A Multi-Task Reinforcement Learning Approach for Optimal Sizing and Energy Management of Hybrid Electric Storage Systems Under Spatio-Temporal Urban Rail Traffic

Guannan Li, *Student Member, IEEE,* Siu Wing Or, *Senior Member, IEEE*

***Abstract*—Passenger flow fluctuation and delay-induced traffic regulation bring considerable challenges to cost-efficient regenerative braking energy utilization of hybrid electric storage systems (HESSs) in urban rail traction networks. This paper proposes a synergistic HESS sizing and energy management optimization framework based on multi-task reinforcement learning (MTRL) for enhancing the economic operation of HESSs under dynamic spatio-temporal urban rail traffic. The configuration-specific HESS control problem under various spatio-temporal traction load distributions is formulated as a multi-task Markov decision process (MTMDP), and an iterative sizing optimization approach considering daily service patterns is devised to minimize the HESS life cycle cost (LCC). Then, a dynamic traffic model composed of a Copula-based passenger flow generation method and a real-time timetable rescheduling algorithm incorporating a traction energy-passenger-time sensitivity matrix is developed to characterize multi-train traction load uncertainty. Furthermore, an MTRL algorithm based on a dueling double deep Q network with knowledge transfer is proposed to simultaneously learn a generalized control policy from annealing task-specific agents and operation environments for solving the MTMDP effectively. Comparative studies based on a real-world subway have validated the effectiveness of the proposed framework for LCC reduction of HESS operation under urban rail traffic.**

***Index Terms*—Hybrid electric storage systems, multi-task learning, optimal sizing and energy management, reinforcement learning, urban rail transits.**

I. Introduction

A. Motivation

With the expansion of urban rail transportation systems to provide high-frequency and green mass travel services [1], periodic and short-distanced train acceleration has resulted in significant traction energy demand, while its deceleration produces considerable regenerative braking energy (RBE) to urban rail traction networks (URTNs) [2], [3]. Although the RBE can be utilized by nearby trains in traction for energy saving, excessive RBE can be wasted, leading to voltage fluctuation and heat management issues [4]. Driven by the transition to a carbon-neutral energy-transportation nexus [5], the stationary hybrid electric storage system (HESS) has been widely considered a potential buffer hub [6] for RBE recovery and consumption management. Recently, the increasing traction energy consumption induced by the rapid growth of passenger demand has underscored the necessity of developing effective HESS configuration and energy management technologies for improving the energy and cost efficiencies of URTNs.

Nevertheless, the inherent dynamic energy interactions among the HESS, traction substations, and multiple trains in different operation states inevitably result in the volatility of traction loads [7]. Moreover, the URTN operation is susceptible to multi-source external uncertainties, such as the daily passenger flow fluctuation and varying line conditions affected by weather, which results in the high nonlinearity of train and network parameters [8]. The temporary traffic regulation caused by stochastic short delays has especially strengthened the spatio-temporal uncertainty of traction load distribution. Furthermore, the spatio-temporal coupling of multi-station passengers adds substantial complexity to the HESS real-time control strategy. Therefore, it is crucial to investigate an optimal HESS sizing and energy management framework considering the dynamic traffic of urban rails for enhancing the economic HESS operation.

B. Background

Extensive efforts have been devoted to the energy management of railway electric storage systems (ESSs) composed of a single storage medium in URTNs. The optimal operation of such ESS can be determined by train and traction network parameters. For instance, in [9], an ESS power management system was developed for urban light railways based on train speed and acceleration estimations to recover maximum RBE. In [10], the optimal stationary ESS control was realized based on catenary voltage. Besides, the ESS parameters, such as the state-of-charge, can also be utilized for optimal switching of working modes [11]. Recently, in order to address the growing electricity consumption, the HESS has been developed to utilize the complementary nature of the high energy density of batteries and the high power density of supercapacitors for increasing energy efficiency. Compared with ESSs, HESSs possess unique control topologies of bidirectional DC-DC converters [12], enabling efficient power allocation and branch current estimation [13] among multiple energy sources. Various studies have investigated optimal HESS power allocation

This work was supported in part by the Research Grants Council of the HKSAR Government under Grant No. R5020-18, and in part by the Innovation and Technology Commission of the HKSAR Government to the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center under Grant No. K-BBY1. (*Corresponding author: Siu Wing Or.)*

Guannan Li and Siu Wing Or are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China, and also with the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center, Hong Kong, China. (emails: 21037965r@connect.polyu.hk, eeswor@polyu.edu.hk).

strategies to mitigate battery aging and achieve economic operation. In [14], an adaptive HESS energy management framework is proposed based on fuzzy logic control. However, considering real-time dynamic traffic conditions, inevitable train operation deviations, and coordination of energy sources, the optimal control rules are difficult to set. In [15], a model predictive control framework was proposed to optimize train and HESS operation to reduce energy consumption. In [16], a two-stage hierarchical model was established to optimize supercapacitor capacity, train operation diagrams, and URTN parameters collaboratively. In [17], a bi-level multi-objective optimization was performed considering substation operation stability. However, these approaches rely on accurate uncertainty forecasting and can be time-consuming addressing highly non-linear URTN parameters. Meanwhile, the voltage thresholds and power allocations of HESSs have not been jointly optimized online, which decreases the cost-saving effect.

DRL [18] has been recognized as a promising solution for high-dimensional sequential decision-making problems through iterative interactions with the environment. The well-trained DRL agent can execute high-quality decisions in milliseconds [7]. Despite advanced DRL algorithms, such as DQN [7], TD3 [19], and VDN [20], have been applied for ESS or HESS control in URTNs, existing works are limited to learn individual strategies for each specific train headway task with fixed train speed profiles. Even if multiple such tasks are involved, the inter-task coupling relationship is ignored, and a large amount of interaction data is required. Since modern train operations adopt different headways and a set of pre-programmed speed profiles [21] to handle peak and off-peak hours, it is necessary to learn a generalized control policy of HESS across multiple tasks to adapt to daily operation requirements. Furthermore, considering the same URTN topology, it is beneficial to learn multiple tasks simultaneously to leverage shareable experience to improve DRL performance and data efficiency [22].

So far, passenger-related and delay-related restrictions have been involved in speed profile design [23], rescheduling plans [24], and rolling stock circulation [25]. For instance, considering passenger-induced RBE difference, a mixed-integer programming model was formulated to optimize speed profile selection for timetable rescheduling [26]. Nevertheless, few studies considered the impacts of passenger flows and temporary traffic regulations on HESS control. For ESSs, a collaborative ESS and train operation mode was proposed in [27] to minimize passenger travel time, RBE waste, and ESS installation positions, while passenger and delay uncertainties were not considered. In [28], a traffic model considering uncertain dwell time and traffic regulation was developed for economic ESS installation. However, the optimal ESS control strategy and passenger uncertainty were neglected.

C. Contribution

Although various methods have been successfully applied to HESS sizing and energy management in URTNs, only few studies [16], [17] focused on synergistic optimization of HESS sizing and real-time control, and none of them considered coordination of voltage threshold adjustments and power allocations, which significantly undermines the cost-effectiveness. Regarding the dynamic traffic and timely decision-making of HESS control, the optimal rules of [9]–[11], [14] can be difficult to set. Besides, the optimization-based strategies [15]–[17] can be time-consuming and rely on accurate uncertainty forecasting. Although DRL-based strategies [7], [19], [20] were employed to address these issues, previous efforts were made to learn individual strategies for a few fixed train headway and mass sets, neglecting shareable experience among different train service patterns. Furthermore, a thorough investigation of the intense multi-source real-time traffic uncertainties that affect HESS operation in URTNs is urgently needed.

In this paper, an MTRL-based HESS sizing and energy management optimization framework considering dynamic spatio-temporal traffic characteristics of urban rails is proposed for the coordinated operation of HESS and the traction substation. The pivotal contributions of this paper are summarized as follows:

1) Compared with [7], [9]–[11], [14], [15], [19], [20] which only focus on energy management strategies, a synergistic sizing and energy management optimization framework is proposed for enhancing the economic HESS operation. An iterative sizing optimization approach considering daily service patterns is devised to minimize the HESS life cycle cost (LCC), and the configuration-specific HESS energy management problem under various spatio-temporal traffic uncertainties is modeled as a multi-task Markov decision process (MTMDP), where voltage thresholds and power allocations of the HESS are coordinated to further reduce the operation cost.
2) Compared with [27], [28], a dynamic traffic model (DTM) comprehensively considering passenger flow fluctuations and delay-induced traffic regulations is formulated to characterize multi-train traction load uncertainty for enhancing HESS energy management decisions. A Copula-based passenger flow scenario generation method is proposed to capture dependencies between multi-station origin-destination (OD) demands. A real-time timetable rescheduling (RTTR) algorithm incorporating the traction energy-passenger-time (TEPT) sensitivity matrix is developed to optimize the energy-efficient rescheduled timetable and train speed profiles under uncertain delays.
3) Compared with single-task DRL-based methods [7], [19], [20], an MTRL algorithm based on a dueling double deep Q network with knowledge transfer (KT-D3QN) is presented for solving the MTMDP effectively. A policy distillation annealing method is developed to learn a generalized multi-task HESS control policy simultaneously and stably from task-specific agents and dynamic train operation environments. Soft modulation and gradient manipulation techniques are employed to handle inter-task parameter sharing and conflicts.

The rest of the paper is organized as follows. Section II illustrates the URTN structure and problem formulation.

Section III presents the proposed synergistic optimization framework. Section IV reports case studies and their results. Section V gives the conclusions.

