

# Energy-Aware Federated Learning in Satellite Constellations

Nasrin Razmi<sup>\*†</sup>, Bho Matthiesen<sup>\*†</sup>, Armin Dekorsy<sup>\*†</sup>, and Petar Popovski<sup>‡\*</sup>

<sup>\*</sup> Dept. of Communications Engineering, University of Bremen, Germany

<sup>†</sup>Gauss-Olbers Space Technology Transfer Center, University of Bremen, Germany

<sup>‡</sup> Dept. of Electronic Systems, Aalborg University, Denmark

Emails: {razmi, matthiesen, dekorsy}@ant.uni-bremen.de, petarp@es.aau.dk

**Abstract**—Federated learning (FL) in satellite constellations, where the satellites collaboratively train a machine learning (ML) model, is a promising technology towards enabling globally connected intelligence and the integration of space networks into terrestrial mobile networks. The energy required for this computationally intensive task is provided either by solar panels or by an internal battery if the satellite is in Earth’s shadow. Careful management of this battery and system’s available energy resources is not only necessary for reliable satellite operation, but also to avoid premature battery aging. We propose a novel energy-aware computation time scheduler for satellite FL, which aims to minimize battery usage without any impact on the convergence speed. Numerical results indicate an increase of more than  $3\times$  in battery lifetime can be achieved over energy-agnostic task scheduling.

**Index Terms**—Federated learning, satellite constellation, battery lifetime, energy-aware, scheduling, orbital edge computing.

## I. INTRODUCTION

Driven by the desire for ubiquitous global connectivity, satellite communications and the space industry are undergoing a major transformation [1]. Especially the paradigm shift towards mega-constellations of interconnected small satellites in low Earth orbits led towards space networks nowadays being regarded as a major component of 6G (and beyond) networks [2], [3]. An integral component, both in managing [4] and utilizing [5] these constellations, is distributed machine learning (ML). Among the architectural approaches towards integrating ML into space networks [6], satellite federated learning (SFL) [7], where the ML training is performed directly on each satellite, appears to be most promising strategy in the long run. A closely related use case is orbital edge computing [8], where terrestrial devices offload computing tasks, including ML training [9], to satellites.

A considerable challenge in implementing computationally intensive processes in satellite systems is their energy consumption. While, theoretically, solar energy is available in abundance in near-Earth space, the amount of energy a satellite can harvest is practically limited by the size of its solar panels. Moreover, the nature of orbital mechanics places

satellites regularly in Earth’s shadow, necessitating the usage of batteries to bridge these outages in solar energy. These aspects lead to a limited availability of energy within the satellite. Since this energy is not only used for (secondary) computational tasks but also to supply critical satellite subsystems, careful energy management and scheduling of computing time is strictly necessary to ensure reliable satellite operations [10]. Furthermore, the lifetime of batteries is limited by the number of charge/discharge cycles they can withstand. Since replacing a failing battery in a deployed satellite is infeasible, inadequate battery management can have detrimental effect on the lifetime of a satellite system [11], [12]. This is not only undesirable from an economical perspective, but also environmentally unsustainable.

To this end, we consider energy-aware task scheduling for SFL with the goal of maximizing the battery lifetime. SFL is the application of the federated learning (FL) paradigm [13] in satellite constellations. It was first studied in [14] for ground-assisted FL, where the limited connectivity between satellites and ground station (GS) was identified as the main impairment towards reasonable training performance. Subsequent works focus primarily on improving the algorithm in [14], e.g., [15]–[19], and enhancing connectivity [20]–[23]. While energy aspects of SFL are considered in [24]–[26], none of these works considers long-term effects of SFL on the battery lifetime. Our proposed algorithm leverages the predictability of sunlight periods and satellite operations to schedule computation time for SFL with respect to the energy demands of more critical satellite subsystems, while using any remaining degrees of freedom to minimize strain on the battery to extend its lifetime. While we focus primarily on SFL, the proposed algorithm should also be applicable to a wide range of related orbital edge computing scenarios.

## II. SYSTEM MODEL

### A. Constellation Configuration

We consider a constellation with  $Q$  orbital planes, where each orbit  $q \in \{1, \dots, Q\}$  contains  $K_q$  satellites. The set of all satellites in the  $Q$  orbital planes is denoted with  $\mathcal{K}$ , with a total number of satellites  $K = \sum_{q=1}^Q K_q$ . Any satellite  $k$  completes an orbit around Earth within a period

This work was funded in part by the German Research Foundation (DFG) under Germany’s Excellence Strategy (EXC 2077 at University of Bremen, University Allowance).

---

**Algorithm 1** Satellite Learning Procedure
 

---

```

1: procedure SATLEARNPROC( $w^n$ )
2:   initialize  $w_k^{n,0} = w^n$ ,  $m = 0$ , learning rate  $\eta$ 
3:   for  $M$  epochs do ▷  $M$  epochs of mini-batch SGD
4:      $\mathcal{D}_k \leftarrow$  Randomly shuffle  $\mathcal{D}_k$ 
5:      $\mathcal{B} \leftarrow$  Partition  $\mathcal{D}_k$  into mini-batches of size  $B$ 
6:     for each batch  $\mathcal{B} \in \mathcal{B}$  do
7:        $w_k^{n,m} \leftarrow w_k^{n,m} - \frac{\eta}{|\mathcal{B}|} \nabla_w (\sum_{z \in \mathcal{B}} g(z, w))$ 
8:     end for
9:      $m \leftarrow m + 1$ 
10:  end for
11:  return  $D_k w_k^{n,M}$ 
12: end procedure
  
```

---

of  $\iota_k = 2\pi\sqrt{\frac{o_k^3}{\mu}}$ , where  $o_k$  is the semi-major axis and  $\mu = 3.98 \times 10^{14} \text{ m}^3/\text{s}^2$  is the geocentric gravitational constant. For circular orbital planes, semi-major axis is  $o_k = r_E + \nu_k$ , where  $r_E = 6371 \text{ km}$  and  $\nu_k$  are the Earth radius and satellite's altitude above the Earth's surface, respectively.

