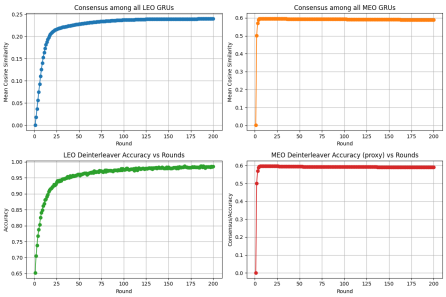

Figure 6: Consensus and deinterleaver accuracy across LEO and MEO layers.

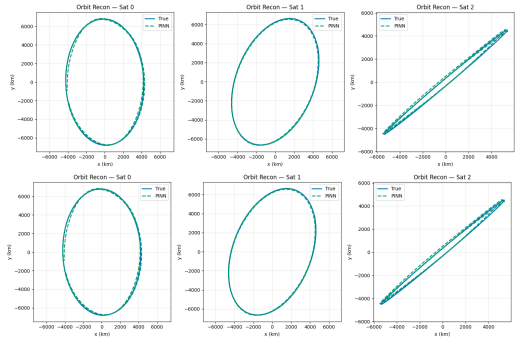

Figure 7: Orbit reconstruction for two representative satellites.

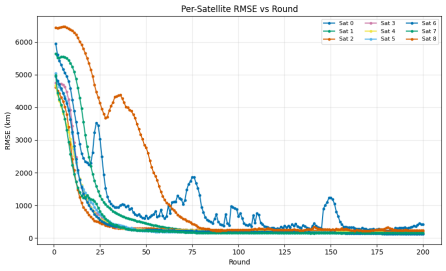

Figure 8: Per-satellite RMSE vs training round.

Discussion

The convergence result indicates that the expected residual error decays geometrically to a steady-state floor, governed by the quantization variance. Graph-DFL remains stable and physically consistent under quantized, decentralized communication because CM×LC weighting preserves spectral coverage, while quantization noise only sets a small error floor without destabilizing learning.

Experimental findings corroborate the theoretical predictions. Both data-fit and physics residual losses drop by over three orders of magnitude (Fig. 5), confirming stable NTK-regime convergence with quantized updates. t-SNE visualizations (Fig. 4) show clear separation of satellite clusters after GRU-based deinterleaving, indicating effective disentangling of interleaved orbital measurements. Consensus analysis (Fig. 6) shows rapid alignment among MEO relays (cosine similarity $\approx$ 0.6 within 10 rounds) and gradual improvement among LEO nodes, validating the benefit of CM×LC-weighted diffusion.

Quantized 4-bit communication achieves nearly $8\times$ bandwidth reduction while preserving convergence; quantization noise merely raises the steady-state floor, consistent with NTK bounds. Orbit reconstructions (Fig. 7) show sub-kilometer phase drift and $\sim$200–300 km mean RMSE across satellites, verifying that decentralized fusion remains physics-consistent and numerically stable.

Overall, Graph-DFL delivers physically grounded, communication-efficient, and privacy-preserving learning across multi-orbit constellations. By combining GRU-based deinterleaving, PINN physics modeling, and CM×LC-weighted consensus, a scalable foundation is formed for autonomous Space Situational Awareness and cooperative on-orbit intelligence.

Conclusion and Future Work

We presented Graph-DFL, a physics-aware decentralized learning framework that unites GRU-based deinterleaving, PINN-based orbital modeling, and CM×LC-weighted consensus for collaborative Space Situational Awareness. Experiments on real TLE/SGP4 trajectories demonstrated lower RMSE, stronger consensus, and stable convergence under 4-bit quantized communication, confirming both theoretical and practical robustness.

Future work will extend this framework to dynamic graphs, biased or adaptive quantizers, and inter-constellation coordination across LEO–MEO–GEO layers. Integrating uncertainty-aware physics priors and transformer-based PINNs could further enhance prediction fidelity and resilience for autonomous, large-scale orbital networks.

References

Alistarh, D.; Grubic, D.; Li, J.; Tomioka, R.; and Vojnović, M. 2017. QSGD: Communication-Efficient SGD via Gradient Quantization and Encoding. In *Advances in Neural Information Processing Systems*, volume 30, 1709–1720.

Blumenthal, L. M. 1953. *Theory and Applications of Distance Geometry*. Oxford University Press.

Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, volume 33, 1877–1901.

Caldas, F.; and Soares, C. 2024. Machine Learning in Orbit Estimation: A Survey. *Acta Astronautica*, 218: 188–206.

Chizat, L.; and Bach, F. 2018. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in Neural Information Processing Systems*, volume 31, 3036–3046.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. ArXiv:1412.3555.

Curtis, H. D. 2013. *Orbital Mechanics for Engineering Students*. Butterworth-Heinemann, 3 edition.