II. URTN Structure and Problem Formulation

A. URTN Structure & Modeling

The studied URTN is composed of substations, trains, 750/1500 V catenary system, rails, and a HESS (see Fig. 1). The substation contains a unidirectional 24-pulse wave diode rectifier. The HESS connects to the substation by DC-DC converters. Generally, the passenger flow prediction is conducted at a large time interval (e.g., 15 min), while the HESS control is carried out at a small time interval (e.g., 1 s). Thus, we use $n = 1, 2, \cdots, N$ to denote the "prediction interval", and $t = 1, 2, \cdots, T$ to denote the "control interval". The following assumptions are considered: *a)* The OD matrix is deterministic, while the passenger arrival rate varies [27]. *b)* The train stops at each station with a pre-determined dwell time and running time. While delays extend the planned dwell time, the total running and dwell times are unchanged [29].

In a scenario ρ, $P^{\mathrm{SUB}}_{i,\rho,t}$ is the power of ith substation, $1 \le i \le I$, $P^{\mathrm{SC,CH}}_{\rho,t}$ and $P^{\mathrm{SC,DIS}}_{\rho,t}$ are the discharging and charging power of the supercapacitor, respectively, $P^{\mathrm{BT,CH}}_{\rho,t}$ and $P^{\mathrm{BT,DIS}}_{\rho,t}$ are the discharging and charging power of the battery, respectively, $P^{\mathrm{TR}}_{k,\rho,t}$ is the power of kth train, $1 \le k \le K$. When a delay time $T^{\mathrm{D}}_{i,\rho}$ is known, the planned running time $T^{\mathrm{PL}}_{i,j,\rho}$ is changed to $T^{\mathrm{RTTR}}_{i,j,\rho}$ by RTTR, and an optimal speed profile is selected from a pre-programmed speed profile set.

The equivalent circuit model of the URTN (see Fig. 2) includes substation (1), train (2), battery (3)-(4), and supercapacitor (5)-(6) models. R^{P} is the pantograph resistance. $X^{\mathrm{L}}_{k,i,\rho,t}$ and $X^{\mathrm{L}}_{k,i+1\rho,t}$ are the distance to station i and $i+1$, respectively, R^{L} is the line resistance per km. $U^{\mathrm{SUB}}_{i,\rho,t}$ and $I^{\mathrm{SUB}}_{i,\rho,t}$ are the substation voltage and current, respectively, U^{SUB}_0 is the no-load substation voltage, R^{SUB} is the substation resistance, $U^{\mathrm{TR}}_{k,\rho,t}$ and $I^{\mathrm{TR}}_{k,\rho,t}$ are the train voltage and current, respectively. $\eta^{\mathrm{BR}}_{k,\rho,t}$ simulates the braking resistor [1], which determines the proportion of braking power delivered to the network. The supercapacitor is a resistor R^{SC} with a capacitor C^{SC}. $U^{\mathrm{SC}}_{\rho,t}$ and $I^{\mathrm{SC}}_{\rho,t}$ are its output voltage and current, respectively, $U^{\mathrm{C}}_{\rho,t}$ is the capacitor voltage. The battery is an open-circuit voltage source $U^{\mathrm{OCV}}_{\rho,t}$ with a resistor $R^{\mathrm{BT}}_{\rho,t}$ [30]. $U^{\mathrm{BT}}_{\rho,t}$ and $I^{\mathrm{BT}}_{\rho,t}$ are its output voltage and current, respectively. η^{SC} and η^{BT} are the efficiency of the supercapacitor and battery, respectively, Δt is the increment of the control interval.

$$U^{\mathrm{SUB}}_{i,\rho,t} = U^{\mathrm{SUB}}_0 - R^{\mathrm{SUB}} I^{\mathrm{SUB}}_{i,\rho,t}, \quad I^{\mathrm{SUB}}_{i,\rho,t} \ge 0, \tag{1}$$

$$U^{\mathrm{TR}}_{k,\rho,t} I^{\mathrm{TR}}_{k,\rho,t} = \eta^{\mathrm{BR}}_{k,\rho,t} P^{\mathrm{TR}}_{k,\rho,t}, \tag{2}$$

$$I^{\mathrm{BT}}_{\rho,t} = \begin{cases} P^{\mathrm{BT}}_{\rho,t}/\eta^{\mathrm{BT}} U^{\mathrm{BT}}_{\rho,t}, & P^{\mathrm{BT}}_{\rho,t} = P^{\mathrm{BT,DIS}}_{\rho,t}, \\ \eta^{\mathrm{BT}} P^{\mathrm{BT}}_{\rho,t}/U^{\mathrm{BT}}_{\rho,t}, & P^{\mathrm{BT}}_{\rho,t} = P^{\mathrm{BT,CH}}_{\rho,t}, \end{cases} \tag{3}$$

$$U^{\mathrm{BT}}_{\rho,t} = U^{\mathrm{OCV}}_{\rho,t} - I^{\mathrm{BT}}_{\rho,t} R^{\mathrm{BT}}_{\rho,t}, \tag{4}$$

$$U^{\mathrm{SC}}_{\rho,t} = U^{\mathrm{C}}_{\rho,t} - I^{\mathrm{SC}}_{\rho,t} R^{\mathrm{SC}}, \; U^{\mathrm{C}}_{\rho,t} = U^{\mathrm{C}}_{\rho,t-1} - I^{\mathrm{SC}}_{\rho,t} \Delta t / C^{\mathrm{SC}}, \tag{5}$$

$$I^{\mathrm{SC}}_{\rho,t} = \begin{cases} \dfrac{U^{\mathrm{C}}_{\rho,t} - \sqrt{\left(U^{\mathrm{C}}_{\rho,t}\right)^2 - 4R^{\mathrm{SC}} P^{\mathrm{SC}}_{\rho,t}/\eta^{\mathrm{SC}}}}{2R^{\mathrm{SC}}}, & P^{\mathrm{SC}}_{\rho,t} = P^{\mathrm{SC,DIS}}_{\rho,t}, \\ -\dfrac{U^{\mathrm{C}}_{\rho,t} - \sqrt{\left(U^{\mathrm{C}}_{\rho,t}\right)^2 - 4R^{\mathrm{SC}} \eta^{\mathrm{SC}} P^{\mathrm{SC}}_{\rho,t}}}{2R^{\mathrm{SC}}}, & P^{\mathrm{SC}}_{\rho,t} = P^{\mathrm{SC,CH}}_{\rho,t}. \end{cases} \tag{6}$$

B. Problem Formulation

1) Energy Management: The energy management is subject to URTN power flows, where $\boldsymbol{Y}$ is the admittance matrix, $c^{\mathrm{OM}}_{\mathrm{SC}}$ and $c^{\mathrm{OM}}_{\mathrm{BT}}$ are the supercapacitor and battery operation cost per MWh, respectively, c_{GRID} is the trading cost per kWh, $J^{\mathrm{GRID}}_{\rho,t}$ is the electricity trading cost, $J^{\mathrm{OM}}_{\rho,t}$ is the HESS operation cost,

$$\boldsymbol{Y}\left[\boldsymbol{U}^{l} \;\; \boldsymbol{U}^{\boldsymbol{s}}\right]^T = \left[\boldsymbol{I}^{\boldsymbol{t}} \;\; \boldsymbol{I}^{\boldsymbol{s}}\right]^T, \tag{7}$$

$$\begin{aligned} \boldsymbol{U}^{\boldsymbol{t}} &= \left[U^{\mathrm{L}}_{1,\rho,t} \cdots U^{\mathrm{L}}_{K,\rho,t}\right]^T, \; \boldsymbol{I}^{\boldsymbol{t}} = \left[I^{\mathrm{TR}}_{1,\rho,t} \cdots I^{\mathrm{TR}}_{K,\rho,t}\right]^T, \\ \boldsymbol{U}^{\boldsymbol{s}} &= \left[U^{\mathrm{SUB}}_{1,\rho,t} \cdots U^{\mathrm{SUB}}_{I,\rho,t}\right]^T, \; \boldsymbol{I}^{\boldsymbol{s}} = \left[I^{\mathrm{S}}_{1,\rho,t} \cdots I^{\mathrm{S}}_{I,\rho,t}\right]^T, \end{aligned} \tag{8}$$

$$U^{\mathrm{L}}_{k,\rho,t} = U^{\mathrm{TR}}_{k,\rho,t} - R^{\mathrm{P}} I^{\mathrm{TR}}_{k,\rho,t}, \tag{9}$$

$$I^{\mathrm{S}}_{i,\rho,t} = I^{\mathrm{SUB}}_{i,\rho,t} + \left(P^{\mathrm{SC}}_{\rho,t} + P^{\mathrm{BT}}_{\rho,t}\right)/U^{\mathrm{SUB}}_{i,\rho,t}, \tag{10}$$

$$J^{\mathrm{GRID}}_{\rho,t} = \sum\nolimits_{i=1}^{I} c_{\mathrm{GRID}} P^{\mathrm{SUB}}_{i,\rho,t} \Delta t, \tag{11}$$

$$J^{\mathrm{OM}}_{\rho,t} = c^{\mathrm{OM}}_{\mathrm{SC}} \left|P^{\mathrm{SC}}_{\rho,t}\right| \Delta t + c^{\mathrm{OM}}_{\mathrm{BT}} \left|P^{\mathrm{BT}}_{\rho,t}\right| \Delta t. \tag{12}$$

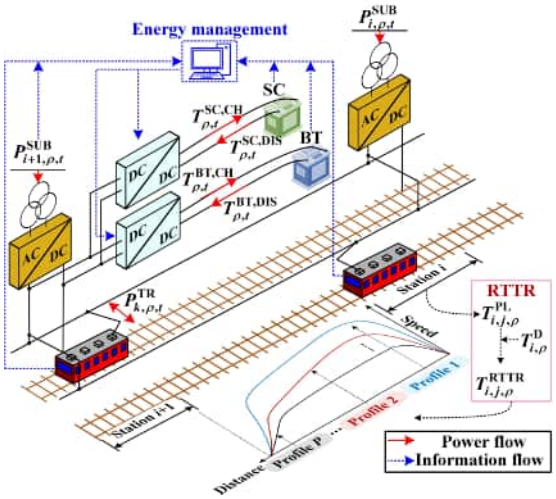

Fig. 1. The structure of URTN with a HESS.

