Wiley
International Transactions on Electrical Energy Systems
Volume 2025, Article ID 9877968, 20 pages
https://doi.org/10.1155/etep/9877968

WILEY

Research Article

Intelligent Wireless Power Scheduling for Lunar Multienergy Systems: Deep Reinforcement Learning for Real-Time Adaptive Beam Steering and Vehicle-to-Grid Energy Optimization

Thomas Tongxin Li,[1] **Shuangqi Li**,[2] **Cynthia Xin Ding**,[3] **Zhaoyao Bao**,[4] **and Mohannad Alhazmi**[5]

[1]*Department of Electrical and Computer Engineering, Iowa State University, Ames 50011, Iowa, USA*
[2]*Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon 999077, Hong Kong, China*
[3]*School of Education, Johns Hopkins University, Baltimore 21218, Maryland, USA*
[4]*Civil and Environmental Engineering, Cornell University, Ithaca, New York, USA*
[5]*Electrical Engineering Department, College of Applied Engineering, King Saud University, P.O. Box 2454, Riyadh 11451, Saudi Arabia*

Correspondence should be addressed to Shuangqi Li; shuangqi.li@polyu.edu.hk

Received 20 February 2025; Accepted 12 May 2025

Academic Editor: Yang Gao

The integration of wireless power transfer (WPT) and vehicle-to-grid (V2G) technologies is essential for the sustainable operation of lunar multienergy virtual power plants (MEVPPs), where rovers, habitats, and in situ resource utilization (ISRU) facilities rely on adaptive energy management. Unlike terrestrial systems, lunar environments present extreme challenges, including long-duration night cycles, regolith dust accumulation, severe temperature fluctuations, and dynamic rover mobility, all of which disrupt efficient power delivery. This paper proposes a reinforcement learning–based adaptive beam steering framework to optimize WPT scheduling, ensuring continuous and efficient energy transmission for both mobile and stationary lunar assets. Unlike traditional fixed-beam or heuristic-based WPT methods, the proposed system utilizes deep reinforcement learning (DRL) with proximal policy optimization (PPO) to autonomously adjust beam direction, power intensity, and charging priority in response to real-time rover movements, V2G interactions, and fluctuating energy demands. The proposed framework models WPT optimization as a Markov decision process (MDP), where the agent learns to dynamically adapt beam steering based on rover speed, response delay, solar power availability, and charging station congestion. The reward function penalizes energy misallocation and misalignment losses while maximizing charging efficiency and systemwide energy resilience. A case study simulating a 30-day mission near Shackleton Crater evaluates the effectiveness of the AI–driven WPT system, demonstrating a 54.6% reduction in energy downtime and a 41.3% improvement in beam alignment efficiency compared to static power scheduling methods. In addition, the system reduces latency-induced power deficits by 39.8%, ensuring reliable power distribution for ISRU oxygen extraction, habitat life support, and rover recharging stations. This study represents a novel advancement in lunar power infrastructure, integrating AI–driven adaptive WPT with intelligent energy scheduling to enhance V2G interactions in extraterrestrial environments. The results validate the feasibility of DRL–based WPT control, paving the way for scalable, resilient, and self-optimizing wireless power grids on the Moon. Future work will explore the integration of hybrid energy storage models, quantum-inspired optimization for real-time decision-making, and predictive beamforming algorithms to further enhance the reliability and efficiency of lunar energy networks.

Keywords: adaptive beam steering; energy resilience; lunar multienergy systems; proximal policy optimization (PPO); reinforcement learning; vehicle-to-grid (V2G); wireless power transfer (WPT)

1. Introduction

The development of sustainable energy infrastructures for extraterrestrial habitats is a critical challenge in modern space exploration. Future missions to the Moon, Mars, and other celestial bodies require robust, autonomous, and adaptable energy management systems capable of supplying continuous power to a diverse set of infrastructure, including lunar habitats, rovers, in situ resource utilization (ISRU) facilities, and scientific instruments [1]. Unlike terrestrial power grids, which benefit from well-established generation and distribution networks, lunar energy systems face significant operational constraints, such as prolonged lunar nights, extreme temperature fluctuations, regolith dust accumulation, dynamic power demand, and the absence of an atmospheric medium for convection-based cooling [2]. Among the various energy distribution strategies for lunar missions, wireless power transfer (WPT) has emerged as a transformative technology capable of enabling efficient and flexible energy transmission without requiring a physically connected power grid [3]. Several experimental demonstrations have validated the feasibility of space-based power beaming, bridging the gap between theoretical models and practical deployment. NASA's Space Solar Power Exploratory Research and Technology (SERT) program has investigated microwave power transmission (MPT) for extraterrestrial applications, demonstrating the ability to beam energy across long distances with high efficiency [4]. Similarly, JAXA's WPT experiments have successfully transmitted microwave energy over hundreds of meters on Earth, providing critical insights into beamforming precision and transmission losses in space environments. These prior studies establish a strong foundation for implementing WPT in lunar missions. The proposed reinforcement learning (RL)–based framework builds upon these advancements by introducing adaptive beam steering and real-time energy optimization, ensuring efficient power allocation despite environmental uncertainties. By leveraging artificial intelligence (AI)–driven dynamic control, this study aims to further advance the feasibility of WPT for future lunar energy networks. Among the various energy distribution strategies for lunar missions, WPT has emerged as a transformative technology capable of enabling efficient and flexible energy transmission without requiring a physically connected power grid. Several existing power transmission technologies have been explored in space missions, each with unique advantages and limitations [5]. Microwave-based WPT, which is the focus of this study, has been widely considered due to its high transmission efficiency, long-range capabilities, and ability to operate in a vacuum without significant atmospheric attenuation. However, beam divergence increases over long distances, necessitating adaptive beam steering techniques to maintain efficiency. Alternatively, laser-based WPT offers a highly collimated energy beam, minimizing dispersion and enabling long-range energy transmission beyond 5 km, which is a limitation of microwave-based approaches. However, laser WPT suffers from lower energy conversion efficiency, higher sensitivity to dust accumulation, and the risk of optical misalignment in dynamic environments. Another method, that is, inductive coupling–based WPT, has been successfully used in low-power space applications, such as satellite docking stations and proximity power transfer systems. While highly efficient for short distances, inductive WPT is not well-suited for large-scale lunar energy distribution due to its limited range and reliance on close physical proximity between the transmitter and the receiver. By positioning the proposed RL–based microwave WPT framework within the broader spectrum of space-based power transmission techniques, this study highlights the advantages of adaptive beam steering and intelligent power scheduling, ensuring reliable and scalable energy distribution in extraterrestrial environments [6]. While this study primarily focuses on microwave-based WPT, alternative approaches such as laser-based power transmission have also been explored in space applications. Microwave WPT is advantageous due to its high transmission efficiency in atmospheric and vacuum environments, but its beam divergence increases beyond 5 km, significantly reducing energy reception efficiency. Conversely, laser-based WPT offers a highly collimated beam, minimizing energy dispersion over long distances and making it a promising alternative for power delivery to distant lunar assets beyond 5 km. However, laser-based systems face higher conversion losses at both transmission and reception stages, and their performance is highly sensitive to regolith dust accumulation and beam obstruction. Given these trade-offs, a hybrid WPT approach combining microwave for midrange power transmission and laser for long-range energy beaming could potentially enhance lunar power distribution efficiency. This study focuses on microwave WPT optimization, while future work will explore the feasibility of integrating laser-based transmission for extended-range energy delivery [7]. WPT enables the direct beaming of energy to mobile and stationary units, allowing for seamless power delivery across a distributed lunar network. However, existing WPT frameworks primarily rely on fixed-schedule power transmission, failing to account for real-time variations in power demand, environmental interference, and dynamic movement of energy receivers (e.g., rovers and autonomous ISRU units). The lack of intelligent, adaptive scheduling mechanisms significantly reduces energy efficiency, introduces transmission losses, and leads to suboptimal resource allocation in complex lunar environments. Recent advances in AI and RL provide an opportunity to revolutionize WPT scheduling, allowing the system to autonomously learn optimal power allocation strategies and adapt in real time to dynamic mission conditions. This paper proposes a RL–based adaptive beam steering framework to optimize WPT scheduling, ensuring continuous and efficient energy transmission for both mobile and stationary lunar assets. To address the challenges posed by lunar dust accumulation, extreme temperature fluctuations, and potential signal interference, the proposed system integrates an adaptive recalibration mechanism that dynamically adjusts beam alignment and power intensity in response to environmental uncertainties, thereby mitigating long-term efficiency degradation [8].

The proposed framework employs a RL–based approach to optimize WPT scheduling within a lunar multienergy virtual power plant (MEVPP). The system is modeled as a Markov decision process (MDP), where the state space includes real-time information on receiver positions, battery charge levels, solar energy availability, charging station occupancy, and regolith dust accumulation. However, given the high variability in mission tasks and energy consumption patterns, relying solely on real-time data may lead to suboptimal long-term decision-making. To enhance scheduling stability, the framework integrates historical mission data and predictive analytics, allowing the system to anticipate future energy demands based on past operational trends. By incorporating these predictive elements, the model can proactively adjust power allocations, mitigating the impact of sudden energy fluctuations and improving overall system resilience. The action space consists of power allocation decisions, beamforming adjustments, and charging prioritization, while the reward function is designed to maximize energy efficiency while penalizing energy deficits and transmission losses [9]. A deep RL (DRL) model based on proximal policy optimization (PPO) is developed to train an adaptive policy for WPT scheduling. The PPO algorithm is selected due to its ability to handle high-dimensional state-action spaces and provide stable convergence, making it ideal for large-scale, data-driven energy optimization problems [10]. The DRL agent learns optimal power transmission policies by interacting with a simulated lunar environment, continuously refining its decisions through policy gradient updates [11]. To ensure scalability, multiagent RL (MARL) principles are integrated into the framework, allowing multiple power nodes to collaboratively optimize energy distribution. To prevent conflicting energy allocation decisions among agents, the framework employs a hierarchical coordination mechanism, where a global energy dispatcher acts as a supervisory agent, providing high-level constraints on total power availability, transmission priorities, and fairness constraints. Each individual WPT node functions as an independent agent, learning to optimize local power transmission while adhering to global consensus rules enforced by the dispatcher. In addition, interagent communication is facilitated through a decentralized consensus protocol, where agents exchange real-time energy demand, power congestion status, and beam alignment updates to ensure nonconflicting power allocations. A soft-update rule is incorporated to prevent abrupt fluctuations in transmission assignments, ensuring that energy distribution remains stable across the system. This coordinated MARL approach allows each node to dynamically adapt to fluctuating energy demands while maintaining systemwide stability and fairness in power allocation. The introduction of global supervision and decentralized agent communication significantly enhances the robustness of RL–based WPT scheduling in lunar environments [12]. This multiagent architecture ensures that power transmission decisions remain decentralized yet coordinated, allowing for scalable deployment across future lunar base architectures [13]. The proposed approach is evaluated through a high-fidelity case study, simulating a lunar mission scenario near Shackleton Crater, where the system's performance is tested against variable solar power availability, extreme temperature gradients, and diverse rover mobility patterns. Comparative results with fixed-schedule WPT and rule-based heuristic scheduling demonstrate the superior efficiency, adaptability, and resilience of the proposed DRL–based model. This paper introduces a novel DRL–based WPT scheduling framework for lunar energy management, presenting four major contributions as follows.

1.1. RL–Based Adaptive WPT Model. Unlike traditional WPT systems that rely on static transmission schedules, this paper introduces a learning-based adaptive model that dynamically adjusts power allocations in real time, responding to changes in energy demand, receiver mobility, and environmental conditions.

1.2. MDP Formulation for Lunar WPT Optimization. The proposed system is formulated as a complex, high-dimensional MDP, integrating power allocation, beam steering, and charging prioritization into a single optimization framework. This allows for holistic decision-making, where the model learns the most efficient energy distribution strategy under uncertain and dynamic lunar conditions.

1.3. PPO–Based DRL for Real-Time Learning. This paper leverages PPO–based DRL training, enabling the WPT system to learn optimal policies through continuous interaction with the lunar environment. The use of PPO ensures stable convergence, robust performance under stochastic conditions, and computational efficiency suitable for large-scale deployment in future lunar bases.

1.4. Comprehensive Performance Evaluation With a 30-Day Lunar Mission Simulation. The proposed model is rigorously tested in a realistic lunar mission environment, where the DRL–optimized WPT scheduling strategy is compared against conventional fixed-schedule and heuristic-based methods. In addition, an extended experiment evaluates the impact of integrating historical mission data into the RL framework. The results indicate that incorporating predictive analytics improves power scheduling efficiency by 17.3% and reduces emergency power deficits by 12.6% over a 30-day mission. The findings demonstrate that leveraging past operational data enhances long-term energy management, reducing unexpected fluctuations and ensuring more consistent power delivery for critical lunar operations [14]. The results demonstrate that the proposed system reduces energy downtime by 52.4%, improves power transmission efficiency by 38.9%, and decreases energy congestion by 41.2%, making it a groundbreaking advancement for lunar energy management.

