TWINERGY: Digital Twin Empowered Practical Energy Saving in Heterogeneous Mobile Networks

Yibo Ma[†], Tianxin Wang[†], Haoye Chai[‡], Yong Li[§], Mahesh K. Marina[†]
The University of Edinburgh[†], Beijing University of Posts and Telecommunications[‡], Tsinghua University§

Abstract

Energy consumption has become a pressing concern as mobile networks expand rapidly, with the radio access network accounting for the dominant share. Among energy-saving techniques, cell shutdown offers large saving potential: it selectively switches off capacity cells in heterogeneous networks while a coverage layer preserves service. However, optimizing cell shutdown policies faces two key challenges: (i) live trial-and-error is prohibitively risky, while traditional simulators cannot capture the complex dynamics needed for reliable evaluation; (ii) operator-deployed threshold policies are interpretable but hard to optimize at scale. On the other hand, binary status optimization is more tractable but produces policies incompatible with operator workflows. We propose Twinergy, a digital twin-empowered framework for practical cell shutdown at scale. Twinergy builds a Network Digital Twin for safe virtual evaluation, and decouples policy design into two stages: first optimizing binary on/off decisions under QoS constraints, and then mapping them into deployable thresholds, with a dedicated module for each. For scalable decision optimization, we design Cable, a multi-agent reinforcement learning framework with shared policies, causal graph-based spatio-temporal encoding, and credit-guided cooperative learning. For threshold synthesis, we design a differentiable mapping module that introduces SoftControl as a surrogate for hard dual-threshold logic, enabling thresholds to be calibrated in continuous space via stable supervised optimization. Experiments on a digital twin built from 14,349 real-world base stations and deployments across three commercial trial networks with over 3,400 cells show that Twinergy achieves up to 26.0% and 12.4% energy savings in 4G and 5G digital-twin evaluation, respectively, and delivers up to 13.07% and 12.20% energy-efficiency improvements in rural and dense 5G live networks without degrading QoS.

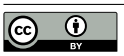
CCS Concepts

• **Networks** → **Mobile networks**; **Network management**.

Keywords

Mobile networks, network digital twins, cell shutdown

ACM Reference Format:
Yibo Ma[†], Tianxin Wang[†], Haoye Chai[‡], Yong Li[§], Mahesh K. Marina[†]. 2026. TWINERGY: Digital Twin Empowered Practical Energy Saving in Heterogeneous Mobile Networks. In *The Twenty-seventh International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '26), November 23-26, 2026, Tokyo, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3842721.3849657



This work is licensed under a Creative Commons Attribution 4.0 International License.
MobiHoc '26, Tokyo, Japan
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2975-1/2026/11
https://doi.org/10.1145/3842721.3849657

1 Introduction

Mobile networks have become critical infrastructure for modern cities and emerging digital services [8]. Their rapid expansion, especially in heterogeneous 4G/5G deployments, has made energy consumption a pressing concern. Since the radio access subsystem dominates this footprint [13], improving network energy efficiency is a first-order problem for operators. Among energy-saving techniques, we focus on *cell shutdown*. Unlike micro-timescale mechanisms such as symbol-level or RF channel shutdown, which require sub-millisecond execution and only optimize limited radio components [1, 16], cell shutdown deactivates an entire cell and offers greater saving potential. Unlike deep sleep, whose high reactivation overhead limits it to stable low-traffic periods [17], cell shutdown operates over practical control intervals (minutes to one hour) and adapts to recurring traffic fluctuations. In layered heterogeneous networks, a low-frequency coverage layer remains active to maintain basic service, while paired high-frequency capacity cells can be selectively switched on or off and their traffic offloaded to neighboring cells [27]. A cell shutdown policy therefore specifies the binary on/off status of cells at each decision interval, making it a practical lever for network-wide energy reduction [8].

Optimizing cell shutdown in real-world networks is challenging. **First**, direct trial-and-error in live networks is prohibitively risky: a poor policy can overload neighboring cells, degrade Quality of Service (QoS), or trigger service disruption. Practical optimization therefore requires high-fidelity network modeling. However, stochastic user association, inter-cell traffic offloading, and heterogeneous energy profiles, are difficult to capture with traditional analytical models [25] or discrete-event simulators [22], which either oversimplify heterogeneity or become inefficient at scale. **Second**, there is a fundamental tension between optimization tractability and deployment compatibility. Threshold-based policies [32] are interpretable and fit existing operator workflows, but discrete logic and inter-cell coupling make heterogeneous thresholds difficult to optimize at scale [4, 6, 19]. Direct binary optimization [5, 20, 29] sidestep this difficulty but produce per-slot action sequences that are not deployable as operator-facing threshold configurations.

To bridge these gaps, we propose Twinergy, a digital twin-empowered framework for scalable cell shutdown policy. For the **first** challenge, Twinergy builds a real-data-driven Network Digital Twin (NDT) [12] to replace risky live trials with virtual evaluation. The NDT combines a Variational Autoencoder (VAE) [10] with 3GPP path-loss-based user association [1] for stochastic connection dynamics, a Granger-causality-guided causal graph for inter-cell traffic offloading, and cell-specific energy regressors calibrated from operational data. For the **second** challenge, we decouple optimization tractability from deployment compatibility. Twinergy

therefore decouples the problem into two stages: Stage 1 optimizes binary cell status for energy savings under QoS constraints, avoiding the intractable threshold space; Stage 2 maps the optimized binary status to deployable threshold configurations that reproduce the optimized decisions. This decomposition separates *decision optimization* from *deployment alignment*, yielding a policy that is both high-performing and directly deployable.

Despite high-fidelity NDT and two-stage formulation, challenges remain. In Stage 1, cell decisions are strongly coupled, since shutting down one cell changes traffic distribution, coverage compensation, and feasible actions of others, causing the joint action space to grow exponentially with network scale. In Stage 2, the operator-defined rules are non-differentiable step functions, making the inverse mapping from binary status to continuous thresholds difficult to optimize. To address these challenges, TWINERGY introduces two key modules. For Stage 1, we design CABLE, a scalable multi-agent reinforcement learning (RL) framework. CABLE models capacity cells as cooperative agents with a shared actor, reduces policy complexity from exponential to linear scaling. It uses a Temporal Graph Convolutional Network (TGCN) [33] with causal inter-cell graphs for spatio-temporal state encoding, and a credit-guided teacher-student dual-critic to disambiguate local decisions from global energy/QoS outcomes, stabilizing cooperative learning. For Stage 2, we develop a differentiable threshold-mapping module (TM) with SoftControl, a differentiable surrogate of the hard dual-threshold logic, so that heterogeneous on/off thresholds can be calibrated in continuous space through stable supervised optimization.

Our main contributions are summarized as follows:

- **A deployable cell shutdown framework (§4).** We design TWINERGY as a unified decision-to-deployment framework that optimizes binary cell status in a high-fidelity NDT and maps them with operator-compatible thresholds. This design resolves the tension between optimization tractability and deployability.
- **A data-driven NDT with high-fidelity (§4.2).** We build a real-data-driven NDT for cell shutdown evaluation. It models stochastic user association, inter-cell coordination, and heterogeneous cell energy profiles, providing a safe and realistic environment for assessing policies without trial-and-error on a live network.
- **Two modules for scalable optimization and deployment alignment (§4.3 and §4.4).** We develop CABLE, a multi-agent RL framework with a shared actor, TGCN-based encoding, and credit-guided dual critics for Stage 1 optimization, together with TM, a differentiable SoftControl-based module that learns heterogeneous thresholds from optimized binary actions for Stage 2.
- **Comprehensive evaluation in digital twin (§5) and live operator networks (§6).** We evaluate TWINERGY on a network digital twin built from a real-world dataset of 14,349 BSs and via deployments in three commercial trial networks with more than 3,400 cells in total. In digital twin evaluation, TWINERGY achieves up to 26.0% and 12.4% energy savings for 4G and 5G under threshold-based policies while keeping reliability degradation within acceptable limits. These savings yield approximately 6.25M annual Operating Expenditure (OPEX) savings and 48.7k tons CO_2e reduction. In real-world deployment, it yields up to 13.07% and 12.20% energy efficiency improvements in rural and dense 5G scenarios, respectively, without degrading QoS.

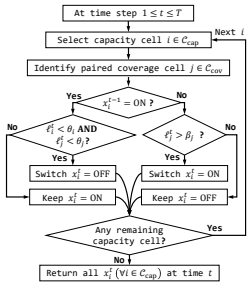


Figure 1: Dual-threshold policy [32].