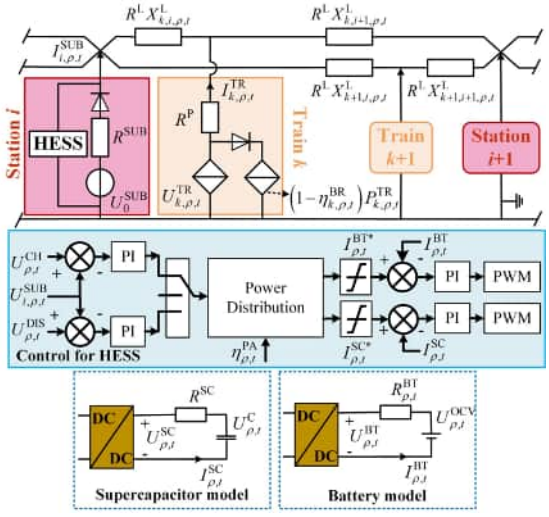

Fig. 2. Equivalent circuit model of URTN with a HESS.

The aim is to optimize real-time HESS voltage thresholds and power allocations for overall operation cost minimization,

$$\min J_\rho = \sum\nolimits_{t=1}^{T} \left(J_{\rho,t}^{\text{GRID}} + J_{\rho,t}^{\text{OM}}\right), \quad \text{s.t. (1)-(12).} \tag{13}$$

2) Sizing Optimization: Considering the match of converters and HESSs, the number of battery and supercapacitor modules in series is fixed. Hence, the sizing optimization problem aims to find the optimal number of battery and supercapacitor modules in parallel.

$$J^{\text{INV}} = \left[c_{\text{SC}}^{\text{INV}} N_{\text{SC}} + c_{\text{BT}}^{\text{INV}} N_{\text{BT}} + c_{\text{DC}}^{\text{INV}} N_{\text{HESS}}^{\text{DC}}\right] \eta^{\text{CR}}, \tag{14}$$

$$J^{\text{REP}} = \sum\nolimits_{m=1}^{N^{\text{REP}}} \frac{c_{\text{BT}}^{\text{INV}} N_{\text{BT}} + c_{\text{DC}}^{\text{INV}} N_{\text{BT}}^{\text{DC}}}{(1+I^R)^{mL_{\text{BT}}}} \eta^{\text{CR}}, \tag{15}$$

$$\eta^{\text{CR}} = I^R\left(1+I^R\right)^L / \left[\left(1+I^R\right)^L + 1\right], \tag{16}$$

where J^{INV} and J^{REP} are the investment and replacement cost, respectively, η^{CR} is the capital recovery factor, $c_{\text{SC}}^{\text{INV}}$, $c_{\text{BT}}^{\text{INV}}$, and $c_{\text{DC}}^{\text{INV}}$ are the investment cost of supercapacitor, battery, and converter per module, respectively, N_{SC} and N_{BT} are the number of supercapacitor and battery modules, respectively, $N_{\text{HESS}}^{\text{DC}}$ is the number of converter modules for HESS, N^{REP} is the replacement frequency, L_{BT} is the estimated battery life, L is the system lifetime, I^R is the interest rate.

The rainflow counting method [30] is utilized to quantify the cycle life loss of the battery. As the supercapacitor generally has a much longer cycle life than the battery, its replacement during system lifetime is ignored.

$$N^{\text{REP}} = \left\lceil \frac{L}{L_{\text{BT}}} - 1 \right\rceil, \; L_{\text{BT}} = \min\left(L, \frac{C_{\text{BT}}^{\text{Norm}}}{365 \cdot C_{\text{BT}}}\right), \tag{17}$$

where $C_{\text{BT}}^{\text{Norm}}$ is the amount of available life cycles, C_{BT} is the number of cycles counted per day.

The objective under scenario probability f_ρ is

$$\min J^{\text{LCC}} = \sum_{\rho} f_\rho J_\rho + J^{\text{INV}} + J^{\text{REP}} + J^{\text{FIX}} \eta^{\text{CR}}, \quad \text{s.t. (13)-(17).} \tag{18}$$

where J^{FIX} is the construction and installation cost.

III. MTRL-Based Synergistic HESS Sizing and Energy Management Optimization Framework

A. Overview of Optimization Process

The proposed framework (see Fig. 3) contains the following steps: 1) Various traction load scenarios are randomly generated based on the DTM, and representative daily traction load scenarios are selected based on clustering algorithms. 2) A HESS size is selected from the size constraint set, and such a HESS energy management problem is reformulated as an MTMDP. 3) The proposed KT-D3QN algorithm solves the MTMDP and trains an intelligent agent for multi-task HESS control. 4) Based on daily service patterns, an LCC analysis is performed, where the daily operation cost is calculated by the well-trained agent. 5) Repeat 2)-4) to traverse all sizes, and the optimal HESS size and energy management strategy is determined with the lowest LCC.

B. Dynamic Traffic Model

1) Copula-Based Passenger Flow Scenario Generation: The spatio-temporal uncertainty of passenger flows is quantified based on the Copula theory. A short description of the scenario generation process is given below, which is based on the thorough explanations provided by [1] (see algorithm 1). First, we estimate the historical passenger data $N_{i,\rho,n}^{\text{B}}$ according to OD and arrival rate tables, where $N_{i,\rho,n}^{\text{B}}$ is the average onboard passengers between station i and $i+1$ at prediction interval n (see (19)-(24)). Then, the temporal correlations of passengers are simplified, where only the temporal correlation between two consecutive time steps is considered. Then, the joint conditional cumulative distribution function (CDF) of passengers in two consecutive time steps is modeled by the conditional Copula function. Finally, multiple pseudo-observations are drawn from the conditional CDF to generate sufficient scenarios.

$$N_{k,i,j,\rho}^{\text{W}} = H_k \alpha_{i,j} \beta_{i,n}, \tag{19}$$

$$N_{k,i,\rho}^{\text{W}} = \begin{cases} \sum\limits_{j=i+1}^{I} N_{k,i,j,\rho}^{\text{W}}, & k=1, \\ N_{k-1,i,\rho}^{\text{W}} - N_{k-1,i,\rho}^{\text{ON}} + \sum\limits_{j=i+1}^{I} N_{k,i,j,\rho}^{\text{W}}, & k \neq 1, \end{cases} \tag{20}$$

$$N_{k,i,\rho}^{\text{ON}} = \begin{cases} \min(N_{k,i,\rho}^{\text{W}}, N_{\max}^{\text{B}}), & i=1, \\ \min(N_{k,i,\rho}^{\text{W}}, N_{\max}^{\text{B}} - \sum\limits_{j=1}^{i-1} N_{k,j,\rho}^{\text{ON}} + \sum\limits_{j=2}^{i} N_{k,j,\rho}^{\text{OFF}}), & i \neq 1, \end{cases} \tag{21}$$

$$N_{k,j,\rho}^{\text{OFF}} = \sum_{i=1}^{j-1} N_{k,i,j,\rho}^{\text{ON}} = \sum_{i=1}^{j-1} \frac{N_{k,i,j,\rho}^{\text{W}}}{N_{k,i,\rho}^{\text{W}}} N_{k,i,\rho}^{\text{ON}}, \tag{22}$$

$$N_{k,i,\rho}^{\text{B}} = \begin{cases} N_{k,i,\rho}^{\text{ON}}, & i=1, \\ N_{k,i-1,\rho}^{\text{B}} + N_{k,i,\rho}^{\text{ON}} - N_{k,i,\rho}^{\text{OFF}}, & i \neq 1, \end{cases} \tag{23}$$

$$N_{i,\rho,n}^{\text{B}} = \frac{1}{K_n} \sum_{k=1}^{K_n} N_{k,i,\rho}^{\text{B}}, \tag{24}$$

where the total waiting passengers $N_{k,i,\rho}^{\text{W}}$ at station i is calculated by (19)-(20), H_k is the headway of train k, $\alpha_{i,j}$ is the OD element, $\beta_{i,n}$ is the arrival rate, $N_{k,i,j,\rho}^{\text{W}}$ is the proportion of $N_{k,i,\rho}^{\text{W}}$ who travel from station i to j. The passengers getting on $N_{k,i,\rho}^{\text{ON}}$ and off $N_{k,i,\rho}^{\text{OFF}}$ is calculated by (21)-(22). $N_{\max}^{\text{B}}$ is the maximum train capacity. The onboard passengers $N_{k,i,\rho}^{\text{B}}$ and $N_{i,\rho,n}^{\text{B}}$ can be obtained by (23)-(24). K_n is the number of trains running at interval n.