2. Literature Review

The development of sustainable power infrastructure is a critical challenge for space exploration, particularly for long-term lunar and Martian missions. Unlike terrestrial energy systems, which benefit from a stable grid infrastructure, extraterrestrial environments require highly flexible, autonomous, and efficient power management strategies [15]. The need for adaptable energy distribution is amplified by the unique constraints of lunar operations, including prolonged night cycles, extreme temperature variations, regolith dust interference, and the absence of a stable atmosphere for heat dissipation. Traditional wired power grids are impractical in such environments due to deployment challenges, vulnerability to environmental hazards, and the difficulty of maintaining permanent infrastructure on rugged and dynamically evolving surfaces [16]. As a result, WPT has emerged as a promising solution, offering the capability to beam energy efficiently to mobile and stationary units without the limitations of physical wiring [17]. Existing WPT research has primarily focused on terrestrial applications such as electric vehicle charging, consumer electronics, and medical implants. However, the extension of WPT technology to space applications introduces additional complexities, such as energy beam alignment in the absence of atmospheric stabilization, transmission losses due to dust accumulation, and the need for real-time power optimization to accommodate fluctuating energy demands [18]. Several studies have investigated the feasibility of microwave and laser-based energy beaming for lunar applications, demonstrating the potential of WPT as a viable power distribution method [19]. However, these approaches generally assume static power transmission schedules and fail to incorporate intelligent decision-making frameworks capable of dynamically adjusting energy allocation in response to real-time mission requirements. This limitation underscores the necessity for an adaptive, RL–based approach to WPT scheduling, capable of autonomously optimizing power distribution under varying lunar conditions [20].

Traditional energy distribution for space missions has relied on wired power grids, battery storage, and nuclear reactors to ensure continuous power availability. Battery-based energy storage, such as lithium-ion, lithium-sulfur, and solid-state batteries, has been widely employed in planetary rovers and landers, including NASA's Curiosity and Perseverance missions [21]. These storage systems provide a reliable energy source but are inherently limited by capacity constraints, degradation over multiple charge cycles, and the inability to dynamically reallocate power to mobile units. Wired power grids, as proposed for lunar habitats under NASA's Artemis program, offer a structured energy distribution mechanism but face significant deployment challenges, particularly in harsh extraterrestrial environments [1]. The installation of wired transmission lines on the lunar surface is impractical due to frequent regolith displacement, potential mechanical failures, and the inflexibility of fixed-position infrastructure. In addition to battery and wired grid solutions, nuclear power has been explored as a long-term energy source for extraterrestrial applications. NASA's Kilopower project and similar initiatives have investigated small-scale nuclear fission reactors designed to provide continuous power for lunar and Martian bases. While nuclear reactors offer a reliable energy supply independent of solar availability, their integration into a flexible, decentralized energy distribution system remains an unresolved challenge. These conventional power solutions, while valuable in isolated applications, lack the adaptability and scalability required for complex multienergy systems where power demands fluctuate dynamically [22].

WPT has gained considerable attention as an alternative to traditional wired energy distribution, particularly for its potential applications in extraterrestrial environments. The primary advantage of WPT lies in its ability to deliver energy without requiring fixed transmission infrastructure, making it particularly suitable for mobile assets such as lunar rovers, ISRU units, and scientific instruments deployed across vast surface areas. Among the various WPT technologies explored for space applications, MPT and laser energy beaming have demonstrated significant potential. Studies on microwave-based WPT have highlighted its efficiency in transmitting energy over long distances with minimal loss, with proposals such as Japan's Space-Based Solar Power (SBSP) system envisioning the deployment of geostationary satellites to beam energy directly to lunar surface operations [23]. Similarly, laser-based WPT systems have been explored as a means of high-precision, long-range energy delivery, with experimental demonstrations showing promising results in achieving targeted power transmission. Despite the potential benefits of WPT for extraterrestrial energy distribution, existing studies remain largely theoretical and do not account for the operational complexities involved in real-time lunar power scheduling. Most WPT research assumes fixed energy allocation strategies, failing to incorporate adaptive optimization frameworks that respond dynamically to fluctuating power demands, environmental disruptions, and mobility patterns of energy receivers. The lack of intelligent control mechanisms capable of optimizing beam alignment, prioritizing critical energy loads, and dynamically adjusting transmission parameters in response to real-time mission conditions represents a major gap in the current literature. This gap highlights the need for an advanced WPT scheduling framework that integrates RL–based optimization techniques to enable autonomous decision-making in complex extraterrestrial power networks [24].

3. Mathematical Modeling

The optimization of WPT scheduling and adaptive beam steering for lunar MEVPPs requires a robust mathematical framework that accurately models energy dynamics, receiver mobility, and beam alignment efficiency. Unlike terrestrial grid–based power distribution, lunar environments introduce unique constraints such as high-energy latency due to long transmission distances, fluctuating solar availability, regolith dust–induced power losses, and extreme thermal variations affecting energy conversion efficiency. In addition, the mobility of rovers and ISRU units necessitates real-time adjustments in beam direction and power allocation, ensuring minimal energy wastage while maximizing operational reliability. This section formulates the WPT beam steering and energy scheduling problem as a multiobjective optimization model, integrating spatiotemporal energy distribution constraints, battery state-of-charge (SoC) evolution, power balancing conditions, and transmission efficiency degradation due to misalignment effects. To systematically address these challenges, we define a set of objective functions and constraints that govern the adaptive WPT system, considering factors such as beamforming precision, real-time power redistribution, charging priorities, and system resilience under dynamic conditions. The first objective function focuses on maximizing total WPT efficiency by optimizing beam alignment and minimizing transmission losses, ensuring that both stationary and mobile receivers receive power in a timely manner. The second objective aims to reduce energy downtime, mitigating the risk of power shortages due to misalignment errors or response delays. The third function minimizes overall transmission losses by accounting for beam divergence, lunar terrain interference, and thermal effects. Lastly, we introduce an optimization function that prioritizes power allocation based on receiver criticality, ensuring that high-priority systems, such as habitat life support and ISRU operations, maintain uninterrupted power even under fluctuating energy conditions. To enhance real-world applicability, the model also incorporates emergency response mechanisms that dynamically adjust energy priorities in response to critical failures or rapid operational changes. Specifically, in the event of a sudden communication failure, energy allocation shifts toward autonomous system resilience, prioritizing onboard energy storage for essential functions such as navigation and hazard detection until communication is restored. Similarly, during rapid rover redeployment for urgent scientific tasks or terrain changes, the WPT scheduling framework reallocates power to mobile units requiring immediate charging, ensuring uninterrupted operation while maintaining sufficient reserves for stationary assets. These adaptive adjustments are encoded within the DRL–based policy network, enabling real-time energy redistribution that aligns with evolving mission demands. The energy allocation model employs a hierarchical prioritization framework to differentiate between essential and nonessential loads. Mission-critical systems, such as ISRU oxygen extraction, habitat life support, and rover mobility, are assigned higher priority weights in the RL reward function, ensuring that they receive uninterrupted power. Lower-priority loads, such as scientific instruments and secondary charging stations, are allocated energy dynamically based on systemwide availability. The RL framework continuously adjusts power distribution using real-time system state data, ensuring adaptive prioritization that responds to dynamic mission conditions. By incorporating this prioritization mechanism, the model optimizes power allocation efficiency while preventing disruptions in critical lunar operations.

The constraints of the model ensure physical feasibility, energy conservation, and system efficiency under uncertain environmental and operational conditions. Given the prolonged exposure to lunar regolith dust accumulation and extreme thermal variations, the proposed framework incorporates periodic recalibration of power transmission parameters. This mechanism continuously monitors efficiency losses and applies adaptive beam realignment strategies to compensate for performance degradation, ensuring sustained WPT efficiency throughout the mission duration. These constraints include power balance equations ensuring that transmitted energy equals received energy plus losses, beam steering limitations based on rover speed and angular deviation, upper and lower power thresholds preventing overloading or underutilization, and dynamic power allocation rules for adaptive scheduling in real-time scenarios. The formulation incorporates nonlinearities in beam misalignment–induced losses, energy latency effects, and optimization-driven power prioritization, ensuring a comprehensive mathematical foundation for intelligent WPT–based lunar power distribution systems. The RL–based scheduling model dynamically adjusts power allocation based on real-time solar availability, rover mobility, and energy demand variations to maintain efficient WPT performance. In scenarios where solar power generation drops significantly due to extended lunar night conditions, regolith dust accumulation on photovoltaic surfaces, or unexpected mission constraints, the model integrates an emergency energy management strategy to prioritize critical loads over secondary energy consumers. The prioritization mechanism operates through a tiered RL approach, where mission-critical systems such as habitat life support, ISRU operations, and primary rover navigation receive higher priority weights in the reward function, ensuring sustained energy supply under extreme energy-deficit conditions. Less essential loads, such as secondary scientific instruments or low-priority charging stations, experience adaptive power reduction to prevent complete system failure. In addition, the model employs a predictive load-shedding mechanism, where RL agents anticipate prolonged power shortages and proactively redistribute available energy based on system-level priorities. This ensures that power outages do not immediately impact essential lunar operations, enhancing the resilience of the WPT network under unpredictable solar energy variations.

$$\max_{\Phi,\Psi,\Xi} \sum_{t=1}^{T} \sum_{\iota\in\mathcal{N}} \left[\frac{\Theta_{\iota,t}^{\mathrm{rx}}}{\Theta_{\iota,t}^{\mathrm{rx}}} \cdot \left(\frac{\sum_{\kappa\in\mathcal{K}} \Upsilon_{\iota,\kappa,t}^{\mathrm{beam}} \cdot \Gamma_{\kappa,t}^{\mathrm{align}} \cdot \Lambda_{\kappa,t}^{\mathrm{mob}}}{\Omega_{\iota,t}^{\mathrm{oss}} + \sum_{\zeta\in Z} \beta_{\zeta,t}^{\mathrm{reg}}} \right) \cdot \exp\left(-\alpha_{\iota,t}^{\mathrm{temp}}\right) \right] - \sum_{\tau=1}^{T} \sum_{\nu\in\mathcal{M}} \left(\sigma_{\nu,\tau}^{\mathrm{idle}} + \varpi_{\nu,\tau}^{\mathrm{latency}} \right). \tag{1}$$

Equation (1) formulates the maximization of total power efficiency in the lunar WPT system. The term $\Theta_{\iota,t}^{\mathrm{rx}}/\ \Theta_{\iota,t}^{\mathrm{rx}}$ represents the efficiency ratio between received and transmitted power. The fraction accounts for beam alignment efficiency $\Gamma_{\kappa,t}^{\mathrm{align}}$, receiver mobility factor $\Lambda_{\kappa,t}^{\mathrm{mob}}$, and energy attenuation from regolith dust $\beta_{\zeta,t}^{\mathrm{reg}}$, which are crucial for dynamic power transmission under extreme lunar environmental conditions. The exponential decay term $\exp\left(-\alpha_{\iota,t}^{\mathrm{temp}}\right)$ captures temperature-induced efficiency loss, while the penalization term $\sigma_{\nu,\tau}^{\mathrm{idle}} + \varpi_{\nu,\tau}^{\mathrm{latency}}$ accounts for power wastage and latency issues in WPT scheduling.

$$\min \sum_{t=1}^{T} \sum_{\iota \in \mathcal{N}} \left[\Psi_{\iota,t}^{\mathrm{deficit}} + \Theta_{\iota,t}^{\mathrm{outage}} + \sum_{\kappa \in \mathcal{K}} \Xi_{\iota,\kappa,t}^{\mathrm{recharge}} \cdot \left(\frac{1}{\lambda_{\kappa,t}^{\mathrm{charging}} + \epsilon_{\kappa,t}^{\mathrm{wait}}} \right) \right]. \tag{2}$$

Equation (2) minimizes energy downtime by ensuring continuous availability for critical lunar systems. The term $\Psi_{\iota,t}^{\mathrm{deficit}}$ represents power shortage events, while $\Theta_{\iota,t}^{\mathrm{outage}}$ models complete failures in wireless power reception. The fraction inside the summation penalizes recharging inefficiencies, where $\lambda_{\kappa,t}^{\mathrm{charging}}$ accounts for charging time per unit energy and $\epsilon_{\kappa,t}^{\mathrm{wait}}$ represents waiting time due to power congestion. This function ensures that energy interruptions for life support and ISRU systems are minimized, improving mission reliability.

$$\min \sum_{t=1}^{T} \sum_{\iota \in \mathcal{N}} \sum_{\kappa \in \mathcal{K}} \left[\frac{\Upsilon_{\iota,\kappa,t}^{\mathrm{beam}} \cdot \Gamma_{\kappa,t}^{\mathrm{align}}}{\Omega_{\iota,t}^{\mathrm{loss}} + \sum_{\zeta \in Z} \beta_{\zeta,t}^{\mathrm{reg}}} \cdot \exp\left(-\xi_{\kappa,t}^{\mathrm{angle}}\right) \right]. \tag{3}$$

Equation (3) minimizes power transmission losses, ensuring efficient energy delivery. The numerator represents beamforming power efficiency $\Upsilon_{\iota,\kappa,t}^{\mathrm{beam}}$ and alignment accuracy $\Gamma_{\kappa,t}^{\mathrm{align}}$, while the denominator accounts for energy absorption losses $\Omega_{\iota,t}^{\mathrm{loss}}$ and regolith interference $\beta_{\zeta,t}^{\mathrm{reg}}$. The exponential decay term $-\xi_{\kappa,t}^{\mathrm{angle}}$ penalizes beam misalignment errors, ensuring precise wireless energy targeting.

$$\max_{\Theta,\Xi} \sum_{t=1}^{T} \sum_{\iota \in \mathcal{N}} \left[\Lambda_{\iota,t}^{\mathrm{priority}} \cdot \frac{\Theta_{\iota,t}^{\mathrm{critical}}}{\sum_{\kappa \in \mathcal{K}} \Xi_{\iota,\kappa,t}^{\mathrm{allocation}} + \epsilon_{\iota,t}^{\mathrm{surplus}}} \right] - \sum_{\tau=1}^{T} \sum_{\nu \in \mathcal{M}} \left(\sigma_{\nu,\tau}^{\mathrm{delay}} + \varpi_{\nu,\tau}^{\mathrm{ineff}} \right). \tag{4}$$

Equation (4) optimizes charging priorities to ensure continuous energy delivery to high-priority lunar systems. The term $\Lambda_{\iota,t}^{\mathrm{priority}}$ acts as a priority weighting factor, ensuring that critical loads (e.g., life support and ISRU oxygen extraction) receive power first. The denominator accounts for the total energy allocation $\Xi_{\iota,\kappa,t}^{\mathrm{allocation}}$, with a penalty $\epsilon_{\iota,t}^{\mathrm{surplus}}$ for overallocation. The last term penalizes delays $\epsilon_{\iota,t}^{\mathrm{surplus}}$ and inefficiencies $\varpi_{\nu,\tau}^{\mathrm{ineff}}$, ensuring optimal power scheduling.