### B. Computation Model

Each satellite  $k$  gathers dataset  $\mathcal{D}_k$  using its on-board instruments and trains an ML model. The overall goal of satellites is to collaboratively solve an optimization problem

$$\min_{w \in \mathbb{R}^d} \frac{1}{D} \sum_{z \in \mathcal{D}} g(z; w) = \min_{w \in \mathbb{R}^d} \sum_{k \in \mathcal{K}} \frac{D_k}{D} \sum_{z \in \mathcal{D}_k} \frac{1}{D_k} g(z; w) \quad (1)$$

and learn the global model parameters  $w$ . In (1),  $\mathcal{D} = \bigcup_{k \in \mathcal{K}} \mathcal{D}_k$  is the set of data samples of all satellites with the size of  $D = \sum_{k \in \mathcal{K}} D_k$ , where  $D_k = |\mathcal{D}_k|$  is the size of dataset of satellite  $k$ . Moreover,  $g(z; w)$  is defined as the loss for a data sample  $z \in \mathcal{D}$  and model parameters  $w$ . The problem (1) is orchestrated by one or several parameter servers (PSs) and is solved in overall  $N$  iterations without sharing datasets between satellites.

In iteration  $n$ , the satellite  $k$  performs  $M$  local epochs of mini-batch stochastic gradient descent (SGD) as Algorithm 1 to minimize the local loss function  $\frac{1}{D_k} \sum_{z \in \mathcal{D}_k} h(z, w)$  [20]. It finally derives the local model parameters  $w_k^{n,M}$  and transmits it to the PS.

### C. Satellite Energy Management

1) *Sunlight and eclipse pattern*: Satellites are equipped with solar panels to harvest the Sun's radiant energy for power generation in sunlight periods. After each sunlight period, the satellite experiences periods of eclipse, as denoted in Fig. 1. In an eclipse period, the satellite relies on the energy stored in its battery, usually a Lithium-ion battery, which is charged during sunlight periods. For any designated duration from  $t_0$  to  $t_1$ , we denote the start of each sunlight period and its subsequent eclipse as  $t_{s,j}$  and  $t_{e,j}$ , respectively, where  $j \in \mathcal{J} = \{1, \dots, J\}$  represents the indices of these periods as depicted in Fig. 1.

2) *Energy Consumption*: Satellites utilize the harvested or battery energy to perform various tasks. We categorize the tasks into two groups: non-training, which include all tasks except ML training, and training which refers only to the ML

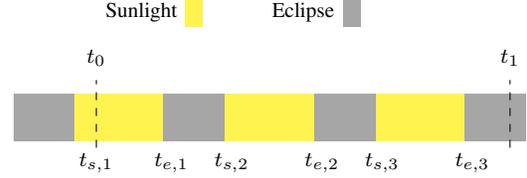


Fig. 1: Satellite's sunlight and eclipse pattern

training task. The energy demand of the non-training tasks is determined a priori by a different subsystem and provided as parameters  $E_{s,j}^d$  and  $E_{e,j}^d$  to the SFL task scheduler, where  $E_{s,j}^d$  and  $E_{e,j}^d$  are the energy demands of non-training tasks during the intervals  $[t_{s,j}, t_{e,j}]$  and  $[t_{e,j}, t_{s,j+1}]$ , i.e., the  $j$ th sunlight and eclipse periods, respectively. For the training task, the satellites need to train an ML model for a duration of  $T_c$ , which results in consuming  $P_c T_c$  amount of energy, where  $P_c$  is the required power for training. We assume that the required training time  $T_c$  is smaller than the available time  $t_1 - t_0$ . Then, the total energy consumption of the satellite is  $E_{s,j}^d + \tau_{s,j} P_c$  during sunlight periods, and  $E_{e,j}^d + \tau_{e,j} P_c$  during eclipse periods, where  $\tau_{s,j}$  and  $\tau_{e,j}$  are the computation times scheduled for ML training during these intervals. Clearly,  $0 \leq \tau_{s,j} \leq t_{e,j} - t_{s,j}$  and  $0 \leq \tau_{e,j} \leq t_{s,j+1} - t_{e,j}$ . Further, to be able to complete ML training, we require

$$\sum_{j=1}^J \tau_{e,j} + \tau_{s,j} = T_c. \quad (2)$$

3) *Satellite battery*: Let  $b(t) \in [0, B_{\max}]$  be the battery's charge level at time instant  $t$ , where  $B_{\max}$  is the battery's capacity. During eclipse periods, the satellite's energy demand is met solely from the battery. Thus, the remaining charge  $b_{s,j+1} = b(t_{s,j+1})$  at the end of eclipse period  $j$  is

$$b_{s,j+1} = b_{e,j} - E_{e,j}^d - P_c \tau_{e,j}, \quad (3)$$

where  $b_{e,j} = b(t_{e,j})$  is the battery level at the beginning of eclipse period  $j$ . During sunlight periods, the battery is recharged from the solar panels. Provided the harvested energy during sunlight period  $j$ , i.e., in the time interval  $[t_{s,j}, t_{e,j}]$  is  $E_{s,j}^h$ , the energy available for charging the battery is  $E_{s,j}^h - E_{s,j}^d - P_c \tau_{s,j}$ . Thus, the battery level at the begin of the following eclipse period is

$$b_{e,j} = \min\{b_{s,j} + E_{s,j}^h - E_{s,j}^d - P_c \tau_{s,j}, B_{\max}\}. \quad (4)$$

This implies that excess energy goes to waste once the battery is fully charged. Since the battery should not be discharged further during sunlight periods, we require

$$E_{s,j}^h - E_{s,j}^d - P_c \tau_{s,j} \geq 0, \quad (5)$$

which implies  $E_{s,j}^h \geq E_{s,j}^d + P_c \tau_{s,j}$ .