Symbol	Definition
$\rho^t_{k,\text{UL}}, \rho^t_{k,\text{DL}}$	Uplink (UL) and Downlink (DL) traffic volumes in cell k at time t
μ^t_k	Number of users associated with cell k at time t
$w^t_{k,\text{UL}}, w^t_{k,\text{DL}}$	UL and DL Physical Resource Block (PRB) usage ratios of cell k at time t
R^t	Users' spatial distribution at time t
γ^t	User channel conditions at time t
$\mathbf{X}^t$	Set of binary on-off statuses of all cells at time t
ℓ^t_k	Load tuple for individual cell k at time t
$f(\cdot)$	Function determining user association
$g(\cdot)$	Function determining traffic offloading
$h(\cdot)$	Function determining PRB usage ratios from traffic and channels
p^{base}_b	Baseline static power consumption of base station b
p^{cell}_k	Per-cell load-dependent power consumption
$\rho^{\max}_{k,\text{UL}}, \rho^{\max}_{k,\text{DL}}$	Cell k's capacity limits for traffic volumes at time t
θ_i	Switch-off threshold vector of capacity cell i
θ_j, β_j	Switch-off and switch-on thresholds of coverage cell j
$c^t_{u,k}$	Binary connection status of user u to cell k at time t

Table 1: List of critical symbols and definitions.

2 Background and Related Work

2.1 Network Digital Twins

A fundamental limitation of current cell shutdown methods is their reliance on idealized or oversimplified models, which fail to capture heterogeneous traffic, spatial coupling, and fast demand shifts. Existing simulators include discrete-event-driven models [22] and analytical models [25] . However, analytical models often miss complex network dynamics, while discrete-event simulators become inefficient at large scales. Similarly, constructing fine-grained models [2] via physics-based platforms like Omniverse [24] provides high-fidelity ray-tracing but suffers from slow physical mechanism simulation that fails to scale. Recent NDT [12] research instead adopts data-driven models, leveraging operational data for low-latency traffic prediction [21] . Although these studies capture the traffic patterns and urban semantics for network simulation, an efficient large-scale NDT is still lacking. Consequently, to safely evaluate cell shutdown policies, high-fidelity NDT must be developed to avoid QoS degradation of trial-and-error in live networks.

2.2 Cell Shutdown

Cell shutdown policies determine the on/off status of cells at each decision epoch [8]. They account for channel conditions and inter-cell offloading, as shutting down a cell transfers its users to overlapping cells that must have enough spare capacity to avoid QoS degradation. Existing works follow two lines: 1) deriving shutdown rules from traffic load thresholds, and 2) directly making binary decisions for each cell. Both aim to balance energy efficiency and QoS, but differ in optimization space, interpretability, and deployability. Henceforth, we use *cell status* to refer to a cell's on/off status.

Cell shutdown with threshold-based policies. A widely deployed policy is the dual-threshold policy [32], which balances energy saving and response time while preventing frequent status flapping through separate on/off thresholds. As shown in Fig. 1, a capacity cell switches status when the corresponding condition is met. In practice, operators define these conditions using multiple traffic indicators, such as PRB usage and user counts, combined through AND/OR rules. This approach is highly interpretable, as operators can directly inspect and tune the resulting decision boundaries.

Dual-threshold policies can be homogeneous, where one threshold set is shared across the network, or heterogeneous, where different cell pairs use different thresholds. Existing studies focus on heterogeneous dual-threshold policies using heuristic methods [32], optimization methods [19], and deep reinforcement learning (RL) [19]. However, this line is difficult to optimize at scale. As in Fig. 1,

threshold-based policies are piecewise step functions. When multiple indicators are combined through rules, the mapping from thresholds to energy and QoS becomes coupled, non-linear, and non-differentiable, with a combinatorial action space that makes exploration sample-inefficient. Gradient-based methods therefore suffer from near-zero gradients, discontinuities at switching boundaries, and unstable convergence. Some studies discretize the action [4] to improve stability, but this further enlarges the search burden and may reduce optimality. Causal modeling [6] can narrow the search space, but they require highly diverse live-network test data and are difficult to maintain in dynamic environments. Consequently, operators often deploy homogeneous threshold policies, sacrificing heterogeneity and potential energy savings.

Cell shutdown with direct binary control. A second line formulates cell shutdown as a binary optimization problem, i.e., deciding the binary status (1 for on and 0 for off) of each cell under QoS constraints. Owing to dynamic traffic, cell heterogeneity, and complex inter-cell coordination, this problem is non-convex and NP-hard [3]. Early studies develop traffic-driven solutions using genetic algorithms [5] and heuristics [20] . Recent work adopts RL to optimize cell status jointly [29]. Compared with threshold optimization, this line avoids learning high-dimensional continuous threshold vectors. However, these methods often oversimplify inter-cell coordination and overlook cell heterogeneity by applying uniform models across diverse network tiers. Their outputs are also direct binary decisions rather than operator-facing control thresholds. As a result, the learned policies are less interpretable and difficult to integrate into current operator networks, where cell shutdown is typically executed through threshold-based logic such as the dual-threshold policy. Deploying such methods would therefore require operators to implement and maintain new decision modules.

Implications and motivation for our design. The above two lines reveal a trade-off between deployability and optimization tractability: threshold-based policies fit current operator practice but are hard to optimize end-to-end, whereas direct binary control is easier to optimize but harder to deploy in live networks. This trade-off motivates a method that preserves both properties. Accordingly, in §3.2 we adopt a two-stage formulation: Stage 1 optimizes binary shutdown decisions in a tractable space, and Stage 2 maps the resulting policy back to thresholds using supervised learning, avoiding inefficient exploration in threshold space. In this way, our design retains the optimization advantage of binary control while remaining compatible with existing threshold-based operation without introducing a new control interface.

3 Problem Statement and Challenges

3.1 System Model

We consider a heterogeneous mobile network consisting of a set of base stations $\mathcal{B}$ serving a set of users $\mathcal{U}$. Let C be the set of all cells and C_b be the set of cells with base station b. As mentioned in §2, there exist two types of cells, i.e., capacity cells C_{cap} and coverage cells C_{cov}, and each capacity cell is paired with a coverage cell. The binary on-off status defined as $x_k^t \in \{0,1\}$ of cell k is determined at each time interval $t = 1, \ldots, T$, with $x_k^t = 0$ indicating cell shutdown. Coverage cells are always active, i.e., $x_j^t = 1, \forall j \in C_{\text{cov}}$.

The number of users associated with cell k at interval t is denoted by μ_k^t. Given an association policy $f(\cdot)$, it is determined by cell status, users' spatial distribution $\mathbf{R}^t$, and channel conditions $\boldsymbol{\gamma}^t = \{\gamma_u^t\}_{u\in\mathcal{U}}$ (measured by RSRP), which is given by

$$\mu_k^t = f(\mathbf{X}^t, \mathbf{R}^t, \boldsymbol{\gamma}^t), \text{ s.t. } \mu_k^t = 0 \text{ if } x_k^t = 0, \tag{1}$$

where $\mathbf{X}^t = \{x_k^t\}_{k\in C}$. We denote $\rho_{k,\text{UL}}^t$ and $\rho_{k,\text{DL}}^t$ as UL and DL traffic volumes in each cell k at interval t in GB, respectively, and denote $\rho_k^t = \{\rho_{k,\text{UL}}^t, \rho_{k,\text{DL}}^t\}$ as per-cell traffic volume. Given a traffic offloading policy denoted by $g(\cdot)$, ρ_k^t is determined by the connected-UE count vector $\mathbf{U}^t = \{\mu_k^t\}_{k\in C}$, cell status $\mathbf{X}^t$, and last-interval traffic volume distribution $\boldsymbol{\rho}^{t-1}$:

$$\rho_k^t = g(\mathbf{U}^t, \mathbf{X}^t, \boldsymbol{\rho}^{t-1}), \text{s.t. } \rho_{k,\text{UL}}^t = 0, \rho_{k,\text{DL}}^t = 0 \text{ if } x_k^t = 0 \tag{2}$$

where $\boldsymbol{\rho}^t = \{\rho_k^t\}_{k\in C}$ depends on the channel conditions through the association $\mathbf{U}^t$ defined by Eq. (1) and the resulting traffic-offloading process. Then, UL and DL PRB usage ratios of cell k at t are denoted as $w_{k,\text{UL}}^t$ and $w_{k,\text{DL}}^t$, respectively, and determined by UL and DL traffic volumes and channel conditions with function $h(\cdot)$:

$$w_{k,\text{UL}}^t = h(\rho_{k,\text{UL}}^t, \boldsymbol{\gamma}^t), \quad w_{k,\text{DL}}^t = h(\rho_{k,\text{DL}}^t, \boldsymbol{\gamma}^t). \tag{3}$$

Taking Eq. (1)-(3) together, the network-wide cell load is denoted by $\mathbf{L}^t = \{\ell_k^t\}_{k\in C}$, where the load ℓ_k^t for an individual cell k is defined as a three-element tuple consisting of the UL PRB usage ratio $w_{k,\text{UL}}^t$, the DL PRB usage ratio $w_{k,\text{DL}}^t$, and the connected UE count μ_k^t.