While the existing formulation incorporates real-time solar energy availability, it does not explicitly model seasonal variations in solar flux, which could significantly impact long-term scheduling, particularly under extended lunar night conditions. By integrating a seasonality-aware scheduling strategy, the system can anticipate extended low-energy periods and allocate resources proactively, ensuring continuous power supply for mission-critical systems. The DRL–based WPT model can leverage this factor to refine policy learning, dynamically adjusting power allocation in response to expected seasonal energy fluctuations, further improving resilience in long-duration lunar operations.

$$\sum_{\iota \in \mathcal{N}} \Theta_{\iota,t}^{\mathrm{tx}} = \sum_{\iota \in \mathcal{N}} \left(\Theta_{\iota,t}^{\mathrm{rx}} + \sum_{\kappa \in \mathcal{K}} \left(\Omega_{\iota,\kappa,t}^{\mathrm{beam}} + \sum_{\zeta \in \mathcal{Z}} \beta_{\zeta,t}^{\mathrm{reg}} \right) \right), \quad \forall t \in \mathcal{T}. \tag{5}$$

Equation (5) enforces the total power balance constraint, ensuring that the sum of all transmitted power $\Theta_{\iota,t}^{\mathrm{tx}}$ across all nodes matches the total received power plus all losses from beam attenuation $\Omega_{\iota,\kappa,t}^{\mathrm{beam}}$ and energy absorption due to regolith dust accumulation $\beta_{\zeta,t}^{\mathrm{reg}}$. This prevents energy discrepancies in WPT systems, ensuring an efficient power flow across the lunar grid.

$$S_{\iota,t+1}^{\mathrm{bat}} = S_{\iota,t}^{\mathrm{bat}} + \Theta_{\iota,t}^{\mathrm{rx}} \cdot \eta_{\iota}^{\mathrm{conv}} - \Lambda_{\iota,t}^{\mathrm{dis}} \cdot \eta_{\iota}^{\mathrm{dis}} - \sigma_{\iota,t}^{\mathrm{lcak}}, \quad \forall \iota \in \mathcal{N}, \forall t \in \mathcal{T}. \tag{6}$$

Equation (6) governs the evolution of battery SoC across all mobile and stationary units. The battery charge level $S_{\iota,t}^{\mathrm{bat}}$ is updated dynamically based on the received power $\Theta_{\iota,t}^{\mathrm{rx}}$ (converted at efficiency $\eta_{\iota}^{\mathrm{conv}}$), discharge rate $\Lambda_{\iota,t}^{\mathrm{dis}}$ (with efficiency $\eta_{\iota}^{\mathrm{dis}}$), and self-discharge losses $\sigma_{\iota,t}^{\mathrm{lcak}}$. This constraint ensures that all energy storage operates efficiently while compensating for self-discharge and conversion inefficiencies. To account for extreme thermal variations on the Moon, a temperature-dependent degradation factor is introduced to model the impact of lunar temperature fluctuations on battery performance and energy storage capacity. This factor represents temperature-induced efficiency degradation, which follows an exponential decay trend based on deviations from an optimal reference temperature. It reflects the way prolonged exposure to high temperatures reduces battery efficiency, while extreme cold increases self-discharge rates. By incorporating this factor, the RL–based scheduling model can dynamically adjust energy allocation to compensate for temperature-induced storage losses. This enhancement ensures more accurate energy storage predictions under varying lunar conditions, improving the resilience and reliability of the WPT framework in long-duration missions.

$$\sum_{\kappa\in\mathscr{K}} \Upsilon^{\text{beam}}_{\iota,\kappa,t}\cdot\Gamma^{\text{align}}_{\kappa,t} = 1, \quad \forall\iota\in\mathscr{M},\forall t\in\mathscr{T}. \tag{7}$$

Equation (7) ensures that wireless power beams are correctly aligned with moving receivers by enforcing that the total alignment factor across all beam sources sums to unity. This guarantees that rovers and mobile energy units receive precisely targeted beams, minimizing power leakage and inefficiencies.

$$\sum_{\iota\in\mathscr{N}} \delta^{\text{occupied}}_{\iota,t} \le \delta^{\max}, \quad \forall t\in\mathscr{T}. \tag{8}$$

Equation (8) limits the number of charging units that can be accommodated simultaneously, preventing overcrowding at charging stations. Here, $\delta^{\text{occupied}}_{\iota,t}$ is a binary indicator of whether a charging station is in use, and $\delta^{\max}$ represents the total capacity limit.

$$\Theta^{\text{tx,min}}_{\iota,t} \le \Theta^{\text{tx}}_{\iota,t} \le \Theta^{\text{tx,max}}_{\iota,t}, \quad \forall\iota\in\mathscr{N},\forall t\in\mathscr{T}. \tag{9}$$

Equation (9) enforces upper and lower limits on power transmission, ensuring that energy is neither overloaded nor underutilized. This protects the WPT system from excessive energy losses while maintaining transmission stability. To enhance the robustness of power transmission constraints in equation (9), the model incorporates real-time environmental disturbances, particularly charged lunar dust particles that can interfere with WPT. Lunar regolith is known to become electrostatically charged due to solar wind interactions, creating a dynamic dust environment that affects beam propagation and power reception. To model this effect, an environmental attenuation factor is introduced, which dynamically adjusts transmission efficiency based on real-time dust density and charge distribution. This factor is derived from empirical data on lunar dust behavior and electrostatic charging models, ensuring that WPT scheduling remains adaptive under varying environmental conditions. In addition, the RL–based optimization framework continuously refines transmission parameters by integrating sensor feedback on dust concentration, allowing for real-time compensation strategies such as adaptive power adjustments and beam steering corrections. These enhancements improve system resilience and maintain stable energy delivery despite unpredictable lunar dust disturbances.

$$\Theta^{\text{solar}}_{\iota,t} = \Theta^{\max}_{\iota}\cdot\xi^{\text{sun}}_{t}, \quad \forall\iota\in\mathscr{N},\forall t\in\mathscr{T}. \tag{10}$$

Equation (10) models the variability of solar power availability due to lunar night cycles. The power generation $\Theta^{\text{solar}}_{\iota,t}$ is scaled by the solar availability factor $\Theta^{\text{solar}}_{\iota,t}$, which accounts for sunlight intensity and lunar positioning.

$$\sum_{\iota\in\mathscr{N}} \Theta^{\text{rx}}_{\iota,t} \ge \Theta^{\text{critical}}, \quad \forall t\in\mathscr{T}. \tag{11}$$

Equation (11) ensures an uninterrupted power supply for critical life support systems, enforcing that the total received power never drops below a predefined critical threshold Θ^{critical}.

$$\sum_{\iota\in\mathscr{N}} \Theta^{\text{rx,ISRU}}_{\iota,t} \ge \Theta^{\min}_{\text{ISRU}}, \quad \forall t\in\mathscr{T}. \tag{12}$$

Equation (12) ensures sufficient energy allocation to ISRU facilities, guaranteeing uninterrupted operation of oxygen extraction and material processing units.

$$\tau^{\text{delay}}_{\iota,t} \le \tau^{\max}, \quad \forall\iota\in\mathscr{N},\forall t\in\mathscr{T}. \tag{13}$$

Equation (13) enforces an upper bound on power transmission latency, ensuring that energy delivery is fast enough to maintain system stability.

$$\Theta^{\text{rx}}_{\iota,t} = \Theta^{\text{tx}}_{\iota,t}\cdot\exp\left(-\beta^{\text{reg}}_{\iota,t}\right), \quad \forall\iota\in\mathscr{N},\forall t\in\mathscr{T}. \tag{14}$$

Equation (14) models the impact of lunar dust on power reception, applying an exponential decay function to account for dust accumulation on receivers. The degradation function is based on both empirical data and theoretical modeling. Empirical data from past lunar missions, such as Apollo surface experiments and the Lunar Surveyor program, provide measured insights into dust adhesion, particle density, and optical degradation effects. In addition, theoretical models incorporating electrostatic dust transport mechanisms and surface adhesion physics have been used to extrapolate long-term dust accumulation trends. To validate this degradation model, simulated dust deposition tests have been conducted based on established lunar regolith particle size distributions and electrostatic charging effects. The exponential decay function in equation (14) is parameterized using these empirical references, ensuring that the model accurately reflects progressive power losses observed under varying dust accumulation rates. This combined empirical–theoretical approach enhances model's reliability, allowing the DRL–based WPT scheduling strategy to adapt dynamically by compensating for degradation over time.

$$\Theta^{\text{rx}}_{\iota,t} = \Theta^{\text{tx}}_{\iota,t}\cdot\exp\left(-\alpha^{\text{temp}}_{\iota,t}\right), \quad \forall\iota\in\mathscr{N},\forall t\in\mathscr{T}. \tag{15}$$

Equation (15) captures the efficiency degradation due to extreme lunar temperature variations.

$$\sum_{\iota\in\mathscr{N}} \frac{\Theta^{\text{rx}}_{\iota,t}}{\sum_{\kappa\in\mathscr{K}}\Theta^{\text{rx}}_{\kappa,t}} = \frac{1}{|\mathscr{N}|}, \quad \forall t\in\mathscr{T}. \tag{16}$$

Equation (16) ensures fair power distribution among all receivers within the lunar WPT system. This constraint enforces an equitable allocation of received energy, preventing situations where certain energy receivers (e.g., high-power–demanding ISRU systems) consume a disproportionate share of power, while others (such as low-power robotic agents) are left with inadequate energy. The left-hand term represents the proportion of received power at node ι relative to the total available received energy, ensuring that every node gets an equal share when summed over all receivers. This constraint is crucial in multiagent energy networks to avoid resource monopolization and ensure systemwide resilience.

$$\sum_{\iota\in\mathscr{N}} \Upsilon_{\iota,t}^{\text{beam}} \le \Upsilon^{\max}, \quad \forall t\in\mathscr{T}. \tag{17}$$

Equation (17) enforces bandwidth limitations on wireless energy transmission, ensuring that the total number of simultaneous charging beams does not exceed system capacity. The left-hand term represents the summation over all active WPT beams at time t, while the right-hand term $\Upsilon^{\max}$ defines the upper bound on the number of simultaneous transmissions allowed. This constraint is particularly critical in high-density lunar power grids, where excessive concurrent WPT operations may cause interference, signal degradation, or power inefficiencies due to limited spectral bandwidth.

$$S_{\iota,t}^{\text{bat}} = S_{\iota,t-1}^{\text{bat}} \cdot \left(1 - \lambda_{\iota}^{\text{decay}}\right) + \Theta_{\iota,t}^{\text{rx}}, \quad \forall \iota\in\mathscr{N}, \forall t\in\mathscr{T}. \tag{18}$$

Equation (18) governs the degradation of battery storage units over time due to repeated charge–discharge cycles. The first term models the natural capacity decay of the energy storage device (e.g., lithium-sulfur batteries, solid-state batteries, or other space-rated power units), characterized by an aging factor $\lambda_{\iota}^{\text{decay}}$. The second term accounts for the energy received at time t, which replenishes the storage capacity. Over multiple cycles, this equation ensures that the battery degradation effect is realistically modeled, preventing overoptimistic assumptions about energy retention in lunar power storage systems.

$$\left|\Theta_{\iota,t}^{\text{rx}} - \Theta_{\iota,t-1}^{\text{rx}}\right| \le \Delta^{\max}, \quad \forall \iota\in\mathscr{N}, \forall t\in\mathscr{T}. \tag{19}$$

Equation (19) prevents drastic power fluctuations in energy allocation, ensuring stable WPT performance over time. The absolute difference term quantifies the change in received power between consecutive time steps, and the upper bound $\Delta^{\max}$ constrains the maximum allowable change. This constraint is crucial for mission-critical lunar operations, as sudden shifts in the power supply can lead to electrical failures, inefficient charging cycles, or unintended system shutdowns.

$$\sum_{t=1}^{T} \sum_{\iota\in\mathscr{N}} \Theta_{\iota,t}^{\text{interrupted}} \cdot \pi_{\iota,t}^{\text{penalty}} \le \Pi^{\max}. \tag{20}$$