2) TEPT-Based RTTR & Speed Profile Optimization: By dividing $N_{i,\rho,n}^{\text{B}}$ into D intervals, the TEPT sensitivity θ^{S} can be written as

$$\theta^{\text{S}} = \begin{bmatrix} \theta_1^{\text{S}} \\ \theta_2^{\text{S}} \\ \vdots \\ \theta_D^{\text{S}} \end{bmatrix}, \theta_d^{\text{S}} = \begin{bmatrix} \theta_{d,\delta_{1,1}}^{\text{S}} & \theta_{d,\delta_{1,2}}^{\text{S}} & \cdots & \theta_{d,\delta_{1,I-1}}^{\text{S}} \\ \theta_{d,\delta_{2,1}}^{\text{S}} & \theta_{d,\delta_{2,2}}^{\text{S}} & \cdots & \theta_{d,\delta_{2,I-1}}^{\text{S}} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{d,\delta_{P-1,1}}^{\text{S}} & \theta_{d,\delta_{P-1,2}}^{\text{S}} & \cdots & \theta_{d,\delta_{P-1,I-1}}^{\text{S}} \end{bmatrix}, \tag{25}$$

$$\theta_{d,\delta_{p,i}}^{\text{S}} = \frac{\Delta E_{d,\delta_{p,i}}}{\Delta T}, \; d = [1,2,\cdots,D], \, i \neq I, p \neq P,$$

where each running section has P pre-programmed speed profiles, $1 \leq p \leq P$, and the speed profiles can be ranked from the highest traction energy consumption to the lowest. $\delta_{p,i}$ denotes the pth of section $(i, i+1)$. $\theta^{\mathrm{S}}_{d,\delta_{p,i}}$ and $\Delta E_{d,\delta_{p,i}}$ are the sensitivity and energy difference between the pth and $p+1$th speed profile and between ith and $i+1$th station at passenger interval d. ΔT is the increment of running time.

A TEPT-based RTTR algorithm can then be developed to obtain the rescheduled timetable and speed profiles after the delay (see algorithm 1). The principle is that, at each iteration, we allocate a ΔT to the section with the lowest sensitivity. Then, the speed profile $\delta_{p,i}$ and sensitivity $\theta^{\mathrm{S}}_{d,\delta_{p,i}}$ for that section are updated, while $T^{\mathrm{D}}_{i,\rho} \leftarrow T^{\mathrm{D}}_{i,\rho} - \Delta T$. The allocation process ends when $T^{\mathrm{D}}_{i,\rho} = 0$. Thus, the rescheduled timetable and speed profiles of each section are determined.

3) Traction Load Calculation: The power of train k is (26), where M^{TR} and M^{P} are the total vehicle mass and passenger mass per person, respectively, $v^{\mathrm{TR}}_{k,\rho,t}$ and $a^{\mathrm{TR}}_{k,\rho,t}$ are the train speed and acceleration, respectively, η^{TR} is the motor efficiency. Since total resistance $F^{\mathrm{TR}}_{k,\rho,t}$ can be uncertain due to weather and line conditions, it is subject to a truncated Normal distribution $\mathcal{N}(\cdot)$ [31], where the standard deviation is 5% of the mean value, and its variance is limited to 10%.

$$P^{\mathrm{TR}}_{k,\rho,t} = \left[\left(M^{\mathrm{TR}} + N^{\mathrm{B}}_{i,\rho,n} M^{\mathrm{P}}\right) a^{\mathrm{TR}}_{k,\rho,t} - \mathcal{N}\left(F^{\mathrm{TR}}_{k,\rho,t}\right)\right] \frac{v^{\mathrm{TR}}_{k,\rho,t}}{\eta^{\mathrm{TR}}}. \quad (26)$$

C. MTMDP Reformulation

1) Task Representation & MTMDP Principle: Since different headways and rescheduled speed profiles significantly change the spatio-temporal distribution of traction loads, each headway and each combination of speed profiles between different stations is a specific task. The task set is $\mathcal{Z} = \{\mathcal{Z}_H, \mathcal{Z}_P\}$, where $\mathcal{Z}_H = \{z_1, \cdots, z_H\}$ contains H headway tasks in one-hot vectors, and $\mathcal{Z}_P = \{\{\delta_{1,1}, \cdots, \delta_{1,I-1}\}, \cdots, \{\delta_{P-1,1}, \cdots, \delta_{P-1,I-1}\}\}$ contains $(P-1)(I-1)$ combination tasks of speed profiles. Hence, the total number of tasks is $H(P-1)(I-1)$. Each task $z \in \mathcal{Z}$ can be formulated as a Markov decision process, and multiple tasks form a MTMDP with components $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma, \mathcal{Z} \rangle$. At time t, the agent with a state $s_t \in \mathcal{S}$ selects an action

Algorithm 1: Passenger flow scenario generation and TEPT-based RTTR in DTM

```
// Passenger flow scenario generation
Input: OD and arrival rate α, β
For prediction interval n = 1, N do
    Update arrival rate β_{i,n} and operation data H_k, K_n
    Estimate historical passengers N^B_{i,ρ,n} by (19)-(24)
Establish the joint conditional CDF by conditional Copula functions, draw multiple pseudo-observations from it to generate sufficient scenarios
Output: onboard passenger N^B
// TEPT-based RTTR
Input: passenger N^B, speed profile δ, delay time T^D, and running time T^PL
Initialize TEPT matrix θ^S
Initialize rescheduling: allocate T^D_{i,ρ} to section (i, i+1), T^D_{i,ρ} ← T^D_{i,ρ} − (p−1)ΔT. Then allocate T^D_{i,ρ} to section (i+1, i+2). Repeat the process till T^D_{i,ρ} = 0. Calculate the total energy consumption E_0 = Σ_{i=1}^{I} E_{d,δ_{p,i}}
While T^D_{i,ρ} ≠ 0 do
    Select [θ^S_{d,δ_{p,1}}, ··· , θ^S_{d,δ_{p,I−1}}] according to interval d and speed profile δ_{p,1}, ··· , δ_{p,I−1}
    Allocate ΔT to the target section with arg min (θ^S_{d,δ_{p,1}}, ··· , θ^S_{d,δ_{p,I−1}}), update the speed profile and sensitivity of the target section
    T^D_{i,ρ} ← T^D_{i,ρ} − ΔT
With the updated speed profiles δ'_{p,1}, ··· , δ'_{p,I−1}, calculate E'_0 = Σ_{i=1}^{I} E_{d,δ'_{p,i}} and compare with E_0, select the timetable with lowest energy consumption
Output: rescheduled time T^RTTR
```

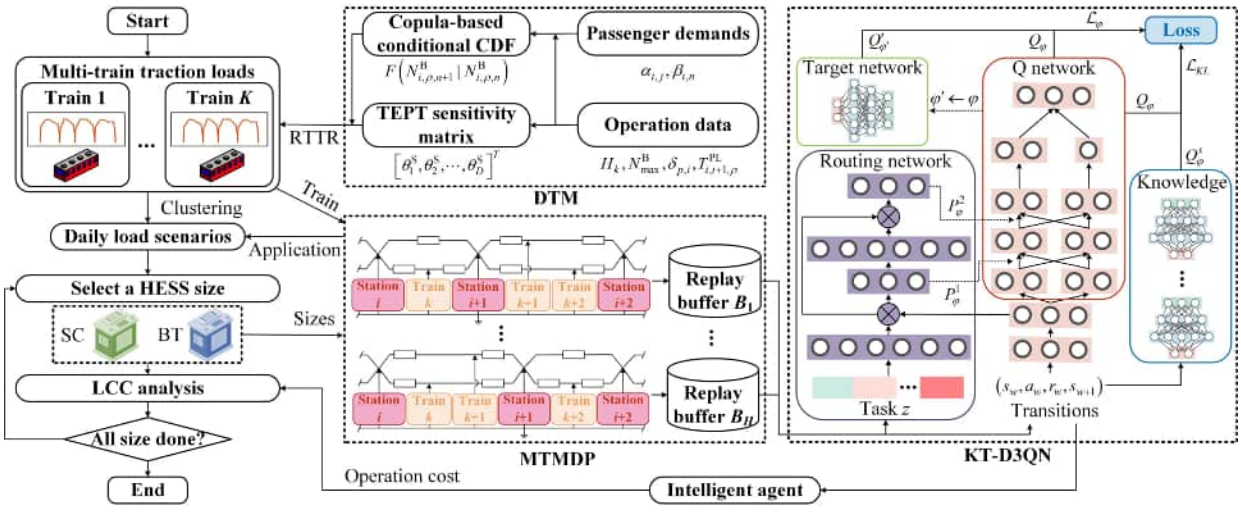

Fig. 3. Framework of MTRL-based synergistic HESS sizing and energy management optimization.

Algorithm 2: KT-D3QN

Randomly initialize Q_φ with φ and routing network with φ_r. Initialize $Q'_{\varphi'}$ with $\varphi' \leftarrow \varphi$, replay buffers $B_1, \cdots, B_H$ and the environment
For *episode = 1, Max* **do**
  Receive the initial state s_0
  For *control interval t = 1, T* **do**
    Select a_t with $\pi(a_t|s_t, z)$ and ϵ-greedy, obtain r_t and s_{t+1} in the environment
    Store transition (s_t, a_t, r_t, s_{t+1}) to B_h based on headway task z_h
    Sample W transitions (s_w, a_w, r_w, s_{w+1}) from each B_h
    Calculate loss $\mathcal{L}$ by (31)-(33), update φ,φ_r,λ
    Soft update: $\varphi' \leftarrow \tau\varphi + (1-\tau)\varphi'$
Output: $Q(s, a)$

$a_t \in \mathcal{A}$ subject to its task-conditioned policy $\pi(a_t|s_t, z)$. Then, s_t transits to the next state s_{t+1} under the state transition $\mathcal{P}(s_{t+1}|s_t, a_t)$, and the agent is rewarded by $r_t \in \mathcal{R}$. The goal of the agent is to maximize the expected return Q^* with a discounted factor γ,

$$Q^*(s,a) = \mathbb{E}_{z\sim p(z)}\left[\mathbb{E}_{\substack{s_{t+1}\sim\mathcal{P}\\ a_t\sim\pi}}\left(\sum_{t=0}^{T}\gamma^t r_t(s_t,a_t)\right)\right]. \quad (27)$$