European Space Agency. 2025. ESA Space Environment Report 2025. Technical Report SDO-TN-10040, ESA Space Safety Programme Office. Issue 9, Rev. 1. ESA/ESOC, Darmstadt, Germany.

Gersho, A.; and Gray, R. M. 1992. *Vector Quantization and Signal Compression*. Springer.

Gray, R. M.; and Neuhoff, D. L. 1998. Quantization. *IEEE Transactions on Information Theory*, 44(6): 2325–2383.

Hoots, F. R.; and Roehrich, R. L. 1980. Spacetrack Report #3: Models for Propagation of the NORAD Element Sets. Technical Report Project Space Track Report No. 3, U.S. Air Force Aerospace Defense Command, Colorado Springs, CO. Approved for public release; distribution unlimited.

Jacot, A.; Gabriel, F.; and Hongler, C. 2018. Neural Tangent Kernel: Convergence and Generalization in Neural Networks. In *Advances in Neural Information Processing Systems*, volume 31, 8571–8580.

Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and Open Problems in Federated Learning. *Foundations and Trends in Machine Learning*, 14(1–2): 1–210.

Karniadakis, G. E.; Kevrekidis, I. G.; Lu, L.; Perdikaris, P.; Wang, S.; and Yang, L. 2021. Physics-informed machine learning. *Nature Reviews Physics*, 3(6): 422–440.

Kessler, D. J.; and Cour-Palais, B. G. 1978. Collision frequency of artificial satellites: The creation of a debris belt. *Journal of Geophysical Research: Space Physics*, 83(A6): 2637–2646.

Koloskova, A.; Stich, S. U.; and Jaggi, M. 2019. Decentralized Stochastic Optimization and Gossip Algorithms with Compressed Communication. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, 3478–3487.

Li, T.; Sahu, A. K.; Talwalkar, A.; and Smith, V. 2020. FedProx: Federated Optimization in Heterogeneous Networks. In *Proceedings of the 3rd MLSys Conference*, 429–450.

Lian, X.; Huang, Y.; Li, Y.; and Liu, J. 2017. Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel SGD. In *Advances in Neural Information Processing Systems*, volume 30, 5330–5340.

McDowell, J. C. 2020. The Low Earth Orbit Satellite Population and Impacts of the Starlink Constellation. Technical report, Harvard–Smithsonian Center for Astrophysics. Accessed: 2025-10-25.

McMahan, H. B.; Moore, E.; Ramage, D.; Hampson, S.; and Agüera y Arcas, B. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54, 1273–1282.

Melis, L.; Song, C.; De Cristofaro, E.; and Shmatikov, V. 2019. Exploiting Unintended Feature Leakage in Collaborative Learning. In *Proceedings of the IEEE Symposium on Security and Privacy*, 691–706.

Menger, K. 1930. Untersuchungen über allgemeine Metrik. *Mathematische Annalen*, 103(1): 466–501.

Montenbruck, O.; and Gill, E. 2000. *Satellite Orbits: Models, Methods, and Applications*. Springer.

Nedić, A.; and Ozdaglar, A. E. 2009. Distributed Subgradient Methods for Multi-Agent Optimization. *IEEE Transactions on Automatic Control*, 54(1): 48–61.

Olfati-Saber, R.; Fax, J. A.; and Murray, R. M. 2007. Consensus and Cooperation in Networked Multi-Agent Systems. *Proceedings of the IEEE*, 95(1): 215–233.

Olfati-Saber, R.; and Murray, R. M. 2004. Consensus Problems in Networks with Switching Topology and Time-Delays. *IEEE Transactions on Automatic Control*, 49(9): 1520–1533.

Raissi, M.; Perdikaris, P.; and Karniadakis, G. E. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378: 686–707.

Reisizadeh, A.; Mokhtari, A.; Hassani, H.; Ribeiro, A.; and Pedarsani, R. 2020. FedPAQ: A Communication-Efficient Federated Learning Method with Periodic Averaging and Quantization. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108, 2021–2031.

Rindler, W. 2006. *Relativity: Special, General, and Cosmological*. Oxford University Press, 2 edition.

Suresh, A. T.; Yu, F. X.; Kumar, S.; and McMahan, H. B. 2017. Distributed Mean Estimation with Limited Communication. In *Advances in Neural Information Processing Systems*, volume 30, 3329–3339.

Vallado, D. A. 2013. *Fundamentals of Astrodynamics and Applications*. Microcosm Press, 4 edition.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, volume 30, 5998–6008.

Xiao, L.; and Boyd, S. 2004. Fast Linear Iterations for Distributed Averaging. *Systems & Control Letters*, 53(1): 65–78.

Zhu, L.; Liu, Z.; and Han, S. 2019. Deep Leakage from Gradients. In *Advances in Neural Information Processing Systems*, volume 32, 14774–14784.