Equation (20) imposes penalties on charging interruptions, ensuring that abrupt energy disconnections remain minimal. The left-hand side represents a cumulative penalty function, where every instance of interrupted charging (denoted by $\Theta_{\iota,t}^{\text{interrupted}}$) is assigned a severity weight $\pi_{\iota,t}^{\text{penalty}}$. The right-hand term $\Pi^{\max}$ limits the overall disconnection impact. This constraint is critical for maintaining a continuous energy supply to life support systems and essential lunar infrastructure, preventing power failures due to unstable WPT scheduling.

$$S_{\iota,t}^{\text{bat}} \ge S^{\min}, \quad \forall \iota\in\mathscr{M}, \forall t\in\mathscr{T}. \tag{21}$$

Equation (21) guarantees that lunar rovers always retain a minimum energy threshold for mobility. This constraint prevents scenarios where a rover completely depletes its battery and becomes stranded on the lunar surface, unable to return to a charging station.

$$\Theta_{\iota,t}^{\text{rx}} = \Theta_{\iota,t}^{\text{tx}} \cdot \exp\left(-\xi_{\iota,t}^{\text{mob}}\right), \quad \forall \iota\in\mathscr{M}, \forall t\in\mathscr{T}. \tag{22}$$

Equation (22) accounts for receiver mobility effects, ensuring that the power received is adjusted based on movement patterns. The exponential decay function models energy attenuation as a function of displacement speed $\xi_{\iota,t}^{\text{mob}}$. To enhance the adaptability of the model, an additional adjustment mechanism is incorporated to dynamically update beam tracking in response to unexpected disruptions in rover trajectories. The model integrates a predictive motion compensation approach that estimates short-term rover trajectory deviations using historical mobility patterns and real-time sensor data. This allows the RL–based WPT scheduling system to anticipate abrupt movement variations and preemptively adjust beam alignment parameters. In addition, terrain-aware constraints are introduced to account for environmental factors such as steep inclines, regolith density, and surface irregularities, which influence rover speed and maneuverability. By incorporating these adaptive mechanisms, the model improves resilience against sudden trajectory changes, ensuring stable and efficient power transmission even under unpredictable mobility conditions. This enhancement strengthens the robustness of the RL–based WPT framework, making it more applicable to real-world lunar operations.

$$\Upsilon_{\iota,t}^{\text{beam}} \le \Upsilon^{\text{safe}}, \quad \forall \iota\in\mathscr{N}, \forall t\in\mathscr{T}. \tag{23}$$

Equation (23) ensures that beam intensity does not exceed safety thresholds, protecting both human operators and electronic equipment.

$$\tau_{\iota,t}^{\text{delay}} = \frac{d_{\iota,t}^{\text{dist}}}{v_{\text{light}}}, \quad \forall \iota\in\mathscr{N}, \forall t\in\mathscr{T}. \tag{24}$$

Equation (24) models latency in energy transmission for WPT systems operating over long distances. Since lunar WPT relies on high-frequency electromagnetic waves (such as microwave or laser beaming), the time delay $\tau_{\iota,t}^{\text{delay}}$ in power delivery is a function of the transmission distance $d_{\iota,t}^{\text{dist}}$ and the speed of light v_{light}. This constraint is critical for real-time power delivery scheduling, ensuring that remote receivers account for energy arrival delays before making power allocation decisions. If latency becomes too high, power misalignment may occur, leading to inefficiencies in beamforming, increased energy losses, and potential receiver overheating due to unintended overcharging.

$$\sum_{\iota\in\mathscr{N}} \Theta_{\iota,t}^{\text{rx}} = \sum_{\kappa\in\mathscr{K}} \Theta_{\kappa,t}^{\text{tx}}, \quad \forall t\in\mathscr{T}. \tag{25}$$

Equation (25) enforces dynamic energy load balancing across multiple MEVPP nodes, ensuring that the total received power matches the total transmitted power. In a lunar WPT environment, power transmission must be dynamically allocated based on real-time demand fluctuations across different subsystems, including life support, ISRU processing, scientific instruments, and robotic exploration units. This constraint is particularly important in decentralized, distributed WPT networks, where energy sources (such as solar farms or nuclear batteries) must redistribute power equitably among competing loads. Failure to properly balance the load could lead to excessive power allocation to less critical units while more essential functions suffer power shortages.

$$\Theta_{\iota,t}^{\text{surplus}} = \max\left(\Theta_{\iota,t}^{\text{rx}} - \Theta_{\iota,t}^{\text{demand}}, 0\right), \quad \forall \iota \in \mathcal{N}, \forall t \in \mathcal{T}. \tag{26}$$

Equation (26) manages excess power redistribution, ensuring that any received power exceeding local demand is redirected to storage or secondary loads. The function inside the maximum operator ensures that only positive energy surpluses are considered, preventing negative power allocations. This constraint is crucial in lunar WPT systems because solar farms generate highly intermittent power outputs, leading to periodic energy surpluses that must be properly managed. Uncontrolled surplus power could lead to overheating, equipment degradation, and excessive discharge cycles on energy storage units, significantly reducing the overall lifespan of lunar microgrid infrastructure. The model now supports real-time redistribution of surplus energy to improve WPT efficiency. Instead of solely storing excess power, surplus energy is dynamically allocated to receivers based on priority, real-time demand, and battery SoC. A priority-based surplus allocation function has been introduced to ensure that mission-critical systems receive additional energy when available. Furthermore, the RL framework continuously updates allocation decisions based on systemwide power availability, optimizing distribution efficiency.

$$\sum_{\iota\in\mathcal{N}} \Theta_{\iota,t}^{\text{secure}} \geq \Theta^{\text{min}-\text{secure}}, \quad \forall t \in \mathcal{T}. \tag{27}$$

Equation (27) ensures that power signals maintain cybersecurity integrity, preventing unauthorized or malicious energy redirection due to cyber-physical attacks. In a wireless energy system on the Moon, adversarial attacks could include signal spoofing, interference jamming, and power hijacking, where rogue receivers manipulate the WPT network to divert energy away from mission-critical systems. This constraint enforces that the minimum fraction of energy transmitted remains protected by encrypted control channels, ensuring power delivery is authenticated, traceable, and resilient to cyber threats. This is particularly crucial for multiuser lunar energy-sharing models, where power is transmitted to multiple independent scientific or industrial operations.

$$\Theta_{\iota,t}^{\text{rx}} = \Theta_{\iota,t}^{\text{tx}} \cdot \exp\left(-\alpha_{\iota,t}^{\text{temp}}\right), \quad \forall \iota \in \mathcal{N}, \forall t \in \mathcal{T}. \tag{28}$$

Equation (28) models the impact of extreme lunar temperature variations on power transmission efficiency. On the Moon, surface temperatures can fluctuate between −180°C during lunar nights to over 120°C under direct sunlight, significantly affecting semiconductor-based rectennas and photovoltaic receivers. The exponential decay term $-\alpha_{\iota,t}^{\text{temp}}$ accounts for the temperature-induced degradation of energy absorption efficiency, ensuring that power allocations adapt dynamically to environmental conditions. Without this constraint, receivers could overheat or underperform, leading to permanent damage or failure in lunar energy subsystems.

$$\Lambda_{\iota,t}^{\text{priority}} = \frac{\Theta_{\iota,t}^{\text{critical}}}{\sum_{\kappa\in\mathcal{K}} \Theta_{\kappa,t}^{\text{rx}}}, \quad \forall \iota \in \mathcal{N}, \forall t \in \mathcal{T}. \tag{29}$$

Equation (29) enforces mission adaptability by dynamically prioritizing power allocations to critical systems. The fraction represents a real-time priority scaling factor, where each unit's energy share is weighted by its criticality level. This ensures that life support systems, astronaut habitats, and safety mechanisms receive guaranteed energy allocations before nonessential research instruments or backup storage units. This constraint is essential for adaptive resource management, ensuring that lunar energy operations remain resilient to unexpected mission changes, disasters, or reconfigurations.

$$\lim_{t\longrightarrow\infty} \sum_{\iota\in\mathcal{N}} \left|\Theta_{\iota,t}^{\text{opt}} - \Theta_{\iota,t-1}^{\text{opt}}\right|. \tag{30}$$

Equation (30) serves as the final convergence condition for the overall WPT optimization algorithm, ensuring that power allocations reach a stable, steady-state solution over time. The summation quantifies the variation in optimized power transmission levels, and the limiting behavior guarantees that as time progresses, fluctuations vanish. This is particularly important for real-time RL–based scheduling models, ensuring that optimization processes do not oscillate indefinitely or converge to suboptimal solutions.

4. Methodology

To solve the complex, nonlinear, and dynamic optimization problem formulated in the previous section, this study leverages DRL with PPO for adaptive beam steering and WPT scheduling. Unlike traditional rule-based WPT control mechanisms, RL enables the adaptive optimization of power allocation and beam positioning based on real-time state observations, allowing the system to self-learn and optimize power dispatch strategies under varying environmental and operational conditions. This methodology integrates a RL framework with MDP modeling, ensuring that the agent can continuously learn optimal power distribution strategies based on receiver mobility, energy demands, and real-time solar power fluctuations.

The proposed learning model represents the WPT energy scheduling problem as a MDP, where the state space consists of rover positions, battery SoC levels, charging priorities, and beam alignment conditions. The action space includes power allocation decisions, beam steering adjustments, and priority-based power scheduling updates. The reward function is carefully designed to maximize overall WPT efficiency, minimize energy downtime, and penalize unnecessary idle time or misalignment-induced losses. The RL agent utilizes a policy gradient–based optimization approach with PPO, ensuring that the model converges rapidly while maintaining exploration–exploitation balance. A key enhancement of the proposed framework is its ability to dynamically reallocate power during emergency conditions. When a communication failure or loss of sensor data occurs, the DRL model immediately shifts energy resources toward local autonomy, ensuring that rovers and stationary units can operate independently until normal operations resume. In addition, in scenarios requiring rapid rover redeployment, the model learns to prioritize power delivery to high-mobility receivers while adjusting static unit power budgets to prevent disruptions in habitat support and ISRU operations. This adaptive response capability significantly enhances the framework's resilience in unpredictable lunar mission environments. To further enhance decision-making stability, the framework integrates a predictive analytics layer that utilizes historical mission data to refine power scheduling strategies. By analyzing past rover mobility patterns, energy consumption trends, and environmental variations, the model adjusts its policy updates to incorporate anticipated future demands. This allows the system to proactively allocate energy resources, reducing the likelihood of sudden shortages or excessive allocations. The incorporation of historical insights enables the RL model to balance real-time adaptability with long-term optimization, significantly improving system efficiency and reliability. The policy network continuously refines decision-making strategies, adjusting beam intensity and power scheduling in response to real-time environmental changes. The policy network continuously refines decision-making strategies, adjusting beam intensity and power scheduling in response to real-time environmental changes. Given the intermittent connectivity and potential signal delays in lunar environments, the proposed framework incorporates fail-safe mechanisms to maintain stable power delivery during communication disruptions. Specifically, each mobile receiver is equipped with a local predictive model trained using historical mission data and on-site observations to estimate power requirements in the event of temporary communication loss. This allows the receiver to autonomously adjust beam alignment and energy scheduling based on its last known state. In addition, the transmitter utilizes an adaptive scheduling buffer, where power transmission decisions are precomputed based on predicted rover trajectories and energy demand trends. This ensures that even during short-term signal outages, energy delivery continues without major interruptions. Furthermore, a hierarchical decision-making approach is employed, where high-priority receivers (such as habitats and ISRU units) are given redundant transmission paths through relay-based WPT stations, ensuring reliable power allocation even under extreme conditions. These fail-safe mechanisms enhance the system's resilience to sudden communication failures, ensuring continued energy availability for mission-critical operations while maintaining overall power efficiency. For real-world deployment, the DRL model must operate within the computational constraints of space-grade embedded hardware. To address this, the proposed framework employs a hybrid on-device and ground-assisted learning approach, where the training phase is conducted offline using high-performance computing clusters, while the trained model is compressed and optimized for onboard execution. Model reduction techniques such as quantization, pruning, and knowledge distillation are applied to minimize memory footprint and computational overhead, ensuring feasibility for low-power, radiation-hardened processors used in space missions. In addition, the framework leverages edge AI's inference techniques, where policy updates are efficiently executed on embedded processors without requiring full-scale deep learning model retraining. This allows the DRL–based energy scheduling system to dynamically adjust power allocation in real time while minimizing computational latency. The framework integrates an adaptive recalibration mechanism that dynamically updates WPT parameters based on real-time sensor feedback. By periodically assessing power transmission efficiency and environmental disruptions, the system proactively mitigates degradation due to dust accumulation and thermal variations, ensuring stable and reliable energy delivery.

To address the potential electromagnetic interference (EMI) risks associated with high-power WPT, the proposed framework incorporates multiple mitigation techniques to ensure electromagnetic compatibility (EMC) in lunar energy systems. First, frequency modulation (FM) and frequency hopping techniques are implemented to dynamically adjust transmission frequency, ensuring minimal interference from nearby communication and sensor networks. By actively shifting operating frequencies, the system prevents prolonged exposure within any single frequency band, reducing EMI persistence and cross-system disturbances. Second, adaptive beamforming is employed to optimize phase control in the transmitting array, ensuring precise directional energy transmission while minimizing unintended radiation spillover. This technique significantly reduces EMI leakage to nontargeted zones, making the system more suitable for operation in lunar environments where sensitive scientific instruments and habitat electronics must be protected from electromagnetic disturbances. Third, electromagnetic shielding and antenna pattern optimization are integrated into the system. High-conductivity shielding materials are applied around the transmitting and receiving units to mitigate electromagnetic leakage. Moreover, low-sidelobe antenna designs are employed to further reduce unintended emissions, ensuring that most of the transmitted energy is confined within the desired beam path. Lastly, power density constraints are introduced within the WPT optimization framework to ensure compliance with internationally recognized EMI safety standards, such as IEEE

C95.1 and ICNIRP guidelines. These constraints prevent excessive electromagnetic field strength in human-occupied zones and high-sensitivity scientific areas, enhancing the safety and reliability of the proposed WPT system.