4) *Battery aging*: To extend the operational life of the satellites, it is crucial to prolong the lifetime of the batteries. An effective parameter to evaluate the battery's lifetime is cycle life, defined as the number of charge and discharge cycles a battery can complete before its performance degradation

[27]. With each charge/discharge cycle, the battery degrades. The battery will last longer with a smaller discharge level [28].

The proportion of energy extracted from the battery within an eclipse duration relative to its full capacity is defined as depth of discharge (DoD), expressed as  $d(t) = \frac{B_{\max} - b(t)}{B_{\max}}$  at any given time  $t$ . In this paper, we focus on the effect of DoD on cycle life [29]. A higher DoD results in a shorter battery lifetime, meaning a higher consumption of cycle life. The cycle life consumption for Lithium-ion batteries is defined as

$$l(t_1, t_2) = \int_{d(t_1)}^{d(t_2)} 10^{a(d-1)} (1 + a \ln 10 \cdot d) dd \quad (6)$$

$$= 10^{a(d(t_2)-1)} d(t_2) - 10^{a(d(t_1)-1)} d(t_1),$$

if  $d(t_2) > d(t_1)$ ; otherwise, it is 0, and  $a$  is a positive constant which is determined by the battery specifications [29].

### III. CYCLE LIFE MINIMIZATION

We aim to schedule the computation time for ML training such that all non-training energy demands are met and the cost on the batteries life is minimized. Since the battery is never discharged during sunlight periods, the total cost on the battery lifetime during  $[t_0, t_1]$  is, following Section II-C4,

$$\sum_{j=1}^J 10^{a(\hat{d}_{e,j}-1)} \hat{d}_{e,j} - 10^{a(d_{e,j}-1)} d_{e,j}, \quad (7)$$

where  $d_{e,j} = d(t_{e,j})$  and  $\hat{d}_{e,j} = d(t_{s,j+1})$  are the DoD at the beginning and end of the  $j$ th eclipse period  $[t_{e,j}, t_{s,j+1}]$ , respectively. Thus, the optimal ML training schedule with respect to the battery lifetime is the solution to

$$\min_{\substack{\forall j: \tau_{s,j}, \tau_{e,j}, b_{e,j}, \\ b_{s,j+1}, d_{e,j}, \hat{d}_{e,j}}} \sum_{j=1}^J 10^{a(\hat{d}_{e,j}-1)} \hat{d}_{e,j} - 10^{a(d_{e,j}-1)} d_{e,j} \quad (8a)$$

$$\text{s.t. } b_{s,j+1} = b_{e,j} - E_{e,j}^d - P_c \tau_{e,j}, \forall j \in \mathcal{J} \quad (8b)$$

$$b_{e,j} = \min\{b_{s,j} + E_{s,j}^h - E_{s,j}^d - P_c \tau_{s,j}, B_{\max}\}, \forall j \in \mathcal{J}, \quad (8c)$$

$$E_{s,j}^h - E_{s,j}^d - P_c \tau_{s,j} \geq 0, \forall j \in \mathcal{J}, \quad (8d)$$

$$d_{e,j} = \frac{B_{\max} - b_{e,j}}{B_{\max}}, \forall j \in \mathcal{J}, \quad (8e)$$

$$\hat{d}_{e,j} = \frac{B_{\max} - b_{s,j+1}}{B_{\max}}, \forall j \in \mathcal{J}, \quad (8f)$$

$$\sum_{j=1}^J \tau_{s,j} + \tau_{e,j} = T_c, \quad (8g)$$

$$0 \leq \tau_{s,j} \leq t_{e,j} - t_{s,j}, \forall j \in \mathcal{J}, \quad (8h)$$

$$0 \leq \tau_{e,j} \leq t_{s,j+1} - t_{e,j}, \forall j \in \mathcal{J}, \quad (8i)$$

$$0 \leq b_{s,j+1} \leq B_{\max}, \forall j \in \mathcal{J}, \quad (8j)$$

$$0 \leq b_{e,j} \leq B_{\max}, \forall j \in \mathcal{J}, \quad (8k)$$

with  $b_{s,1} = B_0$ , where  $B_0$  is the initial battery charge at  $t_0$ . Conditions (8b) and (8c) signify the battery level at the beginning of the sunlight and eclipse periods, respectively.

Additionally, (8d) implies that the battery should not be discharged during sunlight periods. DoD at the beginning and end of the eclipse periods are defined in (8e) and (8f). The training periods in the sunlight and eclipse periods should meet (8g), with constraints defined as (8h) and (8i) for these periods, respectively. Moreover, (8j) and (8k) indicate that the battery level must remain above 0 and below the battery's capacity.