Power consumption of each BS b consists of two parts: i) baseline power P_b^{base} and ii) per-cell load-dependent power $P_k^{\text{cell}}, \forall k \in C_b$, which is defined as a function of cell status x_k^t and cell load ℓ_k^t:

$$P_k^{\text{cell}}(x_k^t, \ell_k^t) = \begin{cases} P_k^{\text{on}}(\ell_k^t), & \text{if } x_k^t = 1, \\ P_k^{\text{off}}, & \text{if } x_k^t = 0, \end{cases} \tag{4}$$

where P_k^{on} and P_k^{off} denote the consumed power levels in the statuses. To this end, total network energy consumption E in each time unit t is determined by the on-off statuses $\mathbf{X}^t$ and corresponding cell loads $\mathbf{L}^t$ of all the cells in the network, expressed as follows:

$$E(\mathbf{X}^t, \mathbf{L}^t) = \sum_{b\in\mathcal{B}} \mathbb{I}\left(\sum_{k\in C_b} x_k^t > 0\right)\left[P_b^{\text{base}} + \sum_{k\in C_b} P_k^{\text{cell}}(x_k^t, \ell_k^t)\right], \tag{5}$$

where $\mathbb{I}(\cdot)$ is the indicator function that returns 1 if the condition holds and 0 otherwise. The main notation used in this work is summarized in Table 1.

3.2 Two-Stage Cell Shutdown Formulation

Motivated by the implications in §2.2, we propose a two-stage formulation to resolve the tradeoff between optimization tractability and deployment feasibility. We first optimize the binary cell statuses to minimize energy consumption (Stage 1), and map these optimal decisions to the heterogeneous thresholds required by operator control logic (Stage 2). Stage 2 is general to any solution that optimizes binary statuses, enabling real-world deployment by mapping such decisions into operator-compatible thresholds.

Stage 1: Cell On-Off Control. The first stage focuses on optimal cell status identification. The goal of Stage 1 is to directly determine the optimal binary operational status of each capacity cell to minimize the theoretical total network energy consumption. In this stage, the decision variable is the cell status of capacity cells at each

time interval. By denoting $\mathbf{X}^t_{\text{cap}} = \{x^t_i\}_{i\in C_{\text{cap}}}$ and $\mathbf{X}^t_{\text{cov}} = \{x^t_i\}_{i\in C_{\text{cov}}}$ ($\mathbf{X}^t = \mathbf{X}^t_{\text{cap}} \cup \mathbf{X}^t_{\text{cov}}$), the objective is formulated as:

$$\min_{\mathbf{X}^t_{\text{cap}}} \quad \sum_{t=1}^{T} E\big(\mathbf{X}^t_{\text{cap}} \cup \mathbf{X}^t_{\text{cov}}, \mathbf{L}^t\big). \tag{6}$$

The optimization must ensure that cell shutdown does not compromise service availability; that is, the total number of connected UEs in the network remains the same as when all cells are active:

$$\sum_{k\in C} \mu^t_k = \eta^t, \forall t = 1, ..., T, \tag{7}$$

where η^t is the total connected UE count with all cells active in interval t. This optimization is also subject to the cell capacity constraint that post-offloading traffic volume of each cell does not exceed its capacity limits $\rho^{\max}_{k,\text{UL}}$ and $\rho^{\max}_{k,\text{DL}}$, expressed by:

$$0 \le \rho^t_{k,\text{UL}} \le \rho^{\max}_{k,\text{UL}}, \quad 0 \le \rho^t_{k,\text{DL}} \le \rho^{\max}_{k,\text{DL}}, \quad \forall k \in C, \forall t. \tag{8}$$

Other feasible value ranges are as follows: $\mu^t_k \in \mathbb{Z}_{\ge 0}$; $w^t_{k,\text{UL}}, w^t_{k,\text{DL}} \in [0, 1]$; $x^t_k \in \{0, 1\}, \forall k \in C, \forall t = 1, ..., T$. Stage 1 is NP-hard even for one slot and one coverage cell, since it can be reduced to a 0–1 Knapsack instance by mapping each item to a capacity cell, its value to saved energy, and its weight to offloaded traffic capacity.

Stage 2: Threshold Mapping. The second stage focuses on threshold optimization. The goal of Stage 2 is to identify the optimal threshold $\boldsymbol{\Theta}$ within predefined feasible ranges such that the resulting threshold-based shutdown best approximates the optimal cell on-off status in each interval $\mathbf{X}^{t*}_{\text{cap}}, t = 1, ..., T$ obtained in Stage 1.

Given each paired capacity cell $i \in C_{\text{cap}}$ and coverage cell $j \in C_{\text{cov}}$, the threshold set consists of the switch-off threshold vector of capacity cell i, denoted by $\boldsymbol{\theta}_i$, and the switch-off threshold and switch-on threshold vectors of coverage cell j, denoted by $\boldsymbol{\theta}_j$ and $\boldsymbol{\beta}_j$, respectively. Therefore we have $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \boldsymbol{\beta}_j\}_{i\in C_{\text{cap}}, j\in C_{\text{cov}}}$. Each threshold vector includes three specific thresholds: UL PRB usage ratio threshold, DL PRB usage ratio threshold, and user count threshold. Specifically, we have $\boldsymbol{\theta}_i = [\theta^{\text{UL}}_i, \theta^{\text{DL}}_i, \theta^{\text{UE}}_i]^T$, $\boldsymbol{\theta}_j = [\theta^{\text{UL}}_j, \theta^{\text{DL}}_j, \theta^{\text{UE}}_j]^T$, $\boldsymbol{\beta}_j = [\beta^{\text{UL}}_j, \beta^{\text{DL}}_j, \beta^{\text{UE}}_j]^T, \forall i \in C_{\text{cap}}, j \in C_{\text{cov}}$.

Following the deployment-aligned threshold-based policy discussed in §2.2, the operating status x^t_i of capacity cell i is determined by its previous status x^{t-1}_i, the threshold set $\boldsymbol{\Theta}$, and the current cell loads ℓ^t_i and ℓ^t_j. For $\forall i \in C_{\text{cap}}, j \in C_{\text{cov}}$, we define this non-differentiable status mapping function as:

$$x^t_i\big(\boldsymbol{\Theta}, \ell^t_i, \ell^t_j, x^{t-1}_i\big) = \begin{cases} 1 - \mathbb{I}(\ell^t_i < \boldsymbol{\theta}_i \wedge \ell^t_j < \boldsymbol{\theta}_j), & x^{t-1}_i = 1, \\ 1 - \mathbb{I}(\ell^t_j < \boldsymbol{\beta}_j), & x^{t-1}_i = 0, \end{cases} \tag{9}$$

where the inequality between vectors is element-wise. The objective is to minimize the discrepancy between the status $x^t_i(\boldsymbol{\Theta}, \ell^t_i, \ell^t_j, x^{t-1}_i)$ induced by the threshold-based policy and the optimal status x^{t*}_i:

$$\min_{\boldsymbol{\Theta}} \sum_{i\in C_{\text{cap}}} \sum_{t=2}^{T} \Big(x^{t*}_i - x^t_i(\boldsymbol{\Theta}, \ell^t_i, \ell^t_j, x^{t-1}_i)\Big)^2, \tag{10}$$

where $\theta^{\text{DL}}_i, \theta^{\text{UL}}_i, \theta^{\text{DL}}_j, \theta^{\text{UL}}_j, \beta^{\text{DL}}_j, \beta^{\text{UL}}_j \in [0, 1]$ and $\theta^{\text{UE}}_i, \theta^{\text{UE}}_j, \beta^{\text{UE}}_j \ge 0$.

3.3 Design Challenges

Safely optimizing the trade-off between network energy savings and service quality in large-scale heterogeneous networks presents three major obstacles. We must first model a highly dynamic network environment with sufficient fidelity. Even with such a model, solving the Stage 1 and Stage 2 decision problems in §3.2 remains difficult because of the complexity of network dynamics.

C0: Characterizing high-fidelity heterogeneous large-scale network environment. Accurately modeling network environments is mathematically intractable; static models misestimate objectives and constraints and thus yield suboptimal policies. The core challenge is to capture three aspects of real-world dynamics. i) Statistical behavior of user association: the function $f(\cdot)$ is hard to model explicitly because user behavior is stochastic, with highly dynamic mobility, varying demand, and localized 3D radio propagation; ii) Complex inter-cell coordination: traffic offloading among heterogeneous cells induces intricate spatial-temporal dependencies driven by overlapping coverage and load balancing, making $g(\cdot)$ highly non-linear. Ignoring these behaviors miscalculates cell loads, skewing switch-on thresholds and constraint evaluation; iii) Cell energy profiles: cells exhibit substantial variance in their energy consumption profiles $P^{\text{cell}}_k(\cdot), \forall k$, so coarse scalar metrics such as average energy efficiency are inadequate. Capturing this power variance is essential for exploiting energy-saving potential.

C1: Resolving Stage 1 in an efficient and scalable way. Dynamic cell sleep control involves multi-step transitions, highly nonlinear traffic dynamics, and strong spatial coupling among cells. Traditional optimization struggles to adapt to real-time network volatility, and heuristic methods cannot adequately capture how joint cell actions affect network-wide energy consumption and constraints, often leading to suboptimal solutions. RL is attractive because it learns directly from environment interaction and optimizes long-term returns. However, directly applying standard or centralized RL to large-scale networks presents three major hurdles. First, state and action spaces grow exponentially with the number of cells, making policy learning computationally prohibitive. Second, standard RL struggles to capture the spatio-temporal patterns of cell collaboration, such as overlapping coverage and dynamic traffic offloading. Third, non-linear dependencies between local actions and global performance create severe learning ambiguity and training instability: individually feasible actions may become infeasible when combined, causing constraint violations.