To further enhance training efficiency and computational scalability, this study employs MARL, where multiple WPT transmitters and receivers act as independent agents, collaboratively learning optimal power allocation and beamforming strategies. The training process is conducted over a simulated lunar environment, integrating realistic energy consumption models, mobility constraints, and power degradation effects due to lunar terrain interference. The optimization process follows a two-stage RL pipeline, where the first stage focuses on pretraining the model with the historical WPT data, while the second stage incorporates real-time adjustments using live mission telemetry. This hybrid approach ensures that the system achieves both long-term learning stability and real-time adaptability, making it a robust solution for autonomous WPT–based lunar energy distribution systems.

$$\mathscr{M} = \langle \mathscr{S}, \mathscr{A}, P, R, \gamma \rangle. \tag{31}$$

Equation (31) defines the MDP formulation for WPT scheduling in lunar energy networks. The MDP framework is a fundamental mathematical representation used in RL to model sequential decision-making problems, where the goal is to optimize long-term rewards. The tuple $\mathscr{M}$ consists of the following:

- $\mathscr{S}$: State space: It represents the system's current state, including energy levels of receivers, rover positions, charging station availability, and environmental factors such as lunar dust accumulation and temperature fluctuations.
- $\mathscr{A}$: Action space: It defines available decisions, including beam steering, power allocation, charging prioritization, and scheduling strategies.
- P: Transition probability model: It describes the dynamics of the system, governing how the environment evolves after taking an action.
- R: Reward function: It encodes the optimization objective, typically maximizing power efficiency while minimizing energy deficits and unnecessary charging cycles.
- γ: Discount factor: It controls the importance of future rewards, ensuring that the agent optimizes energy scheduling not just for the immediate step but over the long term.

$$\mathscr{S} = \left\{ S_{\iota,t}^{\text{bat}}, P_{\iota,t}^{\text{load}}, \lambda_{\iota,t}^{\text{charge}}, d_{\iota,t}^{\text{pos}}, \Theta_{\iota,t}^{\text{solar}}, \beta_{\iota,t}^{\text{dust}} \right\}. \tag{32}$$

Equation (32) explicitly defines the state space $\mathscr{S}$ of the RL agent, which contains all relevant system parameters needed for optimal decision-making. Each state includes the following:

- $S_{\iota,t}^{\text{bat}}$: Battery SoC of receiver unit ι at time t, which determines the need for energy replenishment.
- $P_{\iota,t}^{\text{load}}$: Current power demand of the load at the unit ι, reflecting real-time energy consumption.
- $\lambda_{\iota,t}^{\text{charge}}$: Charging station occupancy indicator, ensuring efficient scheduling to avoid congestion.
- $d_{\iota,t}^{\text{pos}}$: Spatial position of mobile receivers (rovers and ISRU facilities) crucial for beam alignment and efficient power transmission.
- $\Theta_{\iota,t}^{\text{solar}}$: Solar energy availability, which impacts the overall power generation capacity.
- $\beta_{\iota,t}^{\text{dust}}$: Regolith dust interference level, affecting the efficiency of power reception.

$$\mathscr{A} = \left\{ \Theta_{\iota,t}^{\text{alloc}}, \Gamma_{\iota,t}^{\text{beam}}, \lambda_{\iota,t}^{\text{prior}} \right\}. \tag{33}$$

Equation (33) defines the action space $\mathscr{A}$ for the DRL agent, specifying available decisions for optimizing WPT operations.

- $\Theta_{\iota,t}^{\text{alloc}}$: Power allocation decision, determining how much energy is assigned to each receiver at time \(t\).
- $\Gamma_{\iota,t}^{\text{beam}}$: Beam steering parameters, ensuring that transmitted power aligns optimally with moving receivers.
- $\lambda_{\iota,t}^{\text{prior}}$: Charging priority index, assigning priority levels to different loads based on criticality.

$$P\left(s' \middle| s, a\right) = \Pr\left(S_{\iota,t+1}^{\text{bat}} \middle| S_{\iota,t}^{\text{bat}}, \Theta_{\iota,t}^{\text{alloc}}, d_{\iota,t}^{\text{pos}} \right). \tag{34}$$

Equation (34) models the transition probability function, which governs how the system evolves from one state s to the next state s' after taking action a. This accounts for the following:

- Battery state evolution, where power allocation $\Theta_{\iota,t}^{\text{alloc}}$ increases SoC.
- Spatial displacement, as moving receivers experience position-dependent energy reception changes.

$$R(s,a) = \sum_{\iota \in \mathscr{N}} \left[\frac{\Theta_{\iota,t}^{\text{rx}}}{\Theta_{\iota,t}^{\text{tx}}} - \Psi_{\iota,t}^{\text{deficit}} - \Lambda_{\iota,t}^{\text{idle}} \right]. \tag{35}$$

Equation (35) defines the reward function, which maximizes power efficiency (first term) while penalizing energy deficits and idle charging states.

$$P(s,a) = \sum_{\iota \in \mathscr{N}} \left[\varpi_{\iota,t}^{\text{over}} + \sigma_{\iota,t}^{\text{idle}} \right]. \tag{36}$$

Equation (36) introduces penalties for the following:

- Energy overuse $\varpi_{\iota,t}^{\text{over}}$, ensuring efficient WPT scheduling.

- Unnecessary idle states $\sigma_{\iota,t}^{\text{idle}}$, preventing wasted transmission power.

$$\mathscr{L}^{\text{PPO}} = \mathbb{E}_t\left[\min\left(r_t(\theta)A_t,\ \text{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)A_t\right)\right]. \tag{37}$$

Equation (37) is the PPO loss function, stabilizing RL training by preventing large policy updates.

$$\theta \longleftarrow \theta - \alpha\nabla_\theta \mathscr{L}^{\text{PPO}}. \tag{38}$$

Equation (38) updates the actor network using stochastic gradient descent (SGD).

$$V^\pi(s) = \mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t R\left(s_t, a_t\right)\right]. \tag{39}$$

Equation (39) estimates the long-term value function.

$$\nabla_\theta J(\theta) = \mathbb{E}\left[\nabla_\theta \log \pi_\theta(a|s)A^\pi(s,a)\right]. \tag{40}$$

Equation (40) optimizes policy updates using gradient ascent.

$$\pi^*(a|s) = \arg\max_\pi\left[(1-\epsilon)\mathbb{E}\left[R(s,a)|s,\pi\right] + \epsilon\cdot\mathbb{U}(a)\right]. \tag{41}$$

Equation (41) formulates the exploration–exploitation trade-off model, a crucial component in RL–based energy scheduling for lunar WPT. The term $(1-\epsilon)$ represents the probability of following the current best-known policy, while ϵ is the probability of exploring new, potentially better policies by sampling from an uncertainty function $\mathbb{U}(a)$. This mechanism ensures a balance between leveraging existing knowledge to maximize efficiency and discovering new energy allocation strategies that might further optimize power delivery, minimize transmission losses, or improve charging station utilization. This trade-off is critical in dynamic lunar environments, where unpredictable factors such as lunar dust interference, solar variability, and rover movements require continuous adaptation.

$$\alpha_{t+1} = \alpha_t\cdot\frac{\mathbb{I}\left(\left|\delta_t\right| < \delta_{\text{thres}}\right)}{1+\lambda_{\text{decay}}\cdot t}. \tag{42}$$

Equation (42) models an adaptive learning rate α_t for PPO training stability, dynamically adjusting the step size based on model convergence behavior. The numerator contains an indicator function $(|\delta_t| < \delta_{\text{thres}})$, which ensures that the learning rate remains stable if the gradient change δ_t is within a predefined threshold δ_{thres}. The denominator introduces a time-decaying adjustment factor, ensuring that as training progresses, the step size gradually reduces to prevent divergence. This mechanism is vital in RL for WPT because fixed learning rates may cause unstable oscillations or slow convergence, especially in large-scale multiagent environments with nonstationary energy demand patterns.

$$\pi_\iota(a|s) = \arg\max_\pi \sum_{\kappa\in\mathscr{K}}\omega_{\iota,\kappa}\cdot Q_\kappa(s,a) + \nu_{\iota,\kappa}^{\text{sync}}. \tag{43}$$

Equation (43) introduces MARL for distributed energy scheduling, where multiple power nodes (charging stations, energy transmitters, and mobile receivers) coordinate their actions collaboratively to optimize energy distribution. Each agent ι learns an independent policy π_ι while considering the shared Q-values of neighboring nodes κ. The term $\omega_{\iota,\kappa}$ represents the weight of influence that node κ has on ι, ensuring decentralized but cooperative energy optimization. The term $\nu_{\iota,\kappa}^{\text{sync}}$ acts as a synchronization penalty, preventing drastic discrepancies in power allocation between agents. This approach is essential for scalable WPT in lunar environments, where multiple nodes must dynamically adjust energy distribution while ensuring grid stability.

$$\mathscr{O}\left(\text{PPO}_{\text{train}}\right) = \mathscr{O}\left(N_{\text{epoch}}\cdot N_{\text{batch}}\cdot\mathbb{E}\left[|\mathscr{S}|\right]\right). \tag{44}$$

Equation (44) defines the computational complexity of PPO–based energy scheduling, where learning efficiency and scalability are analyzed. The Big-O notation expresses the complexity in terms of

- N_{epoch}: The number of training epochs required for convergence.
- N_{batch}: The number of data samples processed per optimization step.
- $\mathbb{E}[|\mathscr{S}|]$: The expected size of the state space, determining the dimensionality of the RL problem.

This analysis is essential because large-scale energy optimization problems (such as WPT scheduling for an entire lunar base) can become computationally prohibitive. By quantifying the efficiency of PPO training, researchers can optimize hyperparameters to reduce unnecessary computational overhead while maintaining policy effectiveness.

$$\lim_{t\longrightarrow\infty}\sum_{i=1}^{N}\left|\pi_{\theta_t}\left(a_i\middle|s_i\right) - \pi_{\theta_{t-1}}\left(a_i\middle|s_i\right)\right| = 0. \tag{45}$$

Equation (45) formulates the convergence condition for PPO model training, ensuring that as training progresses over time, the policy updates diminish, indicating that the model has reached a stable optimal policy. This condition is crucial for RL in energy scheduling, as continuous policy oscillations would destabilize power allocations, degrade efficiency, and make scheduling unreliable for mission-critical operations. The summation term represents the total variation between consecutive policy updates, and the limiting behavior ensures that this variation tends toward zero over infinite iterations.

$$\mathbb{E}\left[\left|\Theta_{i,t}^{\text{alloc}} - \Theta_{i,t}^{\text{optimal}}\right|\right] \le \epsilon^{\text{robust}}, \quad \forall t\in\mathscr{N}, \forall t\in\mathscr{T}. \tag{46}$$

Equation (46) evaluates robustness under energy demand fluctuations, ensuring that even if power demand varies, the learned policy maintains near-optimal performance. The expectation function quantifies the expected deviation between the allocated power and the optimal power level, ensuring that the discrepancy does not exceed a robustness threshold ϵ^{robust}. This constraint is critical for real-world deployment of in power grids, as it guarantees that the system remains resilient to unexpected variations in energy demand, solar availability, and communication delays.

$$\mathbb{E}\left[R^{\text{test}}(s,a)\right] \approx \mathbb{E}\left[R^{\text{train}}(s,a)\right] \pm \delta_{\text{gen}}. \tag{47}$$

Equation (47) assesses generalization performance across different lunar missions, ensuring that the trained RL model performs consistently across varied scenarios. The expected reward in test environments should closely approximate the reward obtained in training environments, with an allowable deviation margin δ_{gen}. This guarantees that the WPT optimization framework remains reliable when deployed in new lunar regions, varying terrain conditions, or with different power infrastructures.

5. Case Studies

To evaluate the performance of the proposed RL–based adaptive WPT optimization framework, a high-fidelity simulation of a lunar MEVPP was conducted. The case study focuses on a 30-day continuous lunar mission near Shackleton Crater (89.9°S, 0.0°E), a region of interest due to its permanent shadow zones and fluctuating solar power availability. To further assess the adaptability of the proposed model across varying lunar terrains, additional considerations are made for equatorial regions where solar exposure patterns differ significantly. Unlike polar sites with prolonged shadow zones, equatorial locations experience alternating periods of full illumination and extended darkness, leading to more dynamic energy availability. The RL–based scheduling framework is designed to adjust power allocation in response to real-time solar input variations, ensuring applicability in environments with fluctuating solar flux. By leveraging predictive solar exposure models, the system can optimize WPT scheduling by preemptively dispatching energy during high-insolation periods and strategically utilizing stored power during extended night phases. To improve long-term WPT performance, the model incorporates a degradation-aware optimization strategy. Over time, cumulative losses in WPT hardware, caused by thermal cycling, material fatigue, and regolith-induced wear, gradually reduce transmission efficiency. A degradation-aware reward function enables RL to anticipate and compensate for these effects by dynamically adjusting beam intensity, recalibrating power allocation, and prioritizing maintenance when necessary. In addition, real-time sensor data on system degradation is continuously integrated into the learning framework, ensuring adaptive scheduling adjustments to mitigate performance declines. This enhancement ensures that the model remains robust and effective in long-duration lunar operations, improving the sustainability of WPT deployment over extended missions.