#### A. Equivalent Problem

Problem (8) is nonconvex due to (8a) and (8c). However, we can relax (8c) to obtain

$$\min_{\substack{\forall j: \tau_{s,j}, \tau_{e,j}, b_{e,j}, \\ b_{s,j+1}, d_{e,j}, \hat{d}_{e,j}}} \sum_{j=1}^J 10^{a(\hat{d}_{e,j}-1)} \hat{d}_{e,j} - 10^{a(d_{e,j}-1)} d_{e,j} \quad (9a)$$

$$\text{s.t. } b_{e,j} \leq \min\{b_{s,j} + E_{s,j}^h - E_{s,j}^d - P_c \tau_{s,j}, B_{\max}\}, \forall j \in \mathcal{J}, \quad (9b)$$

$$(8b), (8d)-(8k), \quad (9c)$$

which has a convex feasible set and is equivalent to (8).

*Lemma 1:* Any optimal solution to (9) is an optimal solution to (8) (and vice versa).

*Proof:* Consider problem (8). For an arbitrary  $n \in \mathcal{J}$ , replace (8c) by (9b) and let  $\tau_{s,1}^*, \dots, \hat{d}_{e,j}^*$  be an optimal solution to this new problem. Assume this solution satisfies

$$b_{e,n}^* < \min\{b_{s,n}^* + E_{s,n}^h - E_{s,n}^d - P_c \tau_{s,n}^*, B_{\max}\}. \quad (10)$$

Then, we can find a  $\tilde{b}_{e,n} = b_{e,n}^* + \varepsilon$  with  $\varepsilon > 0$  that satisfies (9b) and (8k). From (8b), we obtain  $\tilde{b}_{s,n+1} = b_{s,n+1} + \varepsilon$ . Further,  $\tilde{d}_{e,n} = d_{e,n}^* - \delta$  and  $\tilde{\hat{d}}_{e,n} = \hat{d}_{e,n}^* - \delta$  with  $\delta = \frac{\varepsilon}{B_{\max}} > 0$ . The cycle life consumption for these DoDs is

$$10^{a(\tilde{d}_{e,n}^* - \delta - 1)} (\hat{d}_{e,n}^* - \delta) - 10^{a(d_{e,n}^* - \delta - 1)} (d_{e,n}^* - \delta) \quad (11)$$

$$= 10^{-a\delta} \left( 10^{a(\hat{d}_{e,n}^* - 1)} \hat{d}_{e,n}^* - 10^{a(d_{e,n}^* - 1)} d_{e,n}^* \right) - \delta \left( 10^{a(\hat{d}_{e,n}^* - \delta - 1)} - 10^{a(d_{e,n}^* - \delta - 1)} \right) \quad (12)$$

$$\leq 10^{-a\delta} \left( 10^{a(\hat{d}_{e,n}^* - 1)} \hat{d}_{e,n}^* - 10^{a(d_{e,n}^* - 1)} d_{e,n}^* \right) \quad (13)$$

$$\leq 10^{a(\hat{d}_{e,n}^* - 1)} \hat{d}_{e,n}^* - 10^{a(d_{e,n}^* - 1)} d_{e,n}^* \quad (14)$$

where (13) follows from  $\hat{d}_{e,n} \geq d_{e,n}$  and (14) is due to  $a, \delta > 0$ . Finally, for all  $j > n$ , we also obtain new feasible  $\tilde{b}_{e,j} \geq b_{e,j}^*$  and  $\tilde{b}_{s,j+1} \geq b_{s,j+1}^*$  that further decrease the objective value (by the same argument as before). This implies that  $\tau_{s,1}^*, \dots, \hat{d}_{e,j}^*$  cannot be an optimal solution to (9) and, further, that any optimal solution satisfies (9b) with equality. ■

Due to this equivalence, we can obtain a solution to (8) by solving (9). However, (9) is still a nonconvex optimization problem due to the objective being a difference of convex functions (DC). Thus, we cannot solve it directly with conventional convex optimization methods. Instead, we design

an iterative algorithm based on the majorization-minimization principle [30] to find stationary points of (9).

#### IV. FIRST-ORDER OPTIMAL SOLUTION ALGORITHM

Problem (9) falls in the class of DC programming problems with convex feasible set, i.e., a continuous optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}) = u(\mathbf{x}) - v(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{X}, \quad (15)$$

where  $\mathcal{X}$  is a closed convex subset of  $\mathbb{R}^n$  and  $u, v$  are continuously differentiable convex functions on  $\mathcal{X}$ . The concave-convex procedure [31], [32] obtains stationary points of (15) by solving a sequence of convex programs

$$\mathbf{x}^{(l+1)} \in \arg \min_{\mathbf{x} \in \mathcal{X}} u(\mathbf{x}) - \mathbf{x}^T \nabla v(\mathbf{x}^{(l)}). \quad (16)$$

This is an instance of the more general majorization-minimization framework where the nonconvex part of the objective is linearized. Within this context, the objective of (16) is known as the *surrogate function* of  $f(\mathbf{x})$ .

##### A. Concave-convex procedure for (9)

For ease of notation, define the vectors  $\mathbf{d} = (d_{e,1}, \dots, d_{e,J})$ ,  $\tilde{\mathbf{d}} = (\tilde{d}_{e,1}, \dots, \tilde{d}_{e,J})$ , and  $\mathbf{x}$  to hold all optimization variables in (9). We identify  $\mathcal{X}$  as  $\{\mathbf{x} \mid (8b), (9b), (8d)-(8k)\}$ ,  $u(\mathbf{x}) = \sum_{j=1}^J 10^{a(\tilde{d}_{e,j}-1)} \tilde{d}_{e,j}$ , and, with a minor abuse of notation,  $v(\mathbf{x}) = v(\mathbf{d}) = \sum_{j=1}^J 10^{a(d_{e,j}-1)} d_{e,j}$  in (16). First, observe that this is indeed a DC program since  $\mathcal{X}$  is a convex set (trivial) and  $u, v$  are convex functions, as established next.