C2: Resolving Stage 2 to make shutdown policy deployable. In real-world cellular networks, threshold-based shutdown is governed by the hard step-function status mapping in Eq. (9). Although the forward mapping from thresholds to binary status is explicit, it is non-differentiable, and the inverse calibration from target status to thresholds has no closed-form solution. Consequently, solving Stage 2 effectively remains challenging.

4 Twinergy

4.1 Design Overview

To solve the problem of two stages in §3.2, we propose Twinergy, a digital twin empowered practical framework for energy optimization in heterogeneous mobile networks. As shown in Fig. 2, Twinergy consists of three main components to overcome the aforementioned challenges: a network digital twin system, a cell status control agent, and a threshold mapping module.

The overall workflow of Twinergy is as follows. The network digital twin system first collects real-world data to construct a twin

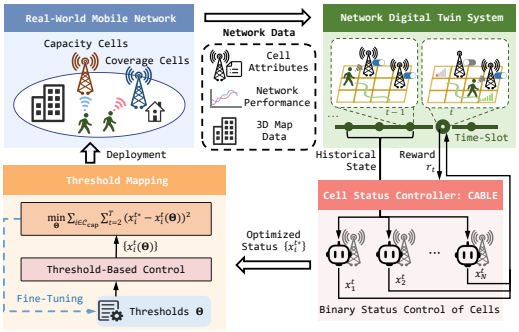


Figure 2: Overview of TWINERGY framework.

model of the heterogeneous network. This addresses the challenge of high-fidelity environment modeling (**C0**) by simulating stochastic user behavior, complex inter-cell offloading dependencies, and diverse energy profiles that static models fail to capture. It then evaluates energy consumption in this dynamic virtual environment, and feeds the resulting constraint violations to the Cell Status Agent module for energy-saving policy optimization.

To overcome the scalability bottlenecks and NP-hard nature of Stage 1 (**C1**), we propose CABLE, a scalable MARL algorithm that learns a unified spatial-temporal representation of cell state under varying local cooperative topologies. It captures both cell heterogeneity and interdependencies to make effective, decentralized on/off decisions for Stage 1. The resulting policies are iteratively fed back into the digital twin to evaluate their cooperative impact.

After multiple iterations, the optimized cell shutdown policy is passed to the Threshold Mapping module. This component resolves the deployment hurdle posed by the non-differentiable, strict rule-based mechanisms of real-world networks (**C2**). It converts the optimized cell-status policy into a practical, rule-compliant threshold-based policy for live deployment, thereby addressing the threshold optimization problem of Stage 2.

4.2 Network Digital Twin System

We present a comprehensive network digital twin that evaluates energy savings and constraint violations for cell shutdown policies. By capturing the complex dynamics of user association $f(\cdot)$ and traffic offloading $g(\cdot)$ defined in our system model, we construct a high-fidelity environment to address C0.

User connection modeling. To model the highly stochastic user association function of Eq. (1), we leverage a VAE [10] trained on real-world trajectories to generate realistic spatial user distributions $\mathbf{R}^t$. User channel conditions $\boldsymbol{\gamma}^t$ are derived via standard 3GPP TR 38.901 path loss models and 3D urban maps. The connection status $c^t_{u,k} \in \{0,1\}$ of user u to cell k depends on an active service demand, modeled as a Poisson process using real-world user count data, and channel conditions. Following the max-RSRP criterion, a user associates with the active cell providing the strongest signal above a minimum connection quality threshold. Consequently, the dynamic active user count is directly aggregated as: $\mu^t_k = \sum_{u\in\mathcal{U}} c^t_{u,k}$, s.t. $\sum_{k\in C} c^t_{u,k} \le 1$ inherently capturing the user coverage state without requiring complex intermediate modeling. Under dynamic cell shutdown policies, $c^t_{u,k}$ is re-evaluated based on the updated network status $\mathbf{X}^t$. The number of offloaded users is calculated by $\Delta\mu^t_{k\to n} = \sum_{u\in\mathcal{U}}[c^t_{u,n} \cdot c^{t-1}_{u,k}]$. This model enables the system to verify whether current connections meet the service level agreement η^t mandated by Eq. (7).

Inter-cell coordination. To capture the non-linear traffic offloading of Eq. (2), we combine spatial neighbor identification with Granger causality tests [7] strictly applied to historical traffic loads. By performing pairwise, bidirectional causality tests across overlapping cells, we construct a directed causal graph g^t_{UL} where edges represent temporal dependencies, weighted by the F-statistic $F_{k\to n}$. When a cell k is off at time t ($x^t_k = 0$), its users are offloaded to its active neighboring cells $\mathcal{N}_k$. The specific offloaded traffic volume $\Delta\rho^t_{k\to n,\mathrm{UL}}$ to an active neighbor $n \in \mathcal{N}_k$ is distributed proportionally using the causal influence weights derived from the graph:

$$\Delta\rho^t_{k\to n,\mathrm{UL}} = \left(\frac{\exp(F_{k\to n}\cdot\Delta\mu^t_{k\to n})}{\sum_{m\in\mathcal{N}_k}\exp(F_{k\to m}\cdot\Delta\mu^t_{k\to m})}\right)\cdot\rho^{t-1}_{k,\mathrm{UL}}. \quad (11)$$

The updated traffic volume for cell n aggregates its intrinsic baseline load with all offloaded traffic volume: $\rho^t_{n,\mathrm{UL}} = \rho^{t-1}_{n,\mathrm{UL}} + \sum_{m\in N_n: x^t_m=0}\Delta\rho^t_{m\to n,\mathrm{UL}}$. While a cell's PRB usage at the second-level timescale is heavily influenced by user channel conditions and the dynamically adapted coding scheme, analysis of real-world network data reveals that these variations are smoothed out by temporal averaging over the longer timescales relevant to cell shutdown decisions (e.g., 30 minutes) [13]. As a result, the average PRB usage ratio can be well approximated by the ratio of traffic volume to cell capacity. Based on this, the corresponding mapping $h(\cdot)$ is derived from this relationship: $w^t_{k,\mathrm{UL}} = h(\rho^t_{k,\mathrm{UL}}, \boldsymbol{\gamma}^t) \approx \rho^t_{k,\mathrm{UL}}/\rho^{\max}_{k,\mathrm{UL}}$. This modeling is symmetrically applied to the DL traffic load, utilizing the causal graph g^t_{DL}, which allows the system to verify the capacity limits in Eq. (8) directly over intervals.

Base station energy profile. The network energy consumption is driven by the dynamic power of Remote Radio Units (RRUs) and the static Baseband Unit (BBU) baseline power P^{base}_b. Based on observations from our real-world dataset, RRU energy consumption P^{cell}_k exhibits a strong linear correlation with the UL PRB usage ratio $w^t_{k,\mathrm{UL}}$, the DL PRB usage ratio $w^t_{k,\mathrm{DL}}$, and the active user count μ^t_k. To accurately reflect real-world energy characteristics and capture hardware-level heterogeneity, we build distinct linear regression models to define P^{on}_k for each cell k:

$$P^{\mathrm{on}}_k(\ell^t_k) = \alpha_{k,\mathrm{UL}}\cdot w^t_{k,\mathrm{UL}} + \alpha_{k,\mathrm{DL}}\cdot w^t_{k,\mathrm{DL}} + \alpha_{k,\mathrm{UE}}\cdot\mu^t_k + \gamma_k, \quad (12)$$

where $\alpha_{k,\mathrm{UL}}$, $\alpha_{k,\mathrm{DL}}$, and $\alpha_{k,\mathrm{UE}}$ are cell-specific energy coefficients for cell load at time t, respectively. The term γ_k represents the static power drawn during on status, ensuring $P^{\mathrm{off}}_k < \gamma_k$. Finally, the BBU contributes a nearly constant power consumption that remains independent of short-term traffic fluctuations, modeled by averaging historical data.

4.3 Cell Status Optimization: CABLE

To systematically address the complexities of applying RL to cell shutdown control (C1), CABLE introduces Credit-guided Agent-Based sleep controL of basE stations (CABLE), a scalable multi-agent reinforcement learning (MARL) framework. First, to overcome the exponential explosion of state and action spaces, we design a decentralized architecture with a shared actor network, reducing policy complexity from exponential to linear. Second, to capture the complex spatio-temporal patterns of cells' collaboration, we propose an embedding module driven by TGCN and Granger causal

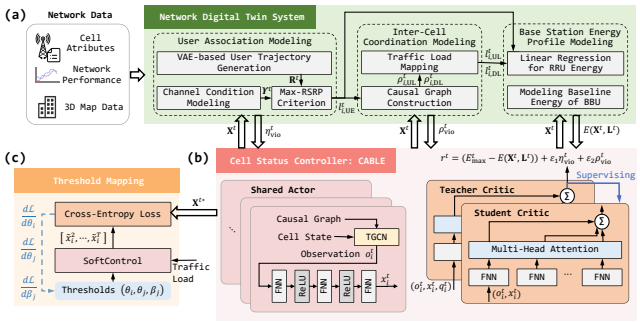


Figure 3: (a) Modules of the network digital twin system. (b) CABLE: Credit-guided Agent-Based sleep controL of basE stations. (c) Threshold Mapping module.

graphs, distilling massive network state into low-dimensional local observations. Finally, to resolve the training instability caused by the non-linear dependencies between local actions and global network performance, we introduce a credit-guided teacher-student dual-critic mechanism. This structure explicitly models the non-linear global impacts of individual cell decisions, ensuring stable convergence and highly coordinated multi-agent cooperation.