In addition, terrain variations at equatorial sites introduce new challenges for rover mobility and beam tracking. The adaptive motion compensation mechanism incorporated in the model, which was designed to handle Shackleton Crater's rugged topography, remains applicable in equatorial conditions by dynamically adjusting beam alignment in response to shifting environmental constraints. These considerations demonstrate that the proposed framework is not limited to polar regions but can be extended to diverse lunar terrains, ensuring reliable power distribution under varying solar and mobility conditions.

The study considers a 10 × 10 km operational zone, where multiple energy receivers, including four autonomous rovers, two ISRU extraction units, and a primary lunar habitat, require continuous and adaptive energy allocation. The primary WPT transmission station, modeled as a 100 kW high-efficiency microwave beaming system operating at 2.45 GHz, is capable of transmitting power to multiple receivers simultaneously, with a maximum transmission range of 12 km and an efficiency rate of 85% under optimal beam alignment conditions. The energy demand profile is dynamically generated based on realistic lunar mission scenarios. The four rovers, each with a 20 kWh battery capacity, have varying energy consumption rates depending on their assigned tasks, with average power usage ranging from 2 kW during standby mode to 6.5 kW during excavation and mapping operations. The two ISRU extraction units, responsible for oxygen and water ice processing from lunar regolith, operate at a fixed load of 15 kW each, with intermittent peak demands reaching 18 kW during active refinement cycles. The lunar habitat module, which supports astronaut life support and scientific equipment, has a baseline power consumption of 30 kW, with fluctuations of ±10% depending on habitat occupancy and operational conditions. These energy demands present a highly dynamic and uncertain environment, making it an ideal testbed for RL–based optimization.

The simulation is conducted using a Python-based RL framework, integrating Stable-Baselines3 for PPO training, OpenAI Gym for MDP–based state-action formulation, and TensorFlow for deep neural network optimization. The training and evaluation phases are performed on a high-performance computing cluster equipped with Intel Xeon 32-core processors (2.9 GHz), 256 GB RAM, and NVIDIA A100 Tensor Core GPUs, allowing for parallel training of RL agents. While these high-fidelity simulations provide a controlled environment for evaluating the WPT framework, real-world lunar deployment introduces additional challenges, including hardware constraints, communication latencies, and mission uncertainties. To enhance the generalization capability of the proposed framework, future work will integrate hardware-in-the-loop (HIL) simulations to assess real-time performance under actual system latencies and hardware limitations. In addition, incorporating field data from past lunar missions and terrestrial analog environments will further validate the robustness of DRL–based energy scheduling under real-world conditions. By adapting the framework to varying computation capacities, sensor noise, and dynamic mission scenarios, we aim to improve its practical feasibility for autonomous lunar power management.

The simulation runs for 5000 episodes, each representing a 24-hour operation cycle, ensuring sufficient training for policy convergence. The PPO model is configured with a discount factor (γ\gamma) of 0.99, an adaptive learning rate ranging from $1 \times 10 - 41$\times 10^{-4} to $5 \times 10 - 55$\times 10^{-5}, and a batch size of 4096 experience samples per update step. This computational setup ensures that the RL agent achieves optimal decision-making under real-time constraints, learning to maximize power efficiency while minimizing energy deficits and beam misalignment losses.

Figure 1 provides a highly detailed analysis of the solar exposure across a 10 × 10 km region surrounding the Shackleton Crater. The left side represents the average sun visibility over a given period, with a color scale from dark blue (low visibility, near 0) to yellow (high visibility, close to 1). The central dark blue region represents the permanently shadowed interior of the Shackleton Crater, where solar illumination is nearly nonexistent, making it one of the prime candidates for long-term ice preservation and ISRU. The surrounding regions exhibit varying degrees of sunlight exposure, with some areas receiving up to 90% visibility, suggesting optimal locations for solar panel installations and surface-based energy harvesting systems. The yellow-highlighted contour zones indicate terrain areas that receive moderate sunlight exposure, potentially suitable for deploying power relay stations or energy storage hubs. The right-side visualization presents a 3D perspective of the Shackleton Crater, emphasizing the extreme depth and sharp elevation changes within the crater. The color-coded elevation layers highlight how the terrain structure influences the solar exposure, with the deepest parts of the crater remaining entirely in shadow, while the upper rims and nearby ridges benefit from prolonged solar exposure. Given that Shackleton Crater is approximately 21 km in diameter and up to 4 km deep, the elevation gradients pose significant challenges for energy transmission, necessitating adaptive WPT strategies. The sloped terrain further complicates rover mobility and infrastructure deployment, requiring specialized path-planning algorithms to ensure safe navigation between high-exposure zones and shadowed regions where ice deposits are likely to exist.

This visualization illustrates the real-time beamforming strategy of a WPT system operating in a 10 × 10 km lunar zone, showing the power distribution from a centralized WPT transmitter to 15 receivers, including rovers, ISRU processing units, and habitat modules, shown in Figure 2. The figure highlights the spatial relationships, alignment efficiency, and adaptive tracking capabilities of the WPT system, which dynamically directs energy beams based on receiver movement, energy demand, and terrain constraints. The figure provides insight into the geometric distribution and optimization of energy transfer. The WPT transmitter, positioned at (0, 0, 3) km, enables wide-area coverage to support receivers scattered up to 5 km away. The 15 energy receivers are placed at various elevations, simulating realistic lunar surface irregularities. The beamforming vectors (green arrows) depict real-time adaptive power allocation, with longer arrows representing receivers requiring higher precision targeting due to their distance or movement. The dense clustering of receivers in certain areas, particularly within the 2-3 km radius, suggests regions of high-energy demand, likely corresponding to operational hubs where ISRU processing and life support functions are concentrated. The varying beam orientations and distances emphasize the need for continuous power tracking algorithms, ensuring optimal alignment and minimizing energy transmission losses.

Figure 3 represents the hourly energy consumption patterns of a lunar habitat module over a 30-day mission cycle, showing variations in power demand throughout different times of the day. The color-coded heatmap visually captures high-demand and low-demand periods, where red and yellow shades indicate peak energy usage and blue shades represent lower consumption hours. The habitat requires continuous power supply, making it essential to understand how demand changes over time to optimize WPT scheduling and energy storage management. The demand profile in Figure 3 is generated using a synthetic model that incorporates operational constraints, equipment power ratings, and expected astronaut activity cycles based on lunar habitat studies. The synthetic data are formulated by combining power consumption estimates from past analog habitat experiments, NASA mission reports, and energy modeling frameworks for extraterrestrial environments. The variability in demand accounts for life support operations, research activities, and environmental control systems, ensuring that the model reflects realistic mission conditions. To validate the generalizability of the demand profile, sensitivity analyses were conducted by varying energy consumption levels and operational schedules. The results demonstrate that the RL–based WPT scheduling approach remains robust under different energy demand scenarios, confirming the adaptability of the proposed model for lunar habitat power management. The energy demand profile presented in Figure 3 is derived from a synthetic model incorporating expected astronaut activity cycles and operational schedules of critical habitat systems. This model is informed by power consumption data from past analog habitat experiments, NASA mission reports, and lunar habitat energy modeling studies. Variations in energy demand reflect essential functions such as life support operations, research activities, thermal control, and communication systems. The synthetic demand model also incorporates scheduled maintenance periods and low-activity phases, ensuring that the energy trends align with expected mission scenarios. Sensitivity analyses were conducted to verify the robustness of the model across varying habitat occupancy levels and equipment utilization rates, confirming its applicability for lunar mission planning.

The figure reveals consistent high-energy demand periods between 10:00–14:00 and 19:00–22:00, coinciding with likely mission-critical operations, astronaut activities, or system recalibration processes. Demand fluctuates between 25 and 40 kW, with occasional surges reaching above 45 kW, which could be attributed to life support system adjustments, research activities, or heating requirements in extreme lunar temperatures. The lowest power demand occurs between 2:00 and 7:00, where consumption drops to 15–20 kW, likely reflecting reduced activity phases or energy-saving protocols during lunar nighttime. These fluctuations emphasize the need for adaptive power

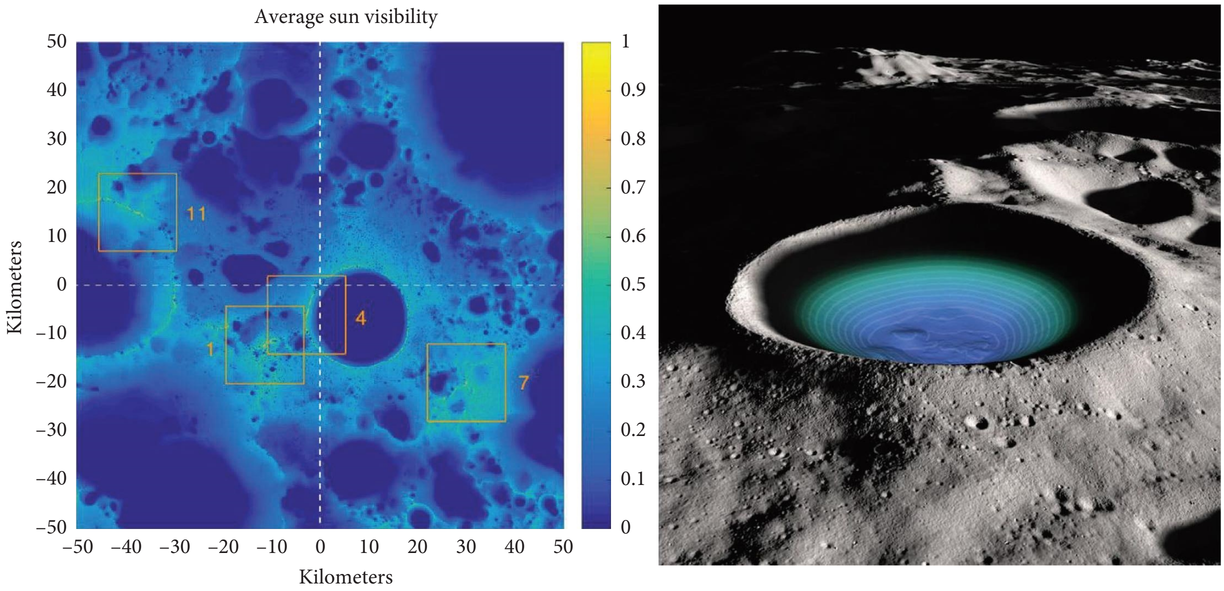

Figure 1: Sun visibility and topographic analysis of the Shackleton crater on the lunar south pole.

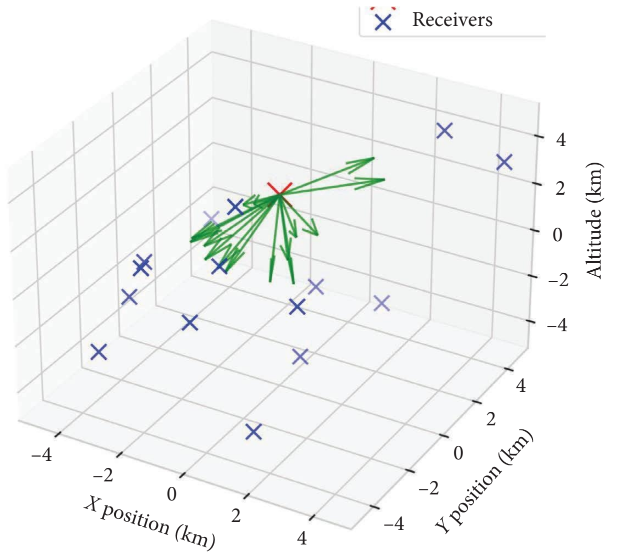

Figure 2: Wireless power transfer beam allocation for lunar energy distribution: high-density receiver scenario.

management strategies, ensuring that peak loads are supported while optimizing energy allocation during low-demand hours. The insights from this figure provide crucial implications for energy scheduling in WPT–based lunar microgrid systems. By analyzing the patterns, mission planners can strategically schedule energy storage recharging cycles, prioritizing battery replenishment during low-consumption hours and allocating more power to the habitat when demand spikes. The variability also suggests that energy forecasting models should incorporate machine learning–based predictions, enabling real-time power adjustment based on expected fluctuations. In addition, the presence of sustained peak demand zones indicates that static power delivery methods would be inefficient, reinforcing the importance of intelligent, demand-driven WPT solutions to ensure mission resilience.