*Lemma 2:* For  $a > 0$ , the function  $\sum_{j=1}^J y_j 10^{a(y_j-1)}$  is strictly convex on  $\mathbb{R}_+^J$  with respect to  $y_1, \dots, y_J$ , i.e., for all  $y_j \geq 0$ ,  $1 \leq j \leq J$ .

*Proof:* The Hessian of  $h = \sum_{j=1}^J y_j 10^{a(y_j-1)}$  is a diagonal matrix with the partial derivatives  $\partial^2 h / \partial y_j^2$  on the diagonal. Thus, it is strictly positive definite if and only if  $\partial^2 h / \partial y_j^2 > 0$  for all  $j$ . From  $\partial^2 h / \partial y_j^2 = a \ln 10 (a y_j \ln 10 + 2) 10^{a(y_j-1)}$ , we see that this is the case if and only if  $a y_j \ln 10 + 2 > 0$  for  $a > 0$ . Thus,  $y_j \geq 0$  is sufficient for strong convexity. ■

*Corollary 1:* The functions  $u$  and  $v$  are convex on  $\mathcal{X}$ .

*Proof:* From (8e), (8f), (8j), (8k), we have  $\tilde{d}_{e,j} \geq 0$  and  $d_{e,j} \geq 0$  for all  $j$ . Since  $a > 0$  by definition, Lemma 2 is applicable. ■

Observe that the nonconvexity in (9) stems solely from  $v(\mathbf{d})$ . Thus, we linearize the objective only in  $\mathbf{d}$  to obtain the following surrogate problem

$$\mathbf{x}^{(l+1)} \in \arg \min_{\mathbf{x} \in \mathcal{X}} u(\mathbf{x}) - \mathbf{d}^T \nabla v(\mathbf{d}^{(l)}), \quad (17)$$

where

$$\mathbf{d}^T \nabla v(\mathbf{d}^{(l)}) = \sum_{j=1}^J d_{e,j} 10^{a(\tilde{d}_{e,j}-1)} (1 + a \tilde{d}_{e,j} \ln 10).$$

This results in the concave-convex procedure stated in Algorithm 2. After initialization, it solves the convex program

---

#### Algorithm 2 Concave-convex procedure for (9)

---

- 1: **Initialize**  $l = 0$ ,  $\lambda > 0$ ,  $\varepsilon > 0$ , and  $\mathbf{x}^{(0)}$  to some feasible point of (9).
  - 2: **repeat**
  - 3:     Set  $\mathbf{x}^{(l+1)}$  to an optimal point of (17)
  - 4:     Update  $l \leftarrow l + 1$
  - 5: **until**  $\|\mathbf{d}^{(l)} - \mathbf{d}^{(l-1)}\| \leq \varepsilon$
- 

in line 3 for  $\mathbf{x}^{(l)}$  and sets the next approximation point  $\mathbf{x}^{(l+1)}$  to an optimal solution of this problem. Note that  $\mathbf{d}^{(l+1)}$  is implicitly defined as it is part of  $\mathbf{x}^{(l+1)}$ . This is repeated until convergence in  $\mathbf{d}^{(l)}$  is observed. Upon termination, the final  $\mathbf{d}^{(l)}$  will be within an  $\varepsilon$ -region of the point  $\mathbf{d}^*$ , where  $\mathbf{d}^*$  is such that the corresponding  $\mathbf{x}^*$  is a stationary point of (9). The initial  $\mathbf{x}^{(0)}$  is set to a feasible point (9), which is easily obtained due to the feasible set  $\mathcal{X}$  being convex.

Convergence of Algorithm 2 is established formally below. Suitable convergence results for (16) are derived in Theorems 4 and 8 of [32]. There, Theorem 8 provides much stronger convergence guarantees than Theorem 4 but requires strong convexity of  $u$  and  $v$  in  $\mathbf{x}$ . This is not the case here, as established in Corollary 1. However, the proof of [32, Thm. 8] can be modified to cover (17) as follows.

*Theorem 1:* For all  $\varepsilon \geq 0$ , the sequence  $\{\mathbf{x}^{(l)}\}_l$  generated by Algorithm 2 converges towards a point  $\bar{\mathbf{x}}$ . For  $\varepsilon = 0$ , this is a stationary point of (9). Otherwise, Algorithm 2 terminates after a finite number of iterations and  $\bar{\mathbf{x}}$  is within an  $\delta$ -region of some stationary point of (9) for some small  $0 < \varepsilon < \delta$ , i.e.,  $\|\mathbf{x}^* - \bar{\mathbf{x}}\| \leq \delta$  for some stationary point  $\mathbf{x}^*$ .

*Proof:* Observe that the feasible set  $\mathcal{X}$  is a finite intersection of affine constraints. By Slater's condition, these constraints are qualified if some feasible point exists [33, Sec. 5.2.3]. Further, define the point-to-set map  $\mathcal{A}$  (cf. [32, Sec. 3]) as the projection of (17) onto  $\mathbf{d}$ . Let  $\mathbf{d}^*$  be a generalized fixed point of  $\mathcal{A}$  and  $\mathbf{x}^*$  the corresponding solution of (17). Then, by the same argument as in [32, Lemma 5],  $\mathbf{x}^*$  is a stationary point of (9). Moreover, note that the right-hand side of (17) is equivalent to

$$\arg \min_{\mathbf{x} \in \mathcal{X}} u(\mathbf{x}) - v(\mathbf{d}^{(l)}) - (\mathbf{d} - \mathbf{d}^{(l)})^T \nabla v(\mathbf{d}^{(l)}) \quad (18)$$