MARL Architecture. To overcome the exponential state and action space explosion inherent in centralized control, CABLE distributes policy execution into a MARL structure. Rather than maintaining distinct models across cell types or relying on intractable centralized policies, all agents utilize a shared actor network to map their local observations o_i^t to binary sleep decisions $x_i^t \in \{0, 1\}$, achieving robust scalability.

The agents collaboratively aim to maximize a global network-level reward r^t. To formalize this efficiently, we define the maximum baseline energy as $E_{\max}^t = E(\{x_k^t = 1\}_{k\in C}, \mathbf{L}^t)$. We quantify the user association constraint violation as user drop ratio $\eta_{\text{vio}}^t = \max(0, \eta^t - \sum_{k\in C} \mu_k^t)/\eta^t$ and the capacity constraint violation as traffic overload $\rho_{\text{vio}}^t = \sum_{k\in C}[\max(0, \rho_{k,\text{UL}}^t - \rho_{k,\text{UL}}^{\max}) + \max(0, \rho_{k,\text{DL}}^t - \rho_{k,\text{DL}}^{\max})]$. The scalar reward is cleanly defined to maximize energy savings while penalizing violations:

$$r^t = \left(E_{\max}^t - E(\mathbf{X}^t, \mathbf{L}^t)\right) - \varepsilon_1 \eta_{\text{vio}}^t - \varepsilon_2 \rho_{\text{vio}}^t, \tag{13}$$

where ε_1 and ε_2 are penalty scaling coefficients. We employ a centralized critic to estimate the global value function, which enables the computation of a global advantage for policy optimization. The shared actor is then trained using the Proximal Policy Optimization (PPO) [28], effectively decoupling policy complexity from the total number of cells.

TGCN-based State Embedding. Cells exhibit strong spatial coupling and temporal dynamics, making it difficult to capture spatio-temporal dependencies. CABLE resolves this by constructing highly informative, low-dimensional state representations leveraging the NDT system. The NDT provides cell-specific energy regression coefficients ($\alpha_{k,\text{DL}}$, $\alpha_{k,\text{UE}}$, and static power γ_k) alongside the base station baseline power P_b^{base}, imbuing each agent with explicit awareness of its energy efficiency. To capture network dynamics, we derive causal graphs g_{UL}^t and g_{DL}^t using Granger causality tests to model cell coordination explicitly. Taking the cell load vector ℓ_i^t, capacity $\rho_{i,\text{UL}}^{\max}$ and $\rho_{i,\text{DL}}^{\max}$, and these energy model coefficients as the local state of cell i, the input—alongside the causal graphs—is passed through a TGCN. The TGCN effectively extracts an embedding that captures both the cell's own state and the aggregated spatio-temporal representation of its neighboring cells. This resulting embedding is formally defined as the local observation o_i^t, enabling autonomous decision-making without requiring a massive global state.

Credit-Guided Teacher-Student Dual-Critic. Relying solely on a global scalar reward lacks modeling of multi-agent cooperation and often induces severe credit assignment ambiguity. The impact of cell actions on the global reward is non-additive and combinatorial, as non-linear cell collaboration, where individually acceptable actions may jointly cause constraint violations, complicates policy optimization. To resolve this, CABLE utilizes a novel teacher-student dual-critic architecture tailored for network energy saving. Empowered by a Multi-Head Attention (MHA) mechanism, this centralized critic structure enables the teacher network to explicitly learn the global impact of local action combinations, while the student network ensures the shared actor policy converges to global optima by mimicking the teacher network.

During interaction with the NDT, an explicit credit signal q_i^t for each cell i is computed by collecting: (i) the saved BS power P_b^{base} where the cell is served, (ii) user association drops, and (iii) traffic overload. The teacher's value estimation for cell i is defined as a localized attention mapping over observations of its own and neighbors: $C_{\phi_{\text{tea}}}(\{(o_i^t, x_i^t, q_i^t)\}_{i\in C}) = \sum_{i\in C} \text{MHA}\Big((o_i^t, x_i^t, q_i^t), \{(o_j^t, x_j^t, q_j^t)\}_{j\in\mathcal{N}_i}\Big)$. where the Query is derived from cell i's inputs, and the Keys and Values are derived from the local neighborhood set $\mathcal{N}_i$. These estimations are aggregated to produce the final global value, evaluated against the temporal global reward r^t: $\mathcal{L}_{\text{teacher}}(\phi_{\text{tea}}) = \mathbb{E}\left[\left(r^t - C_{\phi_{\text{tea}}}(\{(o_i^t, x_i^t, q_i^t)\}_{i\in C})\right)^2\right]$. A student critic $C_{\phi_{\text{stu}}}$ is then trained to approximate this teacher's highly detailed global value estimation without needing the explicit credit signal q_i^t as input during execution. The student network also uses the MHA mechanism and derives its Query, Key, and Value projections purely from the observation and action inputs of the cell and its neighbors. The student is optimized via $\mathcal{L}_{\text{student}}(\phi_{\text{stu}}) = \mathbb{E}\left[\left(C_{\phi_{\text{tea}}}(\{o_i^t, x_i^t, q_i^t\}_{i\in C}) - C_{\phi_{\text{stu}}}(\{o_i^t, x_i^t\}_{i\in C})\right)^2\right]$. By explicitly modeling these individual contributions through the teacher's guidance, this architecture drastically enhances multi-agent cooperation, resolves credit ambiguity, and stabilizes training.

Framework Computational Complexity and Scalability. CABLE bypasses the intractable $O(2^n)$ action space of centralized control, where n is the total number of cells. During training and execution, the computational complexity of the TGCN embedding and the shared actor scales linearly with the total number of cells as $O(n)$. By strictly limiting the attention calculation in both critics to local neighbor cells, the critic complexity is reduced from a global $O(n^2)$ to a localized $O(nk)$, where $k = |\mathcal{N}_i|$ is the neighbor size ($k \ll n$). This guarantees high computational efficiency for large-scale, ultra-dense network deployments.

4.4 Threshold Mapping

To address C2 in §3.3 and enable efficient threshold optimization for Stage 2 in §3.2, we establish a differentiable mapping from cell status optimization to thresholds $\boldsymbol{\Theta}$. Specifically, we design a gradient-continuous function, named *SoftControl*, to replace the hard status

mapping in Eq. (9). By utilizing smooth sigmoid functions, SoftControl produces a probabilistic output over the binary cell status, rendering the mechanism differentiable. The primary objective of this threshold mapping module, as illustrated in Fig. 3 (c), is to force the SoftControl output to closely approximate the optimal binary decisions x_i^{t*} obtained from the CABLE policy in Stage 1, effectively mimicking the reinforcement learning policy within a deployable rule set.

To enable explicit learning of these thresholds, we parameterize the SoftControl function such that the thresholds $(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \boldsymbol{\beta}_j)$ within $\boldsymbol{\Theta}$ are treated as learnable parameters. We first define a vector aggregation operator $Q : \mathbb{R}^D \to \mathbb{R}$ that computes the product of all elements in a D-dimensional vector $\mathbf{v} = [v_1, \ldots, v_D]$, such that $Q(\mathbf{v}) \triangleq \prod_{d=1}^{D} v_d$. The parameterized control policy for capacity cell i and its paired coverage cell j is then formulated as:

$$\begin{aligned} \tilde{x}_i^t &= \textbf{SoftControl}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \boldsymbol{\beta}_j, \ell_i^t, \ell_j^t, x_i^{t-1}) \\ &= \begin{cases} 1 - Q\Big(\sigma_k(\boldsymbol{\theta}_i - \ell_i^t)\Big) \cdot Q\Big(\sigma_k(\boldsymbol{\theta}_j - \ell_j^t)\Big), & x_i^{t-1} = 1, \\ 1 - Q\Big(\sigma_k(\boldsymbol{\beta}_j - \ell_j^t)\Big), & x_i^{t-1} = 0, \end{cases} \end{aligned} \tag{14}$$

where $\sigma_k(\mathbf{z})$ denotes the steep sigmoid function applied element-wise to vector $\mathbf{z}$, with k representing the steepness hyperparameter. This design smoothly approximates the strict network logic: when all load metrics in ℓ_i^t and ℓ_j^t are below their respective switch-off thresholds, the output approaches 0; when any metric exceeds its switch-on threshold $(\boldsymbol{\beta}_j)$, the output approaches 1. Because the output functions as a continuous probability distribution, we adopt a standard cross-entropy loss to minimize the divergence between the soft predictions and the Stage 1 decisions x_i^{t*} across intervals. Using this fully differentiable formulation, the threshold set $\boldsymbol{\Theta}$ is explicitly optimized via backpropagation until convergence, ensuring a smooth, computationally efficient alignment from the CABLE status control policy to deployable network thresholds.