2) State, Action, & Reward: s_t contains two parts: a) local substation operation status, including supercapacitor state-of-energy (SoE) $\mathrm{SoE}^{\mathrm{SC}}_{\rho,t}$, battery SoE $\mathrm{SoE}^{\mathrm{BT}}_{\rho,t}$, local substation outputs $U^{\mathrm{SUB}}_{i,\rho,t}$ and $I^{\mathrm{SUB}}_{i,\rho,t}$ (suppose HESS is in station i), b) train operation status, including position $X^{\mathrm{TR}}_{1,\rho,t}, \cdots, X^{\mathrm{TR}}_{K,\rho,t}$ direction $D^{\mathrm{TR}}_{1,\rho,t}, \cdots, D^{\mathrm{TR}}_{K,\rho,t}$, and power $P^{\mathrm{TR}}_{1,\rho,t}, \cdots, P^{\mathrm{TR}}_{K,\rho,t}$.$a_t$ is the voltage thresholds $U^{\mathrm{CH}}_{\rho,t}$, $U^{\mathrm{DIS}}_{\rho,t}$ and power allocation $\eta^{\mathrm{PA}}_{\rho,t}$. $a_t = \{U^{\mathrm{CH}}_{\rho,t}, U^{\mathrm{DIS}}_{\rho,t}, \eta^{\mathrm{PA}}_{\rho,t}\}$. According to (13), r_t is first set as the minus of $J^{\mathrm{GRID}}_{\rho,t}$ and $J^{\mathrm{OM}}_{\rho,t}$, and then multiplies a constant for reward scaling [22], namely, $r_t = -\left(J^{\mathrm{GRID}}_{\rho,t} + J^{\mathrm{OM}}_{\rho,t}\right)/\eta^{\mathrm{CR}}$.

$$s_t = \left\{\begin{matrix}\mathrm{SoE}^{\mathrm{SC}}_{\rho,t}, \mathrm{SoE}^{\mathrm{BT}}_{\rho,t}, U^{\mathrm{SUB}}_{i,\rho,t}, I^{\mathrm{SUB}}_{i,\rho,t}, X^{\mathrm{TR}}_{1,\rho,t}, \cdots, X^{\mathrm{TR}}_{K,\rho,t},\\ D^{\mathrm{TR}}_{1,\rho,t}, \cdots, D^{\mathrm{TR}}_{K,\rho,t}, P^{\mathrm{TR}}_{1,\rho,t}, \cdots, P^{\mathrm{TR}}_{K,\rho,t}.\end{matrix}\right\} \quad (28)$$

3) State Transition: $\mathcal{P}$ is illustrated as follows. The train operation status is updated by the selected speed profile p, and local substation outputs are updated by power flow calculation according to a_t. The SoEs are updated by

$$\mathrm{SoE}^{\mathrm{SC}}_{\rho,t} = \left(\frac{U^{\mathrm{C}}_{\rho,t}}{U^{\mathrm{C,norm}}}\right)^2, \quad \mathrm{SoE}^{\mathrm{BT}}_{\rho,t} = \mathrm{SoE}^{\mathrm{BT}}_{\rho,t-1} - \frac{I^{\mathrm{BT}}_{\rho,t}\Delta t}{Q^{\mathrm{BT,norm}}}, \quad (29)$$

where $U^{\mathrm{C,norm}}$ is the nominal capacitor voltage, $Q^{\mathrm{BT,norm}}$ is the nominal battery capacity, $\mathrm{SoE}^{\mathrm{SC}}_{\rho,t}$ and $\mathrm{SoE}^{\mathrm{BT}}_{\rho,t}$ are subject to range $\left[\mathrm{SoE}^{\mathrm{SC}}_{\min}, \mathrm{SoE}^{\mathrm{SC}}_{\max}\right]$ and $\left[\mathrm{SoE}^{\mathrm{BT}}_{\min}, \mathrm{SoE}^{\mathrm{BT}}_{\max}\right]$, respectively.

D. Multi-Task Reinforcement Learning Algorithm

1) Dueling Double Deep Q-network: Since the complexity of calculating Q^*, D3QN [18] approximates Q^* by Q_φ with parameter φ, and Q_φ is decoupled with a value estimation $V(s_t)$ and an action advantage estimation $A(s_t, a_t)$. This dueling architecture enables the agent to learn independent state values, which is useful in states where the actions have no effect on the environment,

$$Q_\varphi(s_t,a_t) = V(s_t) + A(s_t,a_t) - \frac{\sum_{a_{t+1}\in\mathcal{A}} A(s_t,a_{t+1})}{|\mathcal{A}|}. \quad (30)$$

Totally H replay buffers are built for all headway tasks, and W transitions (s_w, a_w, r_w, s_{w+1}) are randomly sampled from each buffer for updating D3QN. The loss is

$$\mathcal{L}_\varphi = \frac{1}{HW}\sum_h\sum_w \left(y - Q_\varphi(s_w,a_w)\right)^2, \quad (31)$$

where $y = r_w + \gamma Q'_{\varphi'}(s_{w+1}, \arg\max_{a_{w+1}} Q_\varphi(s_{w+1}, a_{w+1}))$, $Q'_{\varphi'}$ is the target network, φ' is its parameter.

2) Knowledge Transfer & Policy Distillation Annealing: Considering the similarity of different speed profile tasks under a given headway, we develop a knowledge transfer method to rapidly and stably learn the multi-task policy incorporating common knowledge from task-specific agents by policy distillation. For each headway task with several sets of speed profile tasks, a single-task agent is first trained in a learning environment without delay. Then, the Kullback-Leibler divergence is adopted to measure the discrepancy between the policy distributions of single-task agents and the multi-task agent. An annealing strategy is utilized to gradually reduce the knowledge transfer for convergence.

$$\mathcal{L}_{KL} = \sum_h\sum_w \mathrm{softmax}\left(Q^s_\varphi\right)\ln\left(\frac{\mathrm{softmax}\left(Q^s_\varphi\right)}{\mathrm{softmax}\left(Q_\varphi\right)}\right), \quad (32)$$

$$\mathcal{L} = (1-\lambda)\mathcal{L}_\varphi + \lambda\mathcal{L}_{KL}, \quad (33)$$

where Q^s_φ is the Q value of the corresponding single-task agent, λ decreases during training.

3) Soft Modulation with Conflict Gradient Projecting: As the difficulty of learning different tasks varies, soft modulation [32] (See Fig. 3) is introduced to address this issue. To implement soft modulation, the network structure of D3QN is divided into multiple layers, and each layer contains a set of modules. A separate routing network with parameter φ_r is built to estimate the connection probability P^l_φ between modules in layer l and layer $l+1$ according to the task and current state. Hence, for different tasks, the task-conditioned policies can use weighted combinations of shared modules to improve their performance. Moreover, the conflict gradient projecting [33] is adopted to mitigate inter-task conflicts.

IV. Case Studies

A. Setup

The simulation data are from a real-world subway line with 4 elevated stations (see Table I). The practical service pattern is listed in [6], where there are 122 daily train services. We treat all headways during 5:30-9:00 and 16:00-19:00 as 350 s, 5:20-5:30 and 19:00-20:00 as 540 s, and 9:00-16:00 and 20:00-22:05 as 660 s. The train parameters are obtained from [34]. The train has a vehicle mass M^{TR} of 200 t, and the maximum

capacity $N^{\mathrm{B}}_{\max}$ is 1500. The train braking resistor setting follows [21]. The URTN [1] and HESS [16], [30] parameters are listed in Table II, where $U^{\mathrm{SUB}}_{\min}$ and $U^{\mathrm{SUB}}_{\max}$ are the minimum and maximum allowed operation voltage, respectively. For URTN, its operation voltage range is reported in [35]. The HESS is assumed to be installed in station 3. Considering that batteries are not suitable to cover the large traction load power [12], its rated power is roughly taken as the average substation power during one train headway. To meet the peak traction power demand [6], the power difference between the average substation power and the peak substation power is roughly treated as the rated supercapacitor power. Thus, the optimal HESS size is searched in a range near to the above empirical size setting. Historical OD and arrival rate tables are obtained from [27].

For simplicity, the delays are only considered during peak hours (namely, 350 s headway), and only one delay occurs at a random station in each scenario. $T^{\mathrm{D}}_{i,\rho}$ is set between 5-20 s [23] using the log-Normal distribution [36] where the mean and variance are both 5 s. Hence, according to the delay time range, each station has 15 combination tasks of speed profiles. The train speed profiles are generated based on our previous work [34], which minimizes the traction energy consumption and meets multiple objectives of punctuality, safety, and riding comfort. $\Delta T = 1$ s.

The Q network has two fully connected layers both with 128 units and ReLU non-linearity, and followed by 2 layers and 2 modules per layer. Each module has two hidden layers both with 128 units and ReLU non-linearity. The routing network outputs 128 representations for connection probability per layer. The target network and task-specific agents have the same structure as the Q network. $\gamma = 0.998$, $\tau = 5 \times 10^{-4}$, $H = 3$, $W = 43$. The optimizer is Adam with a learning rate of 10^{-4}. ϵ reduces linearly from 0.5 to 0.01 and remains constant at 0.01 after 2000 episodes. λ reduces linearly from 1 to 0.05 and remains constant at 0.05 after 2000 episodes. For energy management, the proposed algorithm is trained and tested by 1000 and 21 random traction load scenarios, respectively. The initial HESS SoEs are randomly set. For sizing optimization, 122 random traction load scenarios are included in a daily operation scenario according to the service pattern, and 1000 such daily operation scenarios are generated. Then, to decrease the computational cost, 10 representative scenarios are retained by K-means clustering. The scenario probabilities are 0.112, 0.137, 0.126, 0.101, 0.096, 0.106, 0.077, 0.073, 0.100, and 0.072. The passenger flow fluctuations and their correlations in the representative scenarios are shown in Fig. 4. All simulations are performed by PyTorch 1.12.1 and Python 3.9.13 with an RTX3070 GPU and 32 GB memory.