Thus, for any  $\mathbf{d}^{(l)} \neq \mathbf{d}^{(l+1)}$ ,

$$\begin{aligned} f(\mathbf{x}^{(l+1)}) &= u(\mathbf{x}^{(l+1)}) - v(\mathbf{d}^{(l+1)}) \\ &< u(\mathbf{x}^{(l+1)}) - v(\mathbf{d}^{(l)}) - (\mathbf{d}^{(l+1)} - \mathbf{d}^{(l)})^T \nabla v(\mathbf{d}^{(l)}) \\ &= g(\mathbf{x}^{(l+1)}, \mathbf{x}^{(l)}) \leq g(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}) = f(\mathbf{x}^{(l)}) \end{aligned}$$

due to  $v$  being strongly convex in  $\mathbf{d}$ . This establishes that the map  $\mathcal{A}$  is strictly monotonic with respect to  $f$ . Since the feasible set  $\mathcal{X}$  is closed and bounded, i.e., compact, its projection onto  $\mathbf{d}$  is as well and, with [32, Remark 7],  $\mathcal{A}$  is uniformly compact and closed. Hence, by virtue of [32, Thm. 3], all limit points of  $\{\mathbf{d}^{(l)}\}_l$  are fixed points of  $\mathcal{A}$  and  $\|\mathbf{d}^{(l)} - \mathbf{d}^{(l-1)}\| \rightarrow 0$ . Further, since  $v$  is a strictly increasing function in  $\mathbf{d}$ , the set of fixed points of  $\mathcal{A}$  is finite and  $\{\mathbf{d}^{(l)}\}_l$  converges to a single limit point  $\mathbf{d}^*$ . The associated point  $\mathbf{x}^*$  is a stationary point of (9) (see argument above).

Finally, since the sequence  $\{f(\mathbf{d}^{(l)})\}_l$  is strictly decreasing and  $\|\mathbf{d}^{(l)} - \mathbf{d}^{(l-1)}\|$  converges continuously to 0, there exists some  $L < \infty$  such that  $\|\mathbf{d}^{(l)} - \mathbf{d}^{(l-1)}\| < \varepsilon$  for all  $l > L$  if  $\varepsilon > 0$ . It is trivial to show that  $\|\mathbf{d}^{(l)} - \mathbf{d}^{(l-1)}\| < \varepsilon$  implies  $\|\mathbf{x}^{(l)} - \mathbf{x}^{(l-1)}\| < \delta$  for some  $\delta > \varepsilon$ . ■

Algorithm 2 does not guarantee that the stationary point is a local minimum of (9). However, unless  $\mathbf{x}^{(0)}$  is stationary, which would result in Algorithm 2 terminating after the first iteration, the obtained point cannot be a local maximum due to  $\{u(\mathbf{x}^{(l)}) - v(\mathbf{d}^{(l)})\}_l$  being strictly decreasing.

Convergence of Algorithm 2 can be strengthened to  $\bar{\mathbf{x}}$  being within an  $\varepsilon$ -region of a stationary point of (9) by replacing the termination criterion with  $\|\mathbf{x}^{(l)} - \mathbf{x}^{(l-1)}\| \leq \varepsilon$ . This, however, requires regularization in (17) with  $\lambda\|\mathbf{x} - \mathbf{x}^{(l)}\|^2$  for some small  $\lambda > 0$  to ensure convergence of  $\{\mathbf{x}^{(l)}\}_l$  to a single limit point.

## V. PERFORMANCE EVALUATION

We evaluate the performance of the proposed energy-aware FL approach by considering 20 satellites from the Starlink constellation [34]. Supported by two GSs, one located in Germany and the other in Japan, these satellites participate in a synchronous FL process for 96 hours [14]. We partition this period into  $N$  equal time-slots, each equivalent to an iteration of FL.

At the end of the  $n$ th time-slot, one of the GSs, which is chosen as the PS, updates the global model parameters as  $\mathbf{w}^{n+1} = \sum_{k=1}^K \alpha_k^n \frac{D_k}{D} \mathbf{w}_k^{n,M}$ , where the local model parameters of the  $k$ th satellite,  $\mathbf{w}_k^{n,M}$ , is received either directly or through the other GS, and  $\alpha_k^n = 1$  if the satellite  $k$  participates in the  $n$ th iteration, otherwise  $\alpha_k^n = 0$ . During each time-slot, only those satellites can participate in FL which are capable of receiving the global model parameters, training the ML model for a period of  $T_c$ , and returning their updated local model parameters to the PS. Note that to communicate the model parameters, the satellite should be visible to one of the GSs. Afterwards, the PS sends back the updated global model parameter  $\mathbf{w}^{n+1}$  to the satellites for the next iteration. Upon receiving the updated parameters, each participating satellite schedules its training, taking into account the predictability of both visibility to GSs and sunlight/eclipse periods.

We compare DoD and consumed cycle life for the battery of the satellites using our proposed approach, which we call energy-aware, with those of the state-of-the-art one [14], which we call energy-agnostic. In the energy-aware approach, in each time-slot, the participating satellites solve (9) to decide how much time they should assign in each sunlight or eclipse for training the ML model. However, in the energy-agnostic one, the participants start the training after receiving the model parameters without considering being in sunlight or eclipse.