5 Evaluation in Network Digital Twin

5.1 Operator Datasets and Baselines

We evaluate TWINERGY using a one-week real-world dataset in City A (July 2023) collected from a commercial operator, comprising 14,349 BSs with 30-minute performance logs. Based on this dataset, we construct our NDT. We train the TWINERGY on 11,425 BSs and evaluate it on the 20% held-out BSs (including 7,185 4G cells and 4,261 5G cells); each cell contains more than 15 shutdown events on average, enabling evaluation beyond ordinary observational traces. We first validate the fidelity of the digital twin system, and then compare the performance of TWINERGY against baseline methods, followed by comprehensive ablation studies.

Baseline Methods. Given the comprehensive review of state-of-the-art prior work in §2, we benchmark TWINERGY against six latest and representative baselines, encompassing both threshold-based policies (CHP, HHTP, PPO4Param) and binary status optimization (REDEEM, PPO-EPO, DRAG). For fair comparison, binary status optimization baselines are augmented with our proposed threshold mapping module to output compatible threshold-based policies.

- **CHP [17].** Current Homogeneous Policy, the operator's homogeneous policy applies uniform thresholds for cell shutdown.
- **HHTP [32].** A heuristic that iteratively fine-tunes load thresholds for batches of high-energy cells until convergence.
- **PPO4Param [4].** A PPO-based state-of-the-art algorithm for end-to-end threshold optimization of cell shutdown.
- **REDEEM [18].** A heuristic partitioning coverage into meshes, greedily switching off cells based on energy efficiency.
- **PPO-EPO [23].** A state-of-the-art PPO framework that achieves energy efficiency while guaranteeing network constraints.
- **DRAG [31].** A DDPG approach integrating action refinement.

Implementation Details. We train the CABLE using standard PPO hyperparameters, including a learning rate of 3e-4, a discount factor of 0.99, a clip ratio of 0.2, and a batch size of 128. For the neural networks, the TGCN module comprises 3 layers with a hidden dimension of 128, and the MHA mechanism in the critic network employs 4 heads. To impose constraint penalties, the coefficients (ϵ_1, ϵ_2) are set to 10^6. Finally, the steepness parameter k in SoftControl is set to 20. Within the network digital twin, the minimum RSRP threshold for valid user connections is set to -110 dBm.

5.2 Fidelity of Network Digital Twin

User association and traffic load. Fig. 4 and Fig. 5 demonstrate the high-fidelity of the methods in both user association and traffic load for 4G and 5G networks. For user association, Our Model closely reproduces the ground truth and achieves a near-perfect correlation, significantly outperforming the Random Shuffle and Point-of-Interest (POI)-Markov method [9]. For DL PRB usage modeling in Fig. 5, the CDF curve of our model closely tracks the ground truth, accurately reproducing the DL PRB usage per cell, while demonstrating a much higher and more consistent correlation coefficient than the Flow-Level [11] and POI-Based GNN [14].

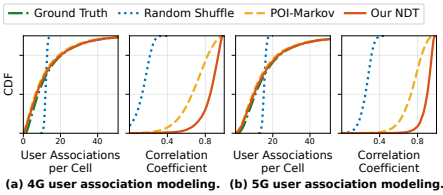
Energy consumption. Fig. 6 present the CDF of the correlation coefficient. The neural network [26] exhibits noticeable bias in 5G scenarios, failing to capture heterogeneous energy profiles and producing overly smoothed predictions. Both methods achieve comparable correlation performance, with our model slightly outperforming on 4G. Our linear model converges in a single pass over all cells, resulting in substantially lower training cost. In contrast, our model closely matches the ground truth distribution for both 4G and 5G, accurately preserving variability across cells.

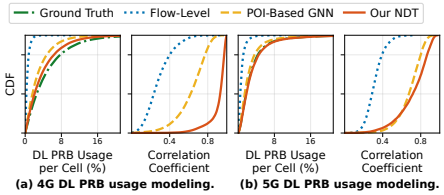
Counterfactual validity. We condition the evaluation on shutdown events to test the NDT under changed cell statuses. As shown in Fig. 7, temporal relative errors are within 5%, with MAPE below 3%; spatial DL-traffic RMSE remains below 20 MB for most cells; and the worst error over shutdown events remains below 5% for most cells, with only a few hotspot cells exceeding 8%.

5.3 Comparison with Baselines

Overall performance evaluation. We evaluate the performance of our proposed methods, CABLE and TWINERGY, against baselines. The primary evaluation metrics include 4G and 5G energy saving rates, as well as constraint violations: Connection Drop (Conn. Drop) and Traffic Overload (TOL) for both 4G and 5G. The constraint violations are measured in parts per million (ppm), representing the ratio of failed connections or overloaded traffic instances to total network data. The energy saving rate is defined as:

$$\delta = \frac{\sum_t \left[E\big(\{x_i^t = 1 \mid i \in C\}, \mathbf{L}^t\big) - E\big(\mathbf{X}^t, \mathbf{L}^t\big)\right]}{\sum_t E\big(\{x_i^t = 1 \mid i \in C\}, \mathbf{L}^t\big)} \times 100\%. \tag{15}$$





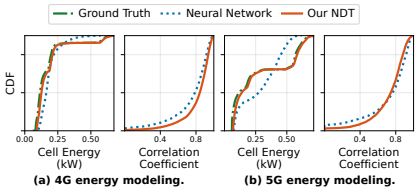
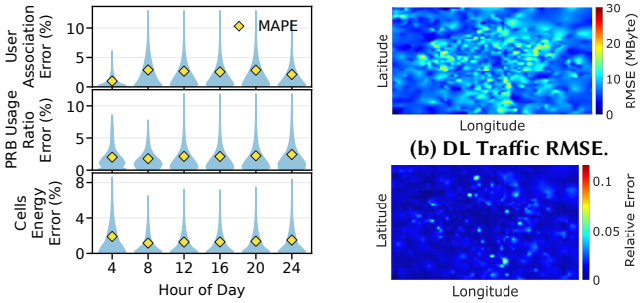


Figure 4: User association modeling fidelity. **Figure 5: DL PRB usage modeling fidelity.** **Figure 6: BS energy modeling fidelity.**



(a) Relative error. (b) DL Traffic RMSE. (c) Worst case (DL traffic).

Figure 7: Validation of NDT under shutdown events.

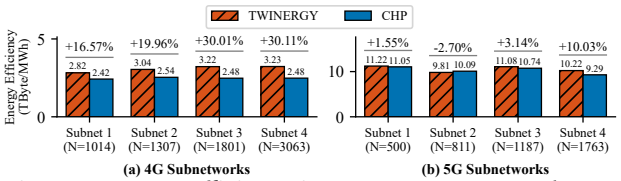


Figure 8: Energy efficiency improvement across subnets.

Table 2a details the cell shutdown performance using binary status control. CABLE significantly outperforms the baseline methods, achieving the highest energy saving rates of 41.5% for 4G and 19.3% for 5G. CABLE utilizes TGCN to effectively model inter-cell collaboration. This ensures low constraint violations while safely unlocking a significantly larger energy-saving margin compared to the baseline methods. Because 5G cells possess much larger capacities, offloading their traffic to the 4G tier significantly increases the risk of 4G TOL. Baselines like REDEEM experience remarkably high 4G TOL because they fail to account for this asymmetric collaboration. Benefiting from the credit assignment mechanism, CABLE successfully learns the distinct collaborative policies required between 4G and 5G cells. This allows it to achieve an optimal trade-off between maximizing energy savings and minimizing TOL.

Table 2b presents the performance using threshold-based policies. We integrate our TM module into these baselines to enable a fair comparison. After applying TM module, all evaluated baselines achieve effective savings with minimal constraint violations and surpass CHP in 4G networks, though 5G performance remains limited. TWINERGY achieves the highest savings, reaching 26.0% for 4G and 12.4% for 5G, while perfectly mitigating 5G traffic overload. Although the conservative CHP yields marginally lower Connection Drops, TWINERGY delivers nearly four times the 4G energy savings with constraints that remain well within acceptable operational limits. These energy savings translate into substantial operational benefits, including an estimated annual OPEX saving of approximately 6.25 million in City A [30], along with a reduction of about 48,706 tons of CO_2e emissions per year [15]. The threshold profile in Table 2b is cell-specific and applied over the evaluation horizon. Table 2c evaluates remapping at shorter intervals: more frequent remapping progressively closes this temporal-mismatch gap, reaching the binary savings at 30-minute granularity.

Energy efficiency improvement across network scale. To evaluate the adaptability of our method under varying topologies, we analyze the energy efficiency, defined as traffic volume served per unit of energy consumed, across different network densities. The test dataset region of City A is divided into four equal-area subnetworks, with the number of cells (N) varying to represent different levels of density, as shown in Fig. 8. We evaluate our approach against CHP to provide a direct perspective on the practical gains, since CHP is the currently deployed policy by operators.