B. Comparative Analysis of HESS Control Behaviors

In this subsection, the impact of different control schemes on the overall operation cost (see Table III) and HESS control behavior (see Fig. 5 and Fig. 6) are analyzed by the test set. As an example, we take the parallel numbers of 15 and 5 for supercapacitor and battery, respectively. The following schemes are compared: *1) Dynamic threshold and power allocation (DTPA, proposed):* Both thresholds and the power allocation of the HESS can be dynamically adjusted. *2) Fixed threshold (FT):* This scheme aims to verify the effectiveness of threshold adjustments. $U^{\mathrm{CH}}_{\rho,t} = 865$V, $U^{\mathrm{DIS}}_{\rho,t} = 855$V. *3) Fixed power allocation (FPA):* This scheme aims to verify the effectiveness of power allocation adjustments. The fixed power allocation is the nominal battery power divided by nominal supercapacitor power. *4) Fixed threshold and power allocation (FTPA):* A conventional rule-based scheme [14], where both thresholds and power allocation are fixed.

Fig. 5a shows the total train power generation (+) and consumption (-). Fig. 5b shows the train displacement curves. When a delay occurs, the rescheduled energy-efficient speed profile prefers higher deceleration and braking power to avoid extra traction energy consumption due to the decreased running time. Hence, the overall power generation under RTTR is higher than that of normal operation. Figs. 5c to 5f show the HESS control parameters under normal operation and RTTR, respectively. From Figs. 5c and 5e, since the battery operation cost is lower than that of the supercapacitor, DTPA and FT utilize more battery capacity for cost-saving by leveraging a higher power allocation ratio than FPA. Besides, from Figs. 6a and 6c, the DTPA and FT release less energy than FPA and FTPA during 50-75 s, which prevents the supercapacitor SoE

TABLE I
TIMETABLE AND RESCHEDULING SETTINGS.

Section	Direction	$T^{\mathrm{PL}}_{i,j,\rho}$ (s)	$T^{\mathrm{RTTR}}_{i,j,\rho}$ (s)	Length (m)	Dwell time (s)
1-2	Down	104	[94, 104]	1354	30
2-3		165	[155, 165]	2337	
3-4		151	[141, 151]	2265	
4-3	Up	151	[141, 151]	2265	
3-2		162	[152, 162]	2337	
2-1		105	[95, 105]	1354	

TABLE II
URTN AND HESS PARAMETERS.

Battery module (LTO 20Ah)			
Nom. voltage	2.3 V	Nom. capacity	20 Ah
No. in series	292	No. in parallel	[5, 10]
Max. discharge rate	5 C	SoE range	0.2-0.8
$c^{\mathrm{OM}}_{\mathrm{BT}}$	1 \$/MWh	$c^{\mathrm{INV}}_{\mathrm{BT}}$	31.51 \$/module
Supercapacitor module (BMOD0083P048)			
Nom. voltage	48 V	Nom. capacity	165 F
Nom. current	130 A	Resistance	6.3×10^{-3} Ω
No. in series	14	No. in parallel	[15, 25]
SoE range	0.25-0.9	$c^{\mathrm{OM}}_{\mathrm{SC}}$	7.5 \$/MWh
$c^{\mathrm{INV}}_{\mathrm{SC}}$	538 \$/module	–	–
DC-DC converter module			
Max. current	400 A	$c^{\mathrm{INV}}_{\mathrm{DC}}$	38500 \$/module
η^{BT}	0.8	η^{SC}	0.95
URTN			
U^{SUB}_0	860 V	$[U^{\mathrm{SUB}}_{\min}, U^{\mathrm{SUB}}_{\max}]$	[500V, 1000V]
U^{BR}	900 V	R^{SUB}	0.0161 Ω
R^{P}	0.015 Ω	R^{L}	0.016 Ω/km
c_{GRID}	0.11 \$/kWh	J^{FIX}	3.2×10^5 \$
L	10	I^R	2.5 %

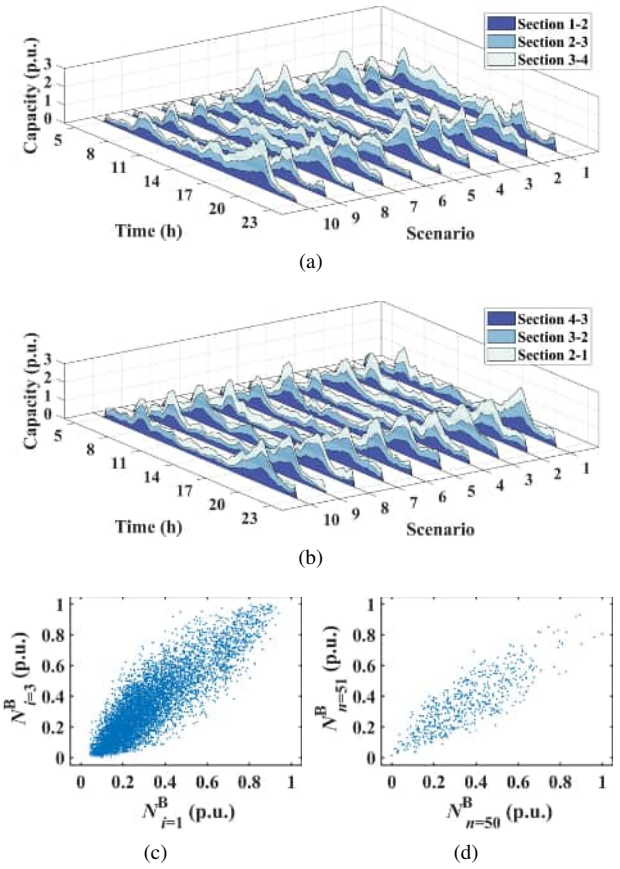

Fig. 4. Passenger flows in (a) down and (b) up directions. Sections 1-2 and 3-4 and time intervals 50 and 51 are selected to show the (c) spatial and (d) temporal correlations.

TABLE III
OVERALL OPEARTION COST OF SCHEMES 1-4.

Performance	DTPA	FT	FPA	FTPA
Overall operation cost ($)	351.57	358.48	368.58	373.56
Electricity trading cost ($)	337.53	344.94	353.95	356.92
HESS operation cost ($)	14.04	13.54	14.63	16.63

from reaching its lowest limit. The continuous supercapacitor power supply of DTPA and FT decreases the substation energy consumption, as shown in Fig. 6e. Moreover, compared with FT, other schemes maintain a reasonable supercapacitor SoE during 275-350 s, which can potentially utilize more supercapacitor energy for further usage. From Figs. 5d and 5f, compared with normal operation, the power allocation ratio of DTPA and FT is closer to FPA. This is to fully utilize the available HESS power to absorb the higher braking power under RTTR. From Figs. 6b and 6d, all schemes show similar performance in maintaining supercapacitor SoE as in normal operation. From Table III, DTPA outperforms other schemes in decreasing the overall operation cost under normal operation and RTTR. The cost reduction is 1.93%-5.89% on average.

C. Impact of Different Optimization Algorithms

In this subsection, four learning-based and two non-learning-based algorithms are compared: *1) KT-D3QN:* proposed. *2) MT-D3QN:* the task set and the routing network are the same as KT-D3QN, while the knowledge transfer is removed. *3) ST-D3QN:* The routing network and the knowledge transfer are not included, and no task set is established. The changes in speed profiles and headways are treated as uncertainties. *4) MTMH-SAC*: the multi-task multi-head soft-actor-critic algorithm which uses an independent head for each task. We revised the realization in [32] to output discrete actions. The above methods are running with 4000 episodes and 3 random seeds. Besides, *5) Genetic algorithm (GA):* GA is directly implemented on the test set, where Q^* is treated as the fitness function. To decrease the computational complexity, we perform GA for each test scenario individually. The population size is 40, the crossover fraction is 0.9, the mutation fraction is 0.1, and the maximum generation is 100. *6) FTPA:* as illustrated in subsection IV-B.

TABLE IV
RBE UTILIZATION AND OVERALL OPERATION COST OF ALGORITHMS 1-6.

Performance	KT-D3QN	MT-D3QN	ST-D3QN
Braking energy (MWh)	25.41	25.41	25.41
RBE (MWh)	17.43	19.32	16.38
Utilization (%)	68.60	76.03	64.46
Cost ($)	351.57	366.39	404.37
Performance	MTMH-SAC	GA	FTPA
Braking energy (MWh)	25.41	25.41	25.41
RBE (MWh)	17.00	15.96	19.74
Utilization (%)	66.90	62.81	77.69
Cost ($)	387.54	403.83	373.56

Fig. 7 shows the reward curves of the test set, where the bold line is the average value, the shaded area is one standard deviation, and the curves of learning-based algorithms are smoothed with a moving average smoothing factor of 0.1 for visual clarity. ST-D3QN gains the lowest reward and shows little improvement with episodes, which indicates that a single-task learning framework is insufficient to handle different headways and multi-source operation uncertainties. KT-D3QN achieves a stable performance and finds a near-optimal policy after 3000 episodes. It obtains the highest reward. Table IV shows the RBE utilization and overall operation cost of the test set with the best performance for each algorithm. Although FTPA achieves the highest RBE utilization, its cost is higher than KT-D3QN and MT-D3QN. This is because the improved RBE recovery of HESS also increases its operation cost. Hence, due to the multi-task learning framework and knowledge transfer, KT-D3QN outperforms other algorithms in improving economic benefits by 4.04%-13.06%, respectively, which verifies its effectiveness.