With  $N = 50$  time-slots and two training duration of  $T_c = 20$  min and  $T_c = 80$  min, Fig. 2 shows DoD for the satellite  $k = 2$  during the time-slots in which it can participate in FL. The power consumption for training the ML model

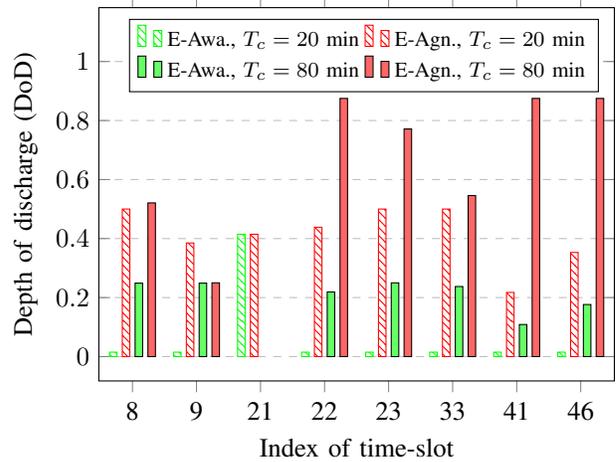


Fig. 2: Depth of discharge (DoD) for the battery of satellite  $k = 2$  with respect to indexes of time-slot. E-Agn. and E-Awa. stand respectively for energy-agnostic and energy-aware FL algorithms and  $T_c$  denotes required time for training the ML model.

is set to 50 W, the battery capacity to 2000 W min, and the battery specification parameter  $a$  to 0.8 as [29]. We assume the satellites consume energy solely for training the ML model. Additionally, the harvested energy during sunlight suffices to charge fully the battery. As we see in Fig. 2, the satellite  $k = 2$  participates in FL process only in the time-slots  $\{8, 9, 21, 22, 23, 33, 41, 46\}$  when  $T_c = 20$  min, and the same time-slots except 21 when  $T_c = 80$  min due to its visibility pattern. During time-slot 21, there is not sufficient time, from the instant that the satellite receives the global model parameters to the latest possible time it can return the updated local model to the PS, to accommodate the required training duration of  $T_c = 80$  min.

As we see in Fig. 2, the energy-aware approach leads to a notably lower DoD compared to the energy-agnostic approach. Specifically, except for time-slot 21, the DoD remains at 0 for all other time-slots in the energy-aware approach. This is because the satellite prioritizes utilizing sunlight periods for training without drawing energy from its battery. Only when sunlight periods are insufficient, the satellite starts to use a portion of the eclipse duration for the remaining training which draws energy from its battery, as observed in the time-slot 21. Moreover, the algorithm strives to evenly distribute the remaining training time across all eclipse periods to minimize the impact on cycle life. However, with the energy-agnostic approach, DoD is significantly higher since the satellite trains the model regardless of the current energy source.

Fig. 3 shows the consumed cycle life of the battery of satellites on average over the total 96 hours with respect to the battery capacity while  $T_c = 80$  min. As we see, the energy-aware approach makes satellites consume lower cycle life of their battery. Specifically, if we consider a battery with capacity of 2000 W min for satellites, by energy-agnostic algorithm, 2.88 cycles are consumed, whereas by the energy-aware algorithm, only 0.76 cycles are used, meaning over

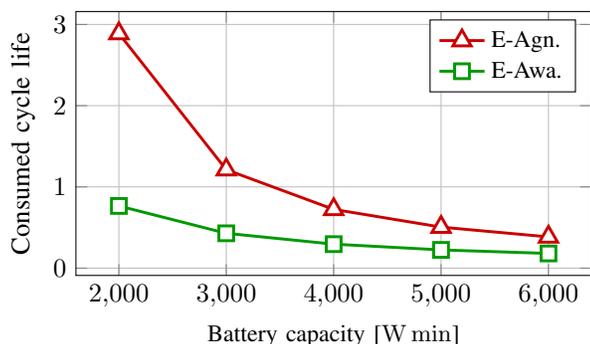


Fig. 3: Consumed cycle life with respect to battery capacity during 96 hours, considering required time for training of  $T_c = 80$  min. E-Agn. and E-Awa. stand respectively for energy-agnostic and energy-aware FL algorithms.

three-fold more cycle life consumption with energy-agnostic. Considering a battery with a total cycle life of 800 [28], if the satellite employs the energy-agnostic approach, it can operate only for approximately 3 years. However, by the energy-aware approach, the satellite’s operational lifetime extends to over 11 years.

## VI. CONCLUSIONS

Satellites use solar energy during sunlight periods but depend on their batteries during eclipses. However, frequent use of batteries decreases their lifetime. To enhance the lifetime of satellite batteries, in this paper, we formulated an optimization problem. The aim is to schedule on-board FL training tasks such that the use of the satellite’s batteries is reduced. We solved the optimization problem with successive convex approximation. Our numerical results show that the introduced approach increases the lifetime of batteries significantly.

## REFERENCES

- [1] M. Sweeting, “Modern small satellites — changing the economics of space,” *Proc. IEEE*, vol. 106, no. 3, pp. 343–361, 2018.
- [2] I. Leyva-Mayorga *et al.*, “LEO small-satellite constellations for 5G and beyond-5G communications,” *IEEE Access*, vol. 8, 2020.
- [3] M. Y. Abdelsadek *et al.*, “Future space networks: Toward the next giant leap for humankind,” *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 949–1007, Feb. 2023.
- [4] G. Fontanesi *et al.*, “Artificial intelligence for satellite communication and non-terrestrial networks: A survey,” Apr. 2023, arXiv:2304.13008. [Online]. Available: <https://arxiv.org/abs/2304.13008v1>
- [5] D. Izzo, G. Meoni, P. Gómez, D. Dold, and A. Zochbauer, “Selected trends in artificial intelligence for space applications,” Dec. 2022. [Online]. Available: <https://arxiv.org/abs/2212.06662>
- [6] H. Chen, M. Xiao, and Z. Pang, “Satellite-based computing networks with federated learning,” *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 78–84, Feb. 2022.
- [7] B. Matthiesen, N. Razmi, I. Leyva-Mayorga, A. Dekorsy, and P. Popovski, “Federated learning in satellite constellations,” *IEEE Netw.*, 2023, early access.
- [8] C. Wu *et al.*, “A comprehensive survey on orbital edge computing: Systems, applications, and algorithms.” [Online]. Available: <http://arxiv.org/abs/2306.00275>
- [9] X. Wang *et al.*, “Convergence of edge computing and deep learning: A comprehensive survey,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 869–904, 2020.