TWINERGY consistently outperforms the CHP baseline in 4G subnetworks, with gains scaling alongside network density, from +16.57% to +30.11%. Higher density inherently creates more potential savings from overlapping coverage and capacity redundancy during low-load periods. TWINERGY exploits this redundancy by capturing inter-cell coordination to safely shut down cells. 5G subnetworks demonstrate a much higher baseline efficiency but narrower optimization margins. TWINERGY's relative performance ranges from a slight -2.70% degradation to a +10.03% improvement. The degradation in subnet1 and subnet2 is mainly caused by sparse 5G deployment: limited overlap among neighboring cells leaves little room for smooth offloading, so switch-off actions are more likely to violate constraints. TWINERGY therefore learns a more conservative policy in these areas to preserve constraints, while CHP applies a homogeneous rule and ignores the local inter-cell coordination. As a result, most 5G TOL under CHP originates from these low-density subnetworks. In denser 5G deployments, TWINERGY can better exploit redundancy to improve efficiency while mitigating overload. This comparison (TWINERGY versus CHP) also maintains consistency with the deployment in §6, highlighting the minimal sim-to-real gap across NDT and real-world environment.

5.4 Ablation Study

All proposed modules are necessary for CABLE. To isolate the contributions of our proposed framework, we conduct an ablation study by comparing the full model against three degraded variants: (1) CABLE w/o hetero, which ignores energy heterogeneity and only takes traffic volume as state; (2) CABLE w/o inter-cell, which disables TGCN and causal graph; and (3) CABLE w/o credit, which removes teacher critic and cell credit. The results are shown in Fig. 9. The full CABLE achieves the highest reward and energy savings. The ablated models suffer significant performance drops:CABLE w/o hetero is unstable and plateaus at low reward ($\approx$ 2.5), resulting in the lowest 4G savings (14.5%) and minimal 5G savings. This demonstrates that homogeneous strategies struggle to converge in diverse environments. The 5G savings CABLE of w/o inter-cell collapse to 1.6%. Given 5G's reliance on 4G coverage, ignoring

Table 2: Overall performance of cell shutdown policies and threshold-update analysis. ppm is parts per million.

(a) Cell shutdown with binary status control.

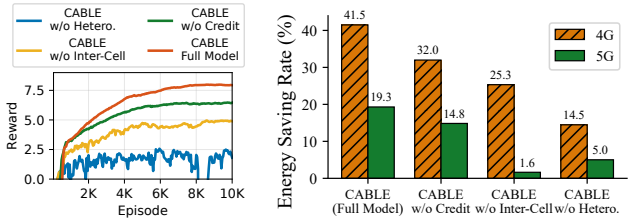
Methods	4G Energy Saving (%)	5G Energy Saving (%)	Conn. Drop (ppm)	4G TOL (ppm)	5G TOL (ppm)
REDEEM	8.0	6.0	175	1392	57
DRAG	11.3	4.4	67	**166**	84
PPO-EPO	12.4	5.4	42	243	12
Cable	**41.5**	**19.3**	**40**	180	**7**

(b) Cell shutdown with threshold-based policies.

Methods	4G Energy Saving (%)	5G Energy Saving (%)	Conn. Drop (ppm)	4G TOL (ppm)	5G TOL (ppm)
CHP	6.6	5.8	**30**	297	98
REDEEM w/ TM	6.9	4.1	149	697	12
DRAG w/ TM	7.8	2.7	47	**109**	37
PPO-EPO w/ TM	8.6	3.9	33	215	0
HHTP	8.7	11.2	99	198	138
PPO4Param	11.8	11.3	34	144	132
Twinergy	**26.0**	**12.4**	31	127	**0**

(c) Threshold adaptation.

Adaptation	4G (%)	5G (%)
Static	26.0	12.4
Every 24 Hours	28.1	13.1
Every 8 Hours	31.2	14.1
Every 2 Hours	37.8	17.5
Every 30 Min	41.5	19.3
Cable (Binary)	41.5	19.3



(a) Reward. **(b) Energy saving.**

Figure 9: Ablation study for Cable.

Table 3: Deployment settings of Twinergy across trials.

	Year-Month	Scenario	# of Cells 4G	# of Cells 5G	Time
Trial 1	2024-10	Rural	12	13	0:00-7:00
Trial 2	2025-04	Urban	234	113	Full-day
Trial 3	2025-07	Dense 5G	N/A	3,049	Full-day

inter-cell collaboration leads to association-constraint violations and inefficient convergence. Energy savings of Cable w/o credit degrade proportionally across both tiers (32.0% for 4G, 14.8% for 5G). Joint agent actions make global reward signals too noisy; removing the credit assignment module prevents the system from making confident decisions necessary for optimal convergence.

Hyperparameter Sensitivity. The reward coefficients (ϵ_1, ϵ_2) balance energy savings and constraints: small values lead to constraint violations, while large values produce conservative policies with limited savings. In the TM module, the steepness parameter k bridges continuous approximation and deployable logic. Small k causes mapping errors, while large k causes gradient instability. We set $k = 20$ to balance fast convergence and accurate rule translation.

6 Evaluation in Real-World Deployment

Collaborating with the operator, we deploy Twinergy across three real-world networks in City B. We implement Twinergy using data collected from these networks and derive a threshold-based policy for each capacity cell. The details are summarized in Table 3. We compare Twinergy with the production CHP across trials with varying traffic loads and cell densities. Operator constraints limit live comparisons to only CHP. Each trial uses the same infrastructure in adjacent weeks with comparable traffic volumes. Trial 3 controls 5G only, while 4G remains active for coverage.

Twinergy achieves substantial improvement in energy efficiency. After deploying Twinergy, network energy efficiency improved across all trials, as shown in Table 4, with varying gains reflecting the characteristics of each scenario. In Trial 1, operating during nighttime with sparse rural traffic, Twinergy capitalizes on temporal inactivity by aggressively putting redundant cells to sleep, yielding a 13.07% efficiency gain for 4G. In the full-day urban center of Trial 2, cells must remain active more frequently to handle continuous high traffic. Consequently, the gains are more modest but still notable, at 8.18% for 4G and 3.42% for 5G. Finally, in the dense, full-day 5G deployment of Trial 3, Twinergy exploits heavily overlapping coverage to unlock a substantial 12.20% gain. This trend is consistent with the network-scale evaluation in Fig. 8, where higher cell density creates larger optimization margins through coverage overlap and capacity redundancy. The smaller 5G gain in Trial 2 also aligns with the earlier observation that 5G networks have higher baseline efficiency but narrower safe shutdown margins under sustained traffic, whereas dense 5G deployments provide more redundancy for Twinergy to exploit without degrading QoS.

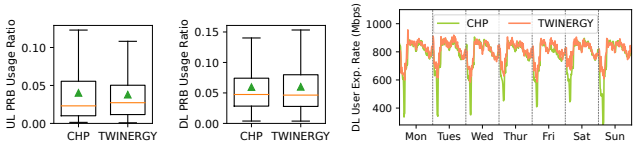
Table 4: Energy efficiency (EE) and QoS across trials. Gains are shown in parentheses.

		4G EE (TByte/MWh)	5G EE (TByte/MWh)	Conn. Succ. Rate	UL User Exp. Rate (Mbps)	DL User Exp. Rate (Mbps)
Trial 1	CHP	2.004	6.184	96.11%	7.72	146.99
	Twinergy	2.266 (+13.07%)	6.414 (+3.71%)	96.01% (-0.10%)	7.85 (+1.68%)	148.59 (+1.09%)
Trial 2	CHP	2.128	8.982	98.88%	17.66	240.75
	Twinergy	2.302 (+8.18%)	9.289 (+3.42%)	98.89% (+0.01%)	17.86 (+1.13%)	246.50 (+2.39%)
Trial 3	CHP	–	10.469	96.021%	31.44	152.90
	Twinergy	–	11.746 (+12.20%)	96.021% (+0.00%)	32.46 (+3.24%)	153.77 (+0.57%)

Twinergy maintains QoS. Table 4 summarizes QoS across trials, Twinergy does not degrade user experience. Network reliability remains stable, with the connection success rate fluctuating by less than 0.1% across all trials. The lower tail is also stable: in Trial 1, q1 changes from 95.46% to 95.01% and q5 from 95.90% to 95.76%, a worst decrease of 0.45 percentage points; q1 in Trials 2 and 3 changes from 97.11% to 97.43% and from 94.66% to 94.69%, respectively. Furthermore, both UL and DL user experience rates actually have slight improvements, especially in high-rate scenarios.

The strong consistency between simulated and live trials. EE gains in Table 4 align with our NDT results in Fig. 8, demonstrating a minimal sim-to-real gap influenced by network density. For 4G, Trial 1's rural deployment yields a 13.07% improvement, which is slightly below the 16.57% in NDT. For 5G, Trial 3's dense environment achieves 12.20%, exceeding the 10.03% in NDT. These live trials validate that Twinergy consistently captures density-driven efficiency improvements across diverse network scales.

Computational cost of deployment. In practice, operators deploy Twinergy by building the digital twin from recent network data, training the policy, translating it into thresholds, and then configuring the resulting thresholds in the live network. The current system has been deployed and is running in a real-world network for over one month. The deployment is not one-shot static: periodic retraining is needed as 5G deployment and usage evolve and seasonal traffic patterns change, but the retraining cycle is on the order of months. Most computational overhead of each retraining is concentrated in the cold-start or retraining stage, including RSRP calculation, trajectory generation, and critic network training. Twinergy can train models for networks with thousands of cells in under 20 hours on a consumer-grade GPU (e.g., RTX 3070), and



(a) UL PRB usage. (b) DL PRB usage. (c) User exp. rate (5G) of trial 2.