D. Optimal HESS Sizing & Impact of Different Traffic Models

In this subsection, the following traffic models are compared: *1) Dynamic traffic model:* proposed. *2) Static traffic model:* Only one most common traction load scenario is generated for each headway, and one daily operation scenario containing 122 such traction load scenarios with different headways is used for sizing optimization. Specifically, this

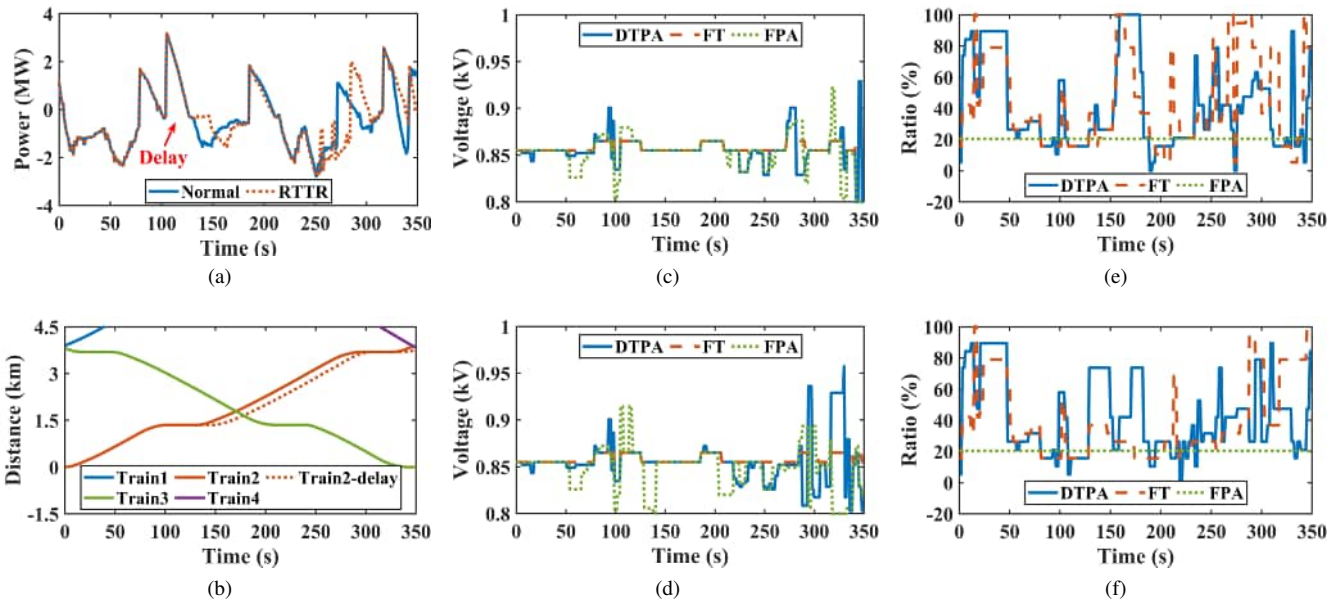

Fig. 5. Train operation and HESS control curves, (a) is total train power, (b) is train displacement, (c)-(d) are HESS thresholds under normal operation and RTTR, respectively. (e)-(f) are HESS power allocations under normal operation and RTTR, respectively.

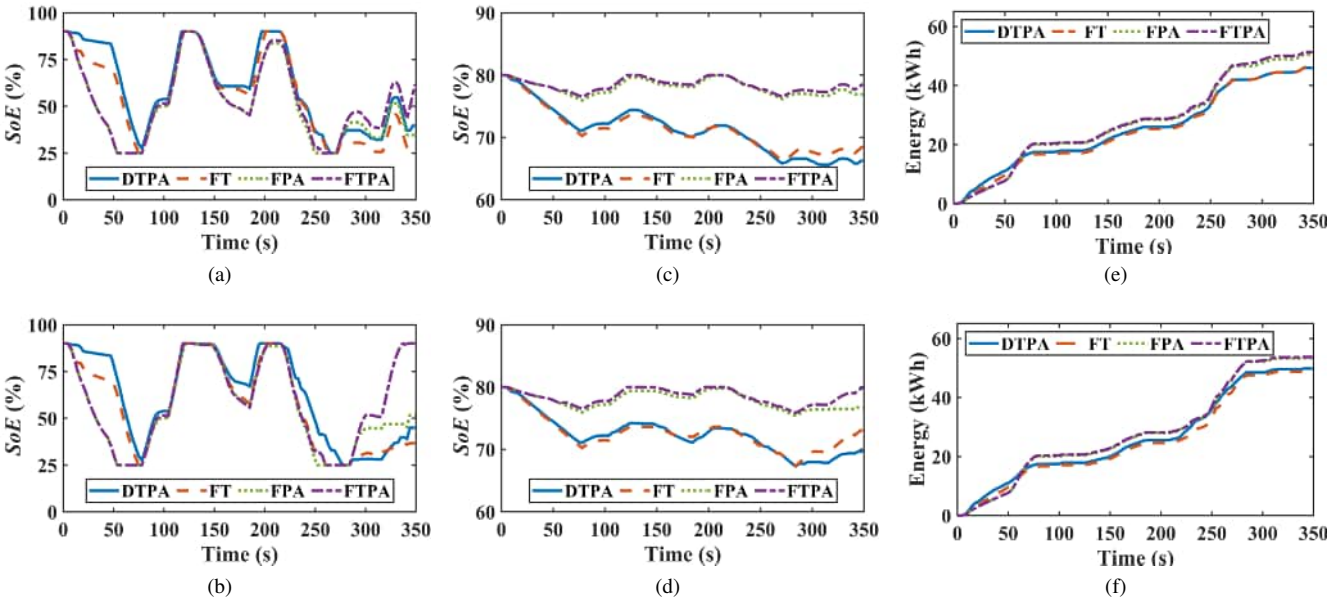

Fig. 6. HESS SoEs and substation energy. (a)-(b) are supercapacitor SoEs under normal operation and RTTR, respectively. (c)-(d) are battery SoEs under normal operation and RTTR, respectively. (e)-(f) are substation energy under normal operation and RTTR, respectively.

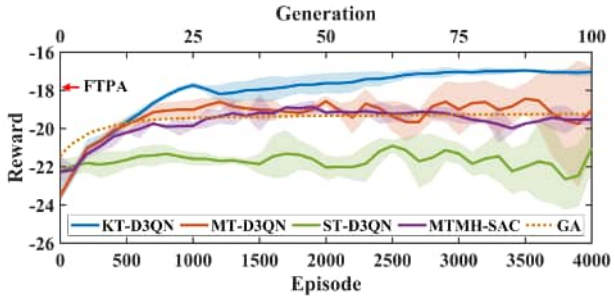

Fig. 7. Reward curves.

daily operation scenario assumes no delays occur and the daily passenger flows follow the historical average daily passenger curve. The train resistance uncertainty is not considered. The initial HESS SoEs are set as the maximum SoE. *3) Static passenger model:* passenger uncertainty is not considered, and the historical average daily passenger curve is used for all daily operation scenarios. *4) No delay model:* the delays and RTTR are ignored, and only the normal operation scenarios in the traction load scenarios are adopted to establish daily operation scenarios. These traffic models are combined with different energy management strategies to optimize the HESS size. Specifically, we use framework F1-F4 to denote the results of combining KT-D3QN with traffic models 1)-4), respectively,

TABLE V
LCC AND OPTIMAL HESS SIZE OF FRAMEWORKS 1-8.

Performance	F1	F2	F3	F4
LCC ($)	1283.53	1184.99	1234.59	1209.40
Supercapacitor capacity (kWh)	14.78	17.74	14.78	17.74
Battery capacity (kWh)	107.46	80.59	107.46	80.59
Supercapacitor power (kW)	1.75	2.10	1.75	2.10
Battery power (kW)	0.54	0.40	0.54	0.40
Battery life (year)	10.00	10.00	10.00	10.00
Performance	F5	F6	F7	F8
LCC ($)	1340.22	1318.48	1301.45	1327.83
Supercapacitor capacity (kWh)	17.74	18.48	17.74	17.74
Battery capacity (kWh)	107.46	120.89	107.46	120.89
Supercapacitor power (kW)	2.10	2.18	2.10	2.10
Battery power (kW)	0.54	0.60	0.54	0.60
Battery life (year)	5.40	5.25	5.37	5.64

and framework F5-F8 to denote the results of combining FTPA with traffic models 1)-4), respectively. F6 (FTPA and static traffic model) is the conventional approach and baseline.

Table V shows the LCC and optimal HESS size under various optimization frameworks. Compared with F1, F2-F4 lacks the consideration of spatio-temporal traction load characteristics on different degrees, which results in the LCC underestimation. Similarly, the LCCs of F6-F8 are lower than F5 due to the lack of the proposed dynamic traffic model. Besides, the LCCs of F5-F8 are significantly higher than F1-F4, which further verifies the effectiveness of KT-D3QN. Compared with the conventional approach F6, the proposed framework F1 reduces the HESS LCC, capacity, and power by 2.65%, 12.29%, and 17.63%, respectively, while increasing the battery life by 86.22%.