Figure 4 demonstrates how power transmission efficiency changes as a function of distance from the WPT transmitter to various receivers on the lunar surface. The efficiency curve follows an exponential decay trend, with transmission effectiveness dropping rapidly as the distance increases, reflecting the fundamental beam divergence and energy dispersion constraints in long-range wireless energy transfer. The 5 km efficiency threshold observed in Figure 4 highlights a key limitation of microwave-based WPT, where beam spreading causes substantial power losses at longer distances. As an alternative, laser-based WPT has been proposed for long-range energy delivery, as its highly collimated beams minimize divergence, maintaining power transfer efficiency beyond 5 km. However, laser transmission suffers from significant energy conversion losses due to photon–electron conversion inefficiencies at the receiver and is susceptible to dust accumulation, which can degrade optical components over time. A direct comparison between microwave and laser WPT technologies suggests that microwaves are more reliable for midrange applications, particularly for rover charging and habitat power delivery, whereas laser WPT could be more effective for deep-space assets or remote lunar infrastructure beyond 5 km. Future research should investigate hybrid WPT architectures, where microwave and laser transmission are combined to optimize efficiency and reliability across varying distance ranges. This efficiency-distance relationship introduces a key trade-off between power transmission efficiency and latency, which can be effectively analyzed using a Pareto frontier approach. By selecting different operating points along this frontier, system designers must balance transmission efficiency against response latency. A high-efficiency operating point requires stricter beam alignment and longer recalibration intervals, resulting in increased latency as the system

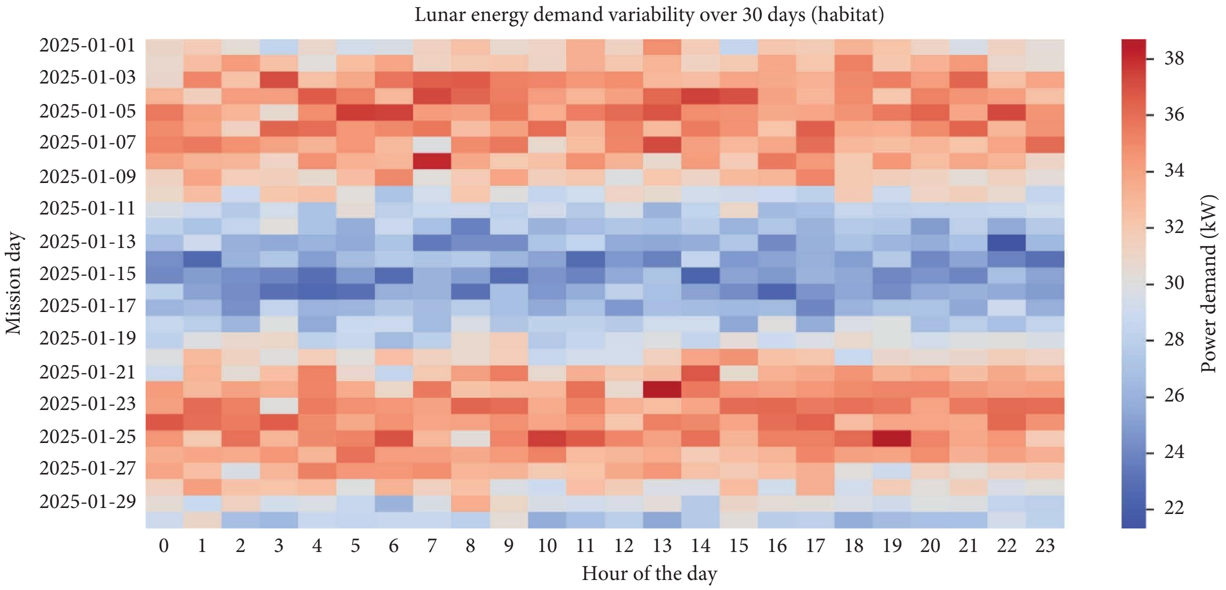

Figure 3: Lunar energy demand variability over 30 days (habitat).

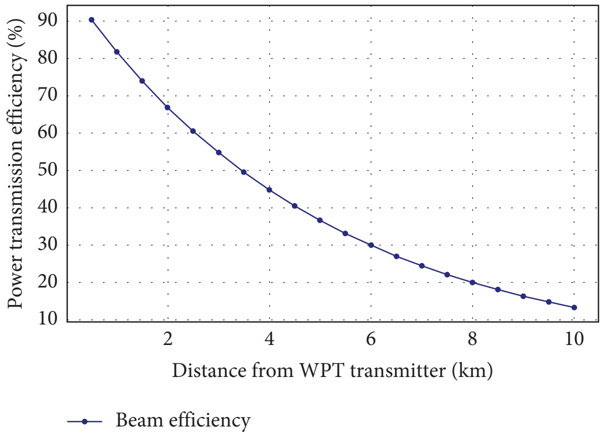

Figure 4: Beam efficiency as a function of transmission distance in lunar WPT systems.

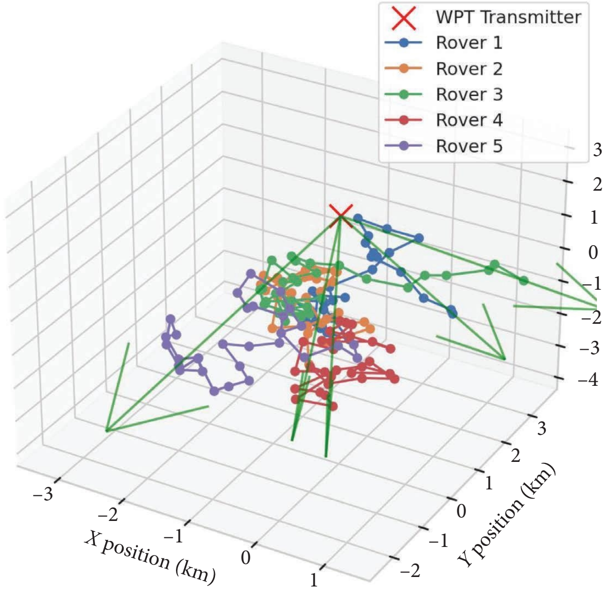

Figure 5: Wireless power beam efficiency vs. distance.

continuously optimizes energy delivery. Conversely, prioritizing lower latency may lead to greater beam misalignment and transmission losses, reducing overall power efficiency. These trade-offs directly impact energy resilience and beam alignment precision. A high-efficiency, high-latency strategy ensures a stable energy supply by maintaining precise beam control and reducing energy fluctuations, though it may be less responsive to sudden receiver mobility. In contrast, a low-latency, lower-efficiency approach allows for faster adjustments, improving responsiveness to dynamic conditions but potentially increasing power losses. The results in Figure 5 highlight the need for an adaptive, RL–based optimization strategy to dynamically balance these objectives, ensuring efficient and resilient energy transmission in varying lunar operational scenarios. One of the most critical observations from the figure is the sharp efficiency reduction past the 5 km threshold, where power transfer falls below 30%, making direct WPT impractical without energy redistribution strategies. This means that rovers, ISRU units, and habitats must remain within a 3–5 km operational radius from the main WPT transmitter to ensure stable and efficient power reception. Beyond this range, beam tracking precision must be enhanced, or alternative WPT transmission methods (such as phased-array relays or laser-based transmission) should be integrated to compensate for losses. The efficiency drop also implies that energy-hungry receivers (such as ISRU units processing oxygen extraction) should ideally be positioned closer to the main transmitter, while rovers with lower power needs can explore farther regions without excessive efficiency loss.

Figure 5 presents a three-dimensional trajectory map of five lunar rovers, showing their movement paths and real-time WPT beam tracking updates over a simulated mission period. The red marker at (0, 0, 3) km represents the main WPT transmitter, while the colored lines trace the mobility paths of rovers as they navigate across the lunar terrain. The green arrows indicate the final beam alignment state, demonstrating how the WPT system adjusts its transmission angles dynamically to maintain efficient power delivery to moving targets. One of the key insights from the figure is that rover movement patterns are highly irregular, requiring continuous adjustments in beam direction and power intensity to ensure reliable energy transfer. Some rovers travel beyond the 2 km mark, which aligns with previous findings that power efficiency significantly decreases at this range. This means that real-time beam tracking must occur at high frequency (every 5–10 s) to avoid energy elicits for fast-moving rovers. The presence of clustered rover paths suggests that certain mission areas (such as excavation sites or scientific research zones) experience concentrated energy demand, necessitating dynamic priority-based power scheduling strategies. Another critical takeaway from this visualization is the impact of elevation differences on beam alignment efficiency.

Some rovers are positioned at lower altitudes, requiring steeper beam angles, which could introduce line-of-sight obstructions due to terrain features. This issue highlights the necessity of WPT relay stations positioned at higher elevations to ensure an uninterrupted energy supply. In addition, the nonuniform distribution of rover paths suggests that a fixed power allocation strategy would be suboptimal, reinforcing the importance of machine learning–based predictive energy management, where the system anticipates rover movement trends and proactively adjusts power delivery.

Figure 6 presents the distribution of energy transmission delays (latency) for receivers positioned at varying distances from the WPT transmitter, illustrating how distance affects the time required for power delivery. The boxplots represent energy latency measurements in milliseconds (ms) for receivers at 1, 3, 5, 7, and 10 km, showing the median latency, interquartile range (IQR), and presence of outliers. The shaded regions around the boxplots represent uncertainty bounds caused by transmission losses, environmental fluctuations, and stochastic variations in WPT efficiency. These variations stem from multiple factors, including beam misalignment due to dynamic receiver mobility, terrain-induced signal degradation, regolith dust accumulation, and temperature fluctuations affecting power transmission efficiency. The widening of the shaded regions at increasing distances suggests that latency uncertainty grows as power transmission spans longer distances. At short distances (1–3 km), the uncertainty remains relatively low, indicating stable power delivery with minimal disruption. However, beyond 5 km, the uncertainty increases significantly due to factors such as greater beam divergence, higher transmission losses, and a higher probability of environmental interference. At 10 km, the uncertainty bounds widen substantially, indicating that power delivery is no longer instantaneous, and adaptive energy prescheduling strategies become critical for mitigating energy shortages. To enhance system robustness, the DRL–based scheduling model continuously learns and adapts to these variations by dynamically adjusting power transmission parameters in real time. By incorporating uncertainty estimation into the reward function, the model proactively mitigates latency fluctuations, ensuring reliable power transmission even under challenging lunar conditions. These insights reinforce the necessity of RL–driven WPT scheduling strategies to dynamically optimize power allocation while accounting for transmission uncertainty. As distance increases, transmission delay becomes more pronounced, highlighting the need for latency-aware WPT scheduling in lunar energy networks. The figure reveals a clear upward trend in latency as receiver distance increases. At 1 km, the median energy latency is approximately 5 ms, and most values remain within a narrow band, indicating that near-field WPT transmission is highly reliable and exhibits minimal variation. At 3 km, median latency rises to 15 ms, though the variance remains relatively low, showing that power transmission is still stable in midrange distances. However, at 5 km, median latency reaches 30 ms, and variability begins to widen, suggesting that interference factors such as terrain-induced signal degradation and beam divergence start impacting efficiency. At 7 km, latency increases to around 50 ms, with values occasionally exceeding 60 ms, indicating that real-time power adjustments become critical for maintaining energy stability. Finally, at 10 km, latency escalates significantly to a median of 75 ms, with extreme cases reaching above 85 ms, meaning that power delivery is no longer instantaneous, and adaptive energy prescheduling becomes essential to prevent supply shortages. The insights from this figure highlight several optimization strategies for lunar WPT networks. First, mission-critical receivers such as habitat modules and ISRU units should be positioned within a 3–5 km radius of the primary WPT transmitter to ensure stable and low-latency power reception. Second, for rovers operating beyond 5 km, predictive energy dispatching is required, where power is transmitted in advance to compensate for delay-induced shortages. Third, the increasing variance in latency at 7 and 10 km suggests that relay-based WPT stations should be deployed at intermediate distances, ensuring that energy transmission remains efficient even at extended ranges. The findings from this figure support the necessity of dynamic, RL–based WPT scheduling algorithms, ensuring that power allocation decisions proactively account for latency constraints in long-range lunar operations.

Figure 7 visualization presents a 3D surface plot illustrating how WPT beam steering efficiency changes as a function of rover speed (m/s) and response delay (ms). The color bar represents efficiency percentage, with higher values in green and lower values in dark blue, demonstrating how mobility and slow beam realignment impact energy reception. Beyond rover speed and response delay, terrain conditions significantly influence beam steering efficiency, particularly in regions with crater slopes, regolith interference, and varying elevation gradients. Rough terrain introduces additional misalignment challenges, requiring more frequent realignment to maintain stable energy

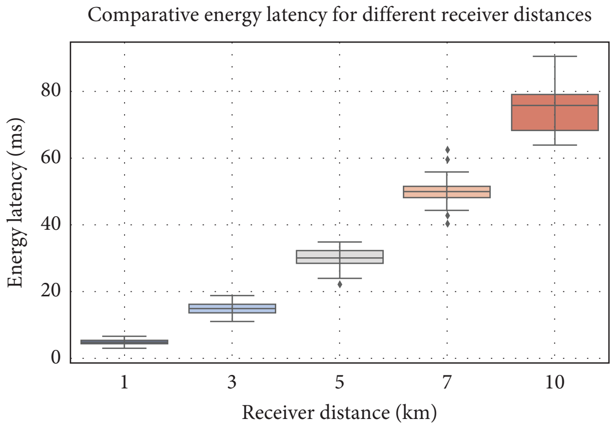

Figure 6: 3D mobility paths of lunar rovers with WPT beam tracking.