Figure 10: Real-world impacts of deploying TWINERGY.

operators can further reduce this cost by excluding stable, always-on capacity cells. In practice, this total overhead is even lower than manual verification and threshold configuration by engineers.

TWINERGY provides practical insights. A primary operator concern is that overly optimistic policies induce excessive offloading, leading to overloads and congestion. Fig. 10a and Fig. 10b illustrate that PRB usage under TWINERGY aligns with the CHP. This stems from TWINERGY's accurate modeling of inter-cell coordination and offloading, which yields precise constraints modelling before decision-making. Furthermore, elucidating the enhanced experience rate observed in Table 4, Fig. 10c details the 5G network DL rate in Trial 2. During nighttime ultra-low traffic periods, CHP deactivates excessive cells, causing degradation. TWINERGY preserves sufficient active cells to guarantee QoS, while leveraging cell energy profiles to shut down high-consumption cells.

7 Conclusions

In this paper, we presented TWINERGY, a digital twin-driven framework designed to resolve the fundamental trade-off between maximizing energy efficiency and preserving network reliability in heterogeneous mobile networks. The proposed reinforcement learning algorithm captures intricate inter-cell coordination via a TGCN and resolves multi-agent cooperation through a credit-guided critic, unlocking significantly larger energy-saving margins under strict constraints. Subsequently, our differentiable SoftControl mapping seamlessly translates these optimal binary decisions into deployable threshold configurations. Extensive digital twin evaluations validate that TWINERGY significantly outperforms existing heuristics and baseline algorithms. Most importantly, deployments in live commercial networks demonstrate substantial and scalable energy efficiency gains, achieving up to 13.07% and 12.20% improvements in rural and dense 5G scenarios, respectively, without degrading user experience, establishing TWINERGY as a highly practical, deployment-ready solution for sustainable radio access networks.

Acknowledgments

Yibo Ma's funding for this research was provided by EPSRC through a PhD studentship within the CDT in Machine Learning Systems hosted in the School of Informatics, University of Edinburgh (EP/Y03516X/1). The work of M. Marina and T. Wang was supported by the UKRI/EPSRC grant UKRI860.

References

[1] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz. Technical Report TR 38.901 V19.3.0, 2026. Release 19.

[2] M. Belgiovine, C. Dick, and K. Chowdhury. Better together: Leveraging multiple digital twins for deployment optimization of airborne base stations. *IEEE Trans. Mobile Comput.*, 2025.

[3] Y. L. Che, L. Duan, and R. Zhang. Dynamic base station operation in large-scale green cellular networks. *IEEE J. Sel. Areas Commun.*, 34(12):3127–3141, 2016.

[4] M. Choi, K. Kim, H. Jang, H. Woo, J. S. Pujol-Roig, Y. Wang, H. Yeon, S. Choi, and S. Jang. Cell on/off parameter optimization for saving energy via reinforcement learning. In *IEEE GLOBECOM Wkshps.*, pages 1–6, 2021.

[5] H. Fourati, R. Maaloul, L. Fourati, and M. Jmaiel. An efficient energy-saving scheme using genetic algorithm for 5G heterogeneous networks. *IEEE Syst. J.*, 17(1):589–600, 2023.

[6] C. Ge, Z. Ge, X. Liu, A. Mahimkar, Y. Shaqalle, Y. Xiang, and S. Pathak. Chroma: Learning and using network contexts to reinforce performance improving configurations. In *Proc. ACM MobiCom*, pages 1–16, 2023.

[7] R. Hoffmann, C.-G. Lee, B. Ramasamy, and M. Yeung. FDI and pollution: a granger causality test using panel data. *J. Int. Dev.*, 17(3):311–317, 2005.

[8] C.-L. I, S. Han, and S. Bian. Energy-efficient 5G for a greener future. *Nat. Electron.*, 3(4):182–184, 2020.

[9] R. Jiang, X. Song, Z. Fan, T. Xia, Z. Wang, Q. Chen, Z. Cai, and R. Shibasaki. Transfer urban human mobility via POI embedding over multiple cities. *ACM Trans. Data Sci.*, 2(1):1–26, 2021.

[10] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[11] H. Klessig, D. Oehmann, A. J. Fehske, and G. P. Fettweis. A performance evaluation framework for interference-coupled cellular data networks. *IEEE Trans. Wireless Commun.*, 15(2):938–950, 2015.

[12] T. Li, Q. Long, H. Chai, S. Zhang, F. Jiang, H. Liu, W. Huang, D. Jin, and Y. Li. Generative AI empowered network digital twins: Architecture, technologies, and applications. *ACM Comput. Surv.*, 57(6):1–43, 2025.

[13] T. Li, L. Yu, Y. Ma, T. Duan, W. Huang, Y. Zhou, D. Jin, Y. Li, and T. Jiang. Carbon emissions of 5G mobile networks in china. *Nat. Sustain.*, 6(12):1620–1631, 2023.

[14] J. Lin, Y. Chen, H. Zheng, M. Ding, P. Cheng, and L. Hanzo. A data-driven base station sleeping strategy based on traffic prediction. *IEEE Trans. Netw. Sci. Eng.*, 11(6):5627–5643, 2021.

[15] Z. Liu, D. Guan, W. Wei, S. J. Davis, P. Ciais, J. Bai, S. Peng, Q. Zhang, K. Hubacek, G. Marland, et al. Reduced carbon emission estimates from fossil fuel combustion and cement production in china. *Nature*, 524(7565):335–338, 2015.

[16] J. X. S. Lozano, J. A. Ayala-Romero, A. Garcia-Saavedra, and X. Costa-Perez. Kairos: Energy-efficient radio unit control for O-RAN via advanced sleep modes. In *IEEE INFOCOM*, pages 1–10, 2025.

[17] Y. Ma, T. Li, Y. Du, S. Dustdar, Z. Wang, and Y. Li. Sustainable connections: Exploring energy efficiency in 5G networks. In *Proc. ACM CoNEXT*, pages 33–40, 2024.

[18] Y. Ma, T. Li, Y. Zhou, L. Yu, and D. Jin. Mitigating energy consumption in heterogeneous mobile networks through data-driven optimization. *IEEE Trans. Netw. Serv. Manag.*, 21(4):4369–4382, 2024.

[19] L. Maggi, C. Mihailescu, Q. Cao, A. Tetich, S. Khan, S. Aaltonen, R. Koblitz, M. Holma, S. Macchi, M. E. Ruggieri, et al. Energy savings under performance constraints via carrier shutdown with bayesian learning. In *EuCNC/6G Summit*, pages 1–6, 2023.

[20] S. S. Mwanje and J. Ali-Tolppa. Distributed energy saving management in multi-layer 4G/5G ultra-dense networks. In *IEEE WCNCW*, pages 143–148, 2018.

[21] H. Nan, R. Li, X. Zhu, J. Ma, and D. Niyato. An efficient data-driven traffic prediction framework for network digital twin. *IEEE Netw.*, 38(1):22–29, 2024.

[22] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Virdis. Simu5G–an OMNeT++ library for end-to-end performance evaluation of 5G networks. *IEEE Access*, 8:181176–181191, 2020.

[23] R. Ntassah, G. M. Dell'Aera, and F. Granelli. PPO-EPO: Energy and performance optimization for O-RAN using reinforcement learning. In *IEEE ICC Wkshps.*, pages 1189–1195, 2025.

[24] NVIDIA. Aerial omniverse digital twin, 2025. Accessed: 2026-04-21.

[25] T. J. O'shea, T. Roy, and N. West. Approximating the void: Learning stochastic channel models from observation with variational generative adversarial networks. In *ICNC*, pages 681–686, 2019.

[26] N. Piovesan, D. López-Pérez, A. De Domenico, X. Geng, H. Bao, and M. Debbah. Machine learning and analytical power consumption models for 5G base stations. *IEEE Commun. Mag.*, 60(10):56–62, 2022.

[27] G. Punz, D. Mur, and K. Samdanis. Energy saving standardisation in mobile and wireless communication systems. *Green Commun.: Princ. Concepts Pract.*, pages 237–256, 2015.

[28] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[29] D. Wu, Y. T. Xu, M. Jenkin, S. Jang, E. Hossain, X. Liu, and G. Dudek. Energy saving in cellular wireless networks via transfer deep reinforcement learning. In *IEEE GLOBECOM*, pages 7019–7024, 2023.

[30] T. Xiao and Q. Zhou. China's industrial power rates 2025: A guide for investors, May 2025. Accessed: 2026-04-21.

[31] J. Ye and Y.-J. A. Zhang. Drag: Deep reinforcement learning based base station activation in heterogeneous networks. *IEEE Trans. Mobile Comput.*, 19(9):2076–2087, 2019.

[32] G. Yu, Q. Chen, and R. Yin. Dual-threshold sleep mode control scheme for small cells. *IET Commun.*, 8(11):2008–2016, 2014.

[33] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.*, 21(9):3848–3858, 2019.