V. CONCLUSION

In this paper, a synergistic MTRL-based HESS sizing and energy management optimization framework is proposed for enhancing the economic operation of HESSs under dynamic spatio-temporal urban rail traffic. A DTM composed of a Copula-based passenger flow generation method and a traction energy sensitivity-based RTTR algorithm is developed to characterize multi-train traction load uncertainty. A KT-D3QN algorithm is proposed to simultaneously learn a generalized multi-task HESS control policy from knowledge of annealing task-specific agents and operation environments. The key findings are summarized as follows: 1) With the joint optimization of voltage thresholds and power allocations in the MTMDP to effectively adjust SoEs, the average overall operation cost is reduced by 5.89% compared with conventional rule-based strategies using fixed thresholds and power allocations. 2) Leveraging the multi-task learning framework and knowledge transfer, the proposed KT-D3QN algorithm decreases the average overall operation cost by 4.04%-13.06% compared with benchmark learning-based and non-learning-based methods. 3) The lack of consideration of spatio-temporal traction load characteristics can result in LCC underestimation up to 6.69%. Compared with the conventional approach, the proposed optimization framework reduces the HESS LCC by 2.65% while increasing the battery life by 86.22%. Future works will focus on improving the economic operation of multiple distributed HESSs considering urban traffic uncertainties.

REFERENCES

[1] G. Li and S. W. Or, "Drl-based adaptive energy management for hybrid electric storage systems under dynamic spatial-temporal traffic in urban rail transits," in *2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG)*, 2023, Conference Proceedings, pp. 1–6.

[2] W. To, P. K. Lee, and T. Billy, "Sustainability assessment of an urban rail system–the case of hong kong," *J. Clean. Prod.*, vol. 253, p. 119961, 2020.

[3] H. Hu, Y. Liu, Y. Li, Z. He, S. Gao, X. Zhu, and H. Tao, "Traction power systems for electrified railways: evolution, state of the art, and future trends," *Railw. Eng. Sci.*, vol. 32, no. 1, pp. 1–19, 2024.

[4] M. Khodaparastan, A. A. Mohamed, and W. Brandauer, "Recuperation of regenerative braking energy in electric rail transit systems," *J. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2831–2847, 2019.

[5] S. Fang, Z. Tian, C. Roberts, and R. Liao, "Guest editorial: Special section on toward low carbon industrial and social economy of energy-transportation nexus," *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 8146–8148, 2022.

[6] G. Li and S. W. Or, "Multi-agent deep reinforcement learning-based multi-time scale energy management of urban rail traction networks with distributed photovoltaic-regenerative braking hybrid energy storage systems," *J. Clean. Prod.*, vol. 466, p. 142842, 2024.

[7] Z. Yang, F. Zhu, and F. Lin, "Deep-reinforcement-learning-based energy management strategy for supercapacitor energy storage systems in urban rail transit," *J. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1150–1160, 2020.

[8] H. Novak, V. Lešić, and M. Vašak, "Energy-efficient model predictive train traction control with incorporated traction system efficiency," *J. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5044–5055, 2022.

[9] F. Ciccarelli, A. D. Pizzo, and D. Iannuzzi, "Improvement of energy efficiency in light railway vehicles based on power management control of wayside lithium-ion capacitor storage," *IEEE Trans. Power Electron.*, vol. 29, no. 1, pp. 275–286, 2014.

[10] D. Ramsey, T. Letrouve, A. Bouscayrol, and P. Delarue, "Comparison of energy recovery solutions on a suburban dc railway system," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 3, pp. 1849–1857, 2021.

[11] H. H. Alnuman, D. T. Gladwin, M. P. Foster, and E. M. Ahmed, "Enhancing energy management of a stationary energy storage system in a dc electric railway using fuzzy logic control," *Int. J. Electr. Power Energy Syst.*, vol. 142, p. 108345, 2022.

[12] J. Wang, B. Wang, L. Zhang, J. Wang, N. Shchurov, and B. Malozyomov, "Review of bidirectional dc-dc converter topologies for hybrid energy storage system of new energy vehicles," *Green Energy Intell. Transp.*, vol. 1, no. 2, p. 100010, 2022.

[13] Q. Yu, Y. Liu, S. Long, X. Jin, J. Li, and W. Shen, "A branch current estimation and correction method for a parallel connected battery system based on dual bp neural networks," *Green Energy Intell. Transp.*, vol. 1, no. 2, p. 100029, 2022.

[14] Y. Liu, Z. Yang, X. Wu, D. Sha, F. Lin, and X. Fang, "An adaptive energy management strategy of stationary hybrid energy storage system," *IEEE Trans. Transp. Electrif.*, vol. 8, no. 2, pp. 2261–2272, 2022.

[15] S. Lu, B. Zhang, J. Wang, Y. Lai, K. Wu, C. Wu, and F. Xue, "Energy-efficient train control considering energy storage devices and traction power network using a model predictive control framework," *IEEE Trans. Transp. Electrif.*, pp. 1–1, 2024.

[16] F. Zhu, Z. Yang, Z. Zhao, and F. Lin, "Two-stage synthetic optimization of supercapacitor-based energy storage systems, traction power parameters and train operation in urban rail transit," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8590–8605, 2021.

[17] H. Dong, Z. Tian, J. W. Spencer, D. Fletcher, and S. Hajiabady, "Bi-level optimization of sizing and control strategy of hybrid energy storage system in urban rail transit considering substation operation stability," *IEEE Trans. Transp. Electrif.*, pp. 1–1, 2024.

[18] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of The 33rd International Conference on Machine Learning*, vol. 48. PMLR, 2016, pp. 1995–2003.

[19] X. Wang, Y. Luo, B. Qin, and L. Guo, "Power allocation strategy for urban rail hess based on deep reinforcement learning sequential decision optimization," *IEEE Trans. Transp. Electrif.*, vol. 9, no. 2, pp. 2693–2710, 2023.

[20] F. Zhu, Z. Yang, F. Lin, and Y. Xin, "Decentralized cooperative control of multiple energy storage systems in urban railway based on multiagent deep reinforcement learning," *IEEE Trans. Power Electron.*, vol. 35, no. 9, pp. 9368–9379, 2020.

[21] A. Fernández-Rodríguez, A. Fernández-Cardador, A. P. Cucala, M. Domínguez, and T. Gonsalves, "Design of robust and energy-efficient ato speed profiles of metropolitan lines considering train load variations and delays," *J. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2061–2071, 2015.

[22] Y. Wang, D. Qiu, X. Sun, Z. Bie, and G. Strbac, "Coordinating multi-energy microgrids for integrated energy system resilience: A multi-task learning approach," *IEEE Trans. Sustain. Energy*, vol. 15, no. 2, pp. 920–937, 2024.

[23] J. Yin, D. Chen, L. Yang, T. Tang, and B. Ran, "Efficient real-time train operation algorithms with uncertain passenger demands," *J. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2600–2612, 2016.

[24] X. Liu, A. Dabiri, Y. Wang, and B. D. Schutter, "Real-time train scheduling with uncertain passenger flows: A scenario-based distributed model predictive control approach," *J. Intell. Transp. Syst.*, vol. 25, no. 5, pp. 4219–4232, 2024.

[25] C.-s. Ying, A. H. F. Chow, and K.-S. Chin, "An actor-critic deep reinforcement learning approach for metro train scheduling with rolling stock circulation under stochastic demand," *Transp. Res. B Methodol.*, vol. 140, pp. 210–235, 2020.

[26] Z. Hou, H. Dong, S. Gao, G. Nicholson, L. Chen, and C. Roberts, "Energy-saving metro train timetable rescheduling model considering ato profiles and dynamic passenger flow," *J. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2774–2785, 2019.

[27] S. Yang, Y. Chen, Z. Dong, and J. Wu, "A collaborative operation mode of energy storage system and train operation system in power supply network," *Energy*, vol. 276, p. 127617, 2023.

[28] D. Roch-Dupré, T. Gonsalves, A. P. Cucala, R. R. Pecharromán, A. J. López-López, and A. Fernández-Cardador, "Determining the optimum installation of energy storage systems in railway electrical infrastructures by means of swarm and evolutionary optimization algorithms," *Int. J. Electr. Power Energy Syst.*, vol. 124, p. 106295, 2021.

[29] L. Zhang, D. He, Y. He, B. Liu, Y. Chen, and S. Shan, "Real-time energy saving optimization method for urban rail transit train timetable under delay condition," *Energy*, vol. 258, p. 124853, 2022.

[30] V. I. Herrera, H. Gaztañaga, A. Milo, A. Saez-de Ibarra, I. Etxeberria-Otadui, and T. Nieva, "Optimal energy management and sizing of a battery–supercapacitor-based light rail vehicle with a multiobjective approach," *IEEE Trans. Ind. Appl.*, vol. 52, no. 4, pp. 3367–3377, 2016.

[31] Q. Zhang, H. Wang, Y. Zhang, and M. Chai, "An adaptive safety control approach for virtual coupling system with model parametric uncertainties," *Transp. Res. C: Emerg. Technol.*, vol. 154, p. 104235, 2023.

[32] R. Yang, H. Xu, Y. Wu, and X. Wang, "Multi-task reinforcement learning with soft modularization," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 4767–4777, 2020.

[33] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, "Gradient surgery for multi-task learning," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 5824–5836, 2020.

[34] G. Li, S. W. Or, and K. W. Chan, "Intelligent energy-efficient train trajectory optimization approach based on supervised reinforcement learning for urban rail transits," *IEEE Access*, vol. 11, pp. 31 508–31 521, 2023.

[35] A. D. Femine, D. Gallo, D. Giordano, C. Landi, M. Luiso, and D. Signorino, "Power quality assessment in railway traction supply systems," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2355–2366, 2020.

[36] D. Roch-Dupré, A. P. Cucala, R. R. Pecharromán, A. J. López-López, and A. Fernández-Cardador, "Evaluation of the impact that the traffic model used in railway electrical simulation has on the assessment of the installation of a reversible substation," *Int. J. Electr. Power Energy Syst.*, vol. 102, pp. 201–210, 2018.