- [10] F. Alagoz and G. Gur, “Energy efficiency and satellite networking: A holistic overview,” *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1954–1979, 2011.
- [11] F. Chen, Q. Wang, and Y. Ran, “Dynamic routing algorithm for maximizing battery life in LEO satellite networks,” in *2022 IEEE 8th Int. Conf. on Computer and Commun. (ICCC)*, 2022, pp. 671–676.
- [12] J. Liu, B. Zhao, Q. Xin, J. Su, and W. Ou, “DRL-ER: An intelligent energy-aware routing protocol with guaranteed delay bounds in satellite mega-constellations,” *IEEE Trans. on Netw. Science and Engineer.*, vol. 8, no. 4, pp. 2872–2884, 2021.
- [13] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Artif. Intell. Statist.*, FL, USA, Apr. 2017.
- [14] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, “Ground-assisted federated learning in LEO satellite constellations,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 717–721, 2022.
- [15] Y. Zhou *et al.*, “Decomposition and Meta-DRL based multi-objective optimization for asynchronous federated learning in 6G-satellite systems,” *IEEE J. on Sel. Areas in Commun.*, pp. 1–1, 2024.
- [16] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, “Scheduling for ground-assisted federated learning in LEO satellite constellations,” in *Eur. Signal Proc. Conf. (EUSIPCO)*, Belgrade, Serbia, Aug. 2022.
- [17] C. Yang *et al.*, “Communication-efficient satellite-ground federated learning through progressive weight quantization,” pp. 1–14.
- [18] Z. Zhai *et al.*, “FedLEO: An offloading-assisted decentralized federated learning framework for low earth orbit satellite networks,” *IEEE Trans. on Mobile Comput.*, vol. 23, no. 5, pp. 5260–5279.
- [19] Z. Lin *et al.*, “FedSN: A novel federated learning framework over LEO satellite networks.” [Online]. Available: <http://arxiv.org/abs/2311.01483>
- [20] N. Razmi, B. Matthiesen, A. Dekorsy, and P. Popovski, “On-board federated learning for satellite clusters with inter-satellite links,” *IEEE Trans. on Commun.*, pp. 1–1, 2024.
- [21] —, “Scheduling for on-board federated learning with satellite clusters,” in *2023 IEEE Globecom Workshops*, 2023, pp. 426–431.
- [22] M. Elmahallawy, T. Luo, and K. Ramadan, “Communication-efficient federated learning for LEO satellite networks integrated with HAPs using hybrid NOMA-OFDM,” *IEEE J. Sel. Areas Commun.*, 2024.
- [23] Y. Shi *et al.*, “Satellite federated edge learning: Architecture design and convergence analysis.” [Online]. Available: <http://arxiv.org/abs/2404.01875>
- [24] Z. Yan and D. Li, “Convergence time optimization for decentralized federated learning with LEO satellites via number control,” *IEEE Trans. on Vehicular Technology*, vol. 73, no. 3, pp. 4517–4522, 2024.
- [25] C. Wu, Y. Zhu, and F. Wang, “DSFL: Decentralized satellite federated learning for energy-aware LEO constellation computing,” in *2022 IEEE Int. Conf. on Satellite Comput.*, 2022, pp. 25–30.
- [26] D.-J. Han, S. Hosseinalipour, D. J. Love, M. Chiang, and C. G. Brinton, “Cooperative federated learning over ground-to-satellite integrated networks: Joint local computation and data offloading,” *IEEE J. on Sel. Areas in Commun.*, pp. 1–1, 2024.
- [27] M. Qardran, N. Jenkins, and J. Wu, “Smart grid and energy storage,” in *McEvoy’s Handbook of Photovoltaics*. Elsevier, 2018, pp. 915–928.
- [28] C. E. Inc. (2020) Bu-808: How to prolong Lithium-based batteries. [Online]. Available: <https://batteryuniversity.com/article/bu-808-how-to-prolong-lithium-based-batteries>.
- [29] Y. Yang, M. Xu, D. Wang, and Y. Wang, “Towards energy-efficient routing in satellite networks,” *IEEE J. on Sel. Areas in Commun.*, vol. 34, no. 12, pp. 3869–3886, 2016.
- [30] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. on Signal Proc.*, vol. 65, no. 3, pp. 794–816, 2017.
- [31] A. L. Yuille and A. Rangarajan, “The concave-convex procedure,” *Neural Comput.*, vol. 15, no. 4, pp. 915–936, 2003.
- [32] B. K. Sriperumbudur and G. Lanckriet, “On the convergence of the concave-convex procedure,” in *Adv. in Neural Inf. Proc. Syst.*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., vol. 22. Curran Associates, Inc., 2009.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [34] Celestrak. Accessed: Feb. 21, 2024. [Online]. Available: <https://celestrak.org/NORAD/elements/gp.php?GROUP=starlink&FORMAT=tle>