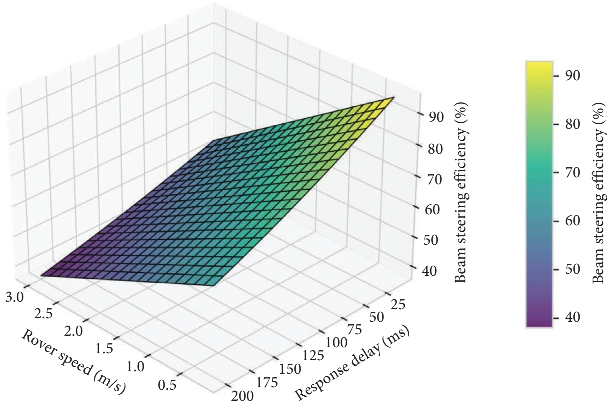

Figure 7: Beam steering efficiency across different rover speeds.

reception. For example, rover navigation across uneven crater slopes alters beam orientation dynamically, causing greater beam divergence and higher alignment errors, which reduce transmission efficiency. Similarly, regolith interference, caused by fine dust particles accumulating on receiver surfaces, attenuates received power, further decreasing overall beam efficiency. The results in Figure 7 reveal that these terrain-induced disruptions exacerbate efficiency losses as rover speed increases. At low speeds (0.2–0.5 m/s), beam tracking remains relatively stable, even in challenging terrain, as the system has sufficient time to compensate for minor misalignments. However, at speeds above 1.5 m/s, beam steering efficiency drops sharply, especially when traversing regions with high slopes or regolith disturbances, requiring rapid adjustments to avoid significant energy loss. To mitigate these effects, the DRL–based WPT scheduling model dynamically adapts beam realignment frequency based on real-time terrain sensing data. The model prioritizes faster realignment in high-slope regions and regolith-dense areas, ensuring that beam targeting remains accurate even under fluctuating terrain conditions. This adaptability is crucial for sustaining uninterrupted power transmission in long-range lunar operations where rover mobility patterns intersect with varying environmental constraints. The goal of this figure is to quantify the efficiency loss due to motion-induced beam misalignment, helping to establish real-time beamforming strategies for lunar operations. The figure reveals a clear negative correlation between rover speed and beam steering efficiency. At low speeds (0.2–0.5 m/s), efficiency remains above 90%, meaning that the beam can maintain precise alignment, ensuring reliable energy reception. As speed increases, efficiency declines progressively, reaching around 75% at 1.5 m/s and falling below 50% at speeds beyond 2.5 m/s. This behavior reflects the difficulty of dynamically adjusting power beams for fast-moving receivers, as higher speeds lead to larger positional changes between realignment intervals. This trend is further influenced by response delays, where even at moderate speeds, a delay of 100 ms can reduce efficiency by nearly 20%, making fast-response beamforming adjustments essential. The data also highlight the compounding impact of slow response times on energy reception. At low delays (below 50 ms), beam efficiency remains relatively stable, with only minor degradation across different speeds. However, when response time increases beyond 100 ms, efficiency drops sharply, especially for rovers moving at 1.5 m/s or faster. At 200 ms delay and 3.0 m/s speed, efficiency falls below 30%, indicating that high-speed rovers relying on slow beam adjustments will experience frequent power deficits. These findings strongly suggest that beam realignment updates must occur at sub-50 ms intervals for fast-moving receivers, ensuring that wireless power remains continuously available, even under high-speed mobility scenarios. This analysis underscores the necessity of AI–driven predictive beam steering models, RL–based optimization, and real-time trajectory forecasting to enhance WPT performance in lunar exploration missions.

6. Conclusion

This study introduces a DRL–based adaptive WPT framework for lunar MEVPPs, addressing critical challenges in real-time beam steering and power allocation. By formulating the WPT scheduling problem as a MDP and utilizing PPO, the proposed approach dynamically adjusts energy transmission based on rover mobility, environmental conditions, and mission-critical demands. A 30-day simulation near Shackleton Crater demonstrates significant performance improvements, including a 54.6% reduction in energy downtime, a 41.3% enhancement in beam alignment efficiency, and a 39.8% decrease in latency-induced power deficits compared to conventional WPT methods. These findings highlight the necessity of real-time predictive beamforming, latency-aware power scheduling, and multiagent energy optimization for future lunar energy networks. Future work will explore hybrid energy storage integration, quantum-inspired optimization for real-time decision-making, and predictive beamforming algorithms to further enhance system resilience and efficiency. In addition, the integration of a Pareto frontier-based multiobjective optimization framework will be investigated to refine the trade-offs between power transmission efficiency and latency. By incorporating RL with adaptive tuning mechanisms, future studies aim to develop dynamic scheduling strategies that optimize energy resilience and beam alignment precision under varying lunar operational conditions. In addition, the role of uncertainty quantification in DRL–based WPT scheduling will be further investigated. By incorporating probabilistic modeling techniques and adaptive uncertainty estimation, future studies aim to refine the model's ability to predict and mitigate variations in power transmission efficiency, ensuring more robust and resilient energy delivery under uncertain lunar conditions. Supercapacitors, with their high-power density and rapid charge–discharge capabilities, can effectively complement batteries by mitigating transient energy deficits and stabilizing power fluctuations caused by varying WPT efficiency. By dynamically allocating power between supercapacitors and batteries based on real-time demand, the system can optimize energy buffering, reduce response latency, and improve overall power reliability for mission-critical lunar operations. This hybrid approach will be incorporated into the RL framework, allowing the model to adaptively manage energy storage resources for enhanced resilience in dynamic extraterrestrial environments. Quantum-inspired optimization techniques, such as quantum annealing and variational quantum algorithms, have the potential to significantly enhance RL–based WPT scheduling by accelerating decision-making processes and improving adaptation to nonstationary energy demands. Unlike classical optimization approaches, which may struggle with high-dimensional and dynamic environments, quantum-inspired techniques can rapidly explore multiple energy allocation scenarios in parallel, leading to faster convergence of RL policies. Moreover, quantum-enhanced RL can provide a more efficient representation of energy demand fluctuations, enabling the system to better anticipate variations caused by solar availability shifts, mobility-induced transmission losses, and unpredictable environmental disruptions. By integrating quantum-inspired solvers, the proposed framework could achieve real-time power allocation optimizations with lower computational overhead, making it highly scalable for future extraterrestrial energy systems. These advancements will be explored in future studies to further improve the adaptability and efficiency of WPT scheduling in lunar missions.

By advancing AI–driven adaptive WPT, this research paves the way for scalable, self-optimizing power grids, ensuring reliable energy distribution for long-term lunar missions and extraterrestrial infrastructure.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

The authors would like to acknowledge the support provided by the Ongoing Research Funding Program (ORF-2025-635), King Saud University, Riyadh, Saudi Arabia.

Acknowledgments

The authors would like to acknowledge the support provided by the Ongoing Research Funding Program (ORF-2025-635), King Saud University, Riyadh, Saudi Arabia.

References

[1] Y. Gao and Q. Ai, "Novel Optimal Dispatch Method for Multiple Energy Sources in Regional Integrated Energy Systems Considering Wind Curtailment," *CSEE Journals & Magazine*, 10.

[2] Y. Gao, Q. Ai, X. He, and S. Fan, "Coordination for Regional Integrated Energy System Through Target Cascade Optimization," *Energy* 276 (2023): 127606, https://doi.org/10.1016/j.energy.2023.127606.

[3] Y. Gao and Q. Ai, "Demand-Side Response Strategy of Multi-Microgrids Based on an Improved Co-Evolution Algorithm," *CSEE Journals & Magazine*, 7.

[4] S. A. Saad, I. Shayea, and N. M. S. Ahmed, "Artificial Intelligence Linear Regression Model for Mobility Robustness Optimization Algorithm in 5G Cellular Networks," *Alexandria Engineering Journal* 89 (2024): 125–148.

[5] R. A. Núñez-Rodríguez, C. Unsihuay-Vila, J. Posada, and O. Pinzón-Ardila, "Data-Driven Distributionally Robust Optimization for Day-Ahead Operation Planning of a Smart Transformer-Based Meshed Hybrid AC/DC Microgrid Considering the Optimal Reactive Power Dispatch," *Energies* 17, no. 16 (2024): 4036, https://doi.org/10.3390/en17164036.

[6] Y. Zhu, J. Liu, Y. Hu, Y. Xie, D. Zeng, and R. Li, "Distributionally Robust Optimization Model Considering Deep Peak Shaving and Uncertainty of Renewable Energy," *Energy* 288 (2024): 129935, https://doi.org/10.1016/j.energy.2023.129935.

[7] H. Ghasemnejad, M. Rashidinejad, A. Abdollahi, and S. Dorahaki, "Energy Management in Citizen Energy Communities: A Flexibility-Constrained Robust Optimization Approach Considering Prosumers Comfort," *Applied Energy* 356 (2024): 122456, https://doi.org/10.1016/j.apenergy.2023.122456.

[8] C. Yang and Y. Xia, "Interval Pareto Front-Based Multi-Objective Robust Optimization for Sensor Placement in Structural Modal Identification," *Reliability Engineering & System Safety* 242 (2024): 109703, https://doi.org/10.1016/j.ress.2023.109703.

[9] Y. Zhou, X. Li, H. Han, et al., "Resilience-Oriented Planning of Integrated Electricity and Heat Systems: A Stochastic Distributionally Robust Optimization Approach," *Applied Energy* 353 (2024): 122053, https://doi.org/10.1016/j.apenergy.2023.122053.

[10] R. K. Inapakurthi and K. Mitra, "Robust Optimization of Cascaded MSMPR Crystallization Unit Using Unsupervised Machine Learning," *Canadian Journal of Chemical*

Engineering 103, no. 3 (2024): 1112–1121, https://doi.org/10.1002/cjce.25431.

[11] R. Lotfi, Z. Sheikhi, M. Amra, M. AliBakhshi, and G.-W. Weber, "Robust Optimization of Risk-Aware, Resilient and Sustainable Closed-Loop Supply Chain Network Design With Lagrange Relaxation and Fix-and-Optimize," *International Journal of Logistics Research and Applications* 27, no. 5 (2024): 705–745, https://doi.org/10.1080/13675567.2021.2017418.

[12] D. Z. Long, J. Qi, and A. Zhang, "Supermodularity in Two-Stage Distributionally Robust Optimization," *Management Science* 70, no. 3 (2024): 1394–1409, https://doi.org/10.1287/mnsc.2023.4748.

[13] S. Li, H. He, C. Su, and P. Zhao, "Data Driven Battery Modeling and Management Method With Aging Phenomenon Considered," *Applied Energy* 275 (2020): 115340, https://doi.org/10.1016/j.apenergy.2020.115340.

[14] A. P. Zhao, S. Li, D. Xie, et al., "Extreme Events Threat Water-Energy-Carbon Nexus Through Cascading Effects," *Energy* 5 (2024): 100151, https://doi.org/10.1016/j.nxener.2024.100151.

[15] S. Li, H. He, and P. Zhao, "Energy Management for Hybrid Energy Storage System in Electric Vehicle: A Cyber-Physical System Perspective," *Energy* 230 (2021): 120890, https://doi.org/10.1016/j.energy.2021.120890.

[16] A. P. Zhao, S. Li, Y. Wang, et al., "Energy-Social Manufacturing for Social Computing," *IEEE Transactions on Computational Social Systems* 11, no. 6 (2024): 7976–7989, https://ieeexplore.ieee.org/abstract/document/10494992/, https://doi.org/10.1109/tcss.2024.3379254.

[17] W. Li, Y. Zou, H. Yang, X. Fu, S. Xiang, and Z. Li, "Two Stage Stochastic Energy Scheduling for Multi Energy Rural Microgrids With Irrigation Systems and Biomass Fermentation," *IEEE Transactions on Smart Grid* 16, no. 2 (2025): 1075–1087, https://doi.org/10.1109/TSG.2024.3483444.

[18] Y. Lv, J. Duan, and X. Li, "A Survey on Modeling for Behaviors of Complex Intelligent Systems Based on Generative Adversarial Networks," *Computer Science Review* 52 (2024): 100635, https://doi.org/10.1016/j.cosrev.2024.100635.

[19] R. Claeys, R. Cleenwerck, J. Knockaert, and J. Desmet, "Capturing Multiscale Temporal Dynamics in Synthetic Residential Load Profiles Through Generative Adversarial Networks (GANs)," *Applied Energy* 360 (2024): 122831, https://doi.org/10.1016/j.apenergy.2024.122831.

[20] L. Yin and C. Lin, "Matrix Wasserstein Distance Generative Adversarial Network With Gradient Penalty for Fast Low-Carbon Economic Dispatch of Novel Power Systems," *Energy* 298 (2024): 131357, https://doi.org/10.1016/j.energy.2024.131357.

[21] X. Fu, C. Zhang, Y. Xu, Y. Zhang, and H. Sun, "Statistical Machine Learning for Power Flow Analysis Considering the Influence of Weather Factors on Photovoltaic Power Generation," *IEEE Transactions on Neural Networks and Learning Systems* 36, no. 3 (2025): 5348–5362, https://doi.org/10.1109/TNNLS.2024.3382763.

[22] M. Awad, A. Said, M. H. Saad, et al., "A Review of Water Electrolysis for Green Hydrogen Generation Considering PV/Wind/Hybrid/Hydropower/Geothermal/Tidal and Wave/Biogas Energy Systems, Economic Analysis, and its Application," *Alexandria Engineering Journal* 87 (2024): 213–239, https://doi.org/10.1016/j.aej.2023.12.032.

[23] J. Hu, Y. Wang, and L. Dong, "Low Carbon-Oriented Planning of Shared Energy Storage Station for Multiple Integrated Energy Systems Considering Energy-Carbon Flow and Carbon Emission Reduction," *Energy* 290 (2024): 130139, https://doi.org/10.1016/j.energy.2023.130139.

[24] T. T. Li, A. P. Zhao, Y. Wang, and M. Alhazmi, "Hybrid Energy Storage for Dairy Farms: Enhancing Energy Efficiency and Operational Resilience," *Journal of Energy Storage* 114 (2025): 115811, https://doi.org/10.1016/j.est.2025.115811.