Front. Eng. Manag. 2019, 6(2): 139–151
https://doi.org/10.1007/s42524-019-0030-7

REVIEW ARTICLE

**Ziyou GAO, Lixing YANG**

# Energy-saving operation approaches for urban rail transit systems



**Abstract** With the accelerated urbanization in China, passenger demand has dramatically increased in large cities, and traffic congestion has become serious in recent years. Developing public urban rail transit systems is an indispensable approach to overcome these problems. However, the high energy consumption of daily operations is an emerging issue due to increased rail transit networks and passenger demands. Thus, reducing the energy consumption and operational cost by using advanced optimization methodologies is an urgent task for operation managers. This work systematically introduces energy-saving approaches for urban rail transit systems in three aspects, namely, train speed profile optimization, utilization of regenerative energy, and integrated optimization of train timetable and speed profile. Future research directions in this field are also proposed to meet increasing passenger demands and network-based urban rail transit systems.

**Keywords** urban rail transit, energy saving, speed profile optimization, regenerative energy, train timetable

## 1 Background

With the accelerated urbanization in China, traffic congestion has become increasingly serious in recent years. This problem hinders the sustainable development of the Chinese economy. Urban rail transit has become an effective transport mode in many large and medium cities due to its advantages, such as high capacity, high speed, punctuality, safety, and environment friendliness. In December 2018, a total of 32 cities (e.g., Beijing, Tianjin, Shanghai, Nanjing, and Guangzhou) in China operated 156 transit lines with a mileage of 5066 km and 3202 stations. According to the Beijing Urban Metro Plan (2015–2021) by the Beijing Development and Reform Commission, Beijing's urban rail transit system plans to operate 27 lines with a total length of 998.5 km by 2021, and 62% of passengers are expected to use this traffic mode as their travel tools.

As the urban rail transit system continues to develop, an increasing number of citizens prefer to use this system for commuting. Thus, this traffic mode not only attracts passenger travel demands but also mitigates road traffic congestion and enhances road traffic conditions. The daily passenger demand in Beijing Metro is approximately 11.95 million at present and ranks first in China. Forty percent of traveling activities occur at morning and evening peak hours. The maximum transect passenger volume has exceeded 50,000 per hour in Beijing Metro Lines 4 and 6. Meanwhile, the maximum transect passenger volume has exceeded 35,000 per hour in Beijing Metro Lines 1, 5, 10, and 13. Therefore, urban rail transit systems contribute considerably to mitigating road traffic congestion.

The energy consumption of urban rail transit systems increases rapidly with the expansion of transit networks and increased service frequency. For instance, Beijing Metro consumed 1.43 billion kWh of electricity in 2014. At the end of 2015, the total energy consumption of Beijing Metro exceeded 1.6 billion kWh, among which trains consumed approximately 0.933 billion kWh in their traction operations, accounting for 58% of the total energy consumption. Note that the total length of Beijing Metro is expected to reach approximately 1000 km by 2021. This situation would drastically increase the energy consumption of this transport system and cause serious financial pressure and carbon emission issues for urban rail transit companies.

In such circumstances, improving the carrying capacity and reducing the operational energy consumption are crucial research directions in the management of urban metro systems to achieve secure, high-efficiency, energy-

Received January 6, 2019; accepted March 20, 2019

Ziyou GAO, Lixing YANG (✉)
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
E-mail: lxyang@bjtu.edu.cn

saving operations. In practice, train traction energy consumption accounts for half of the total energy consumption. Thus, reducing this type of energy consumption is vital from the perspective of train motion. Since the movement trajectories and timetables of trains are closely related to traction energy consumption, a promising approach to decrease energy consumption is to optimize train speed trajectories and timetables. Additionally, the traction motor can convert kinetic energy into electricity during braking operations to produce regenerative energy. The regenerative energy is usually fed back to the overhead contact line and can be used for other accelerating trains. Thus, improving the utilization of regenerative braking energy is another efficient approach to decrease train energy consumption. In this following discussion, this work first introduces the basic concepts of energy consumption in urban rail transit systems. Then, three optimization techniques, namely, train speed trajectory optimization, optimization of regenerative energy utilization, and integrated optimization of timetable and speed trajectory, are introduced to reduce energy consumption.

## 2 Analyses of energy consumption in urban rail transit systems

The energy consumption of trains is composed of the power consumption of various facilities and equipments for guaranteeing the operations of trains and providing passengers high-quality services. The energy consumption of urban rail transit systems can be divided into two parts, namely, auxiliary and traction energy consumption. Auxiliary energy consumption is caused by activities related to train stations and energy facilities, such as ventilation and air-conditioning systems, lighting systems, escalators, weak systems (including communication, monitoring, and signal systems), water supply, and drainage systems. Traction energy consumption refers to the energy consumption of the movement of trains along the railway track during operations and related auxiliary activities. Specifically, this consumption involves the traction energy utilized during train operation and the energy consumption of a train's own auxiliary equipment. Table 1 shows the composition of traction energy consumption in an urban rail transit system.

As shown in Table 1, energy losses typically include traction, braking, and energy losses due to resistances. Traction loss mainly includes losses from the transformer, converter, gearbox, and motor during the traction process. Resistance loss is primarily caused by overcoming frictional, gradient, and curve resistances. Braking loss commonly occurs during the train braking phase in transformers, converters, gearboxes, and motors. Statistical data indicate that the energy required for this part is approximately 58% of the total energy consumption of an urban rail transit system. In addition, the braking operation of a train can convert mechanical energy into electrical energy. Accordingly, regenerative energy (accounts for approximately 42% of the total energy) is generated and can be fed back to the contact network. If the regenerative energy is absorbed or stored and reused by other trains that are simultaneously accelerating in the traction state, then the total traction energy consumption can be significantly reduced. Otherwise, the energy will be consumed by the heating resistor to stabilize the grid voltage. Two approaches can decrease the total energy consumption. First, for a single train, the train speed trajectory can be optimized to reduce traction energy, where the switching points of the train speed curve or train running state (i.e., traction, cruising, coasting, and braking) should be optimized to achieve energy saving. Second, the operation of multiple trains can be coordinated to increase the utilization rate of regenerative energy. Here, the utilization rate of regenerative energy can be enhanced by increasing the overlapping time of simultaneous traction and braking of trains in the same power supply substation. In addition, some energy storage devices (ESDs) can also be installed along the rail line to store regenerative braking energy. Subsequently, the storage and release of regenerative energy can be increased through coordinating train operations.

## 3 Energy-efficient speed profile optimization

Speed profile optimization is one of the most important methods to reduce traction energy consumption in urban rail transit systems. The train speed profile can be optimized to achieve the guaranteed punctuality and precise stopping through a sequence of operation modes and relevant traction/braking accelerations at each position. Trains in urban rail transit systems are generally operated in four modes, namely, traction, cruising, coasting and braking. Traction energy consumption is associated with operation modes (or switching points of operation modes), track conditions, planned running time, and other

**Table 1** Composition of traction energy consumption

| Energy loss | | | Regenerative energy | Total energy |
|---|---|---|---|---|
| Braking loss | Resistance | Traction loss | | |
| 10.9% | 17% | 30.1% | 42% | 100% |

factors. For instance, a train consumes energy in the traction mode, but regenerative energy is produced in the braking mode. In the following content, we determine the energy-saving sequence of operation modes with guaranteed punctuality and precise stopping.

### 3.1 Driving mode description

An illustration of the train operation process is depicted in Fig. 1, where a section $[0,s_3]$ is divided into three stages, namely, traction, cruising/coasting, and braking. The recommended and actual speed profiles are denoted by blue and black lines, respectively. In the traction interval $[0,s_1]$, the train departs from the station and uses the maximal traction acceleration to reach the cruising speed. Then, traction and coasting modes are switched by ATO/the driver to control the operational speed with the guidance of the cruising speed (blue line). At the final stage, the braking rate is dynamically adjusted to stop the train at the next station. In practice, this type of driving sequence is suitable for short-distance operations in metro systems.

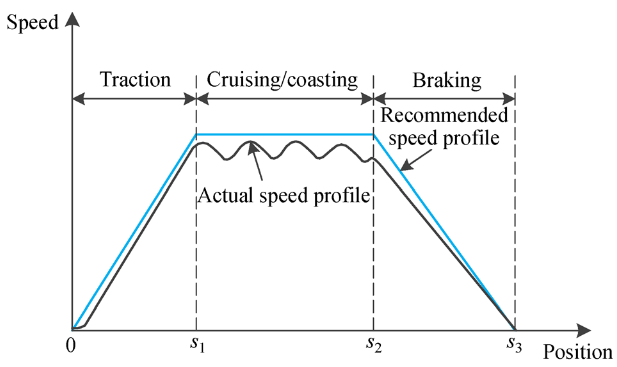


**Fig. 1** Recommended and actual speed profiles

The train movement model is described here to analyze the energy consumption in the operation process. Different forces, including traction, braking, and resistance, are imposed on the train to drive its movement during operations. The computational method for the resultant force is as follows:

$$F_a = u_f F(v) - u_b B(v) - F_{res}(s),$$

where $F_a$ represents the resultant force; $F(v)$ is the maximum traction force; $B(v)$ is the maximum braking force; $F_{res}(s)$ is the resistance force; $u_f$ and $u_b$ are the control ratios of maximum traction force and maximum braking force, respectively; $s$ represents the position; and $v$ is velocity. The relevant energy consumption in each operation mode is discussed below (Wang, 2016).

**Traction mode.** In this mode, the traction ratio is set as $u_f \in (0,1]$, and braking ratio $u_b$ is set as 0. Thus, resultant force $F_a$ can be expressed as $u_f F(v) - F_{res}(s)$. Traction power is provided by the traction supply grid through the adjustment of control ratio $u_f$ to drive train movement. The consumed energy is called traction energy consumption.

**Cruising mode.** In this mode, the train needs to maintains a steady speed, that is, resultant force $F_a = 0$. The consumed energy is determined by resistance force. Specifically, when $F_{res}(s)>0$ or $F_{res}(s) < 0$, traction or braking force is required for keeping the steady speed, which is presented as $u_f \in (0,1]$ or $u_b \in (0,1]$. Notably, energy is consumed when traction force is adopted, and the braking operation generates regenerative energy that is fed back to the traction supply grid.

**Coasting mode.** By removing traction and braking forces (i.e., $u_f = 0$ and $u_b = 0$), natural deceleration or acceleration occurs in the train operation. Deceleration or acceleration is dependent on the resistance force and resultant force $F_a = -F_{res}(s)$. In this mode, consumed energy can be set as 0.

**Braking mode.** In the braking mode, traction control ratio $u_f$ is set as 0, and braking control ratio $u_b$ can be adjusted from 0 to 1. The resultant force is calculated as $F_a = -u_b B(v) - F_{res}(s)$. Specifically, acceleration also occurs when resistance force is less than 0. The regenerative energy can be used by on-board auxiliary devices, and the rest is fed back to the traction supply grid for other traction trains.

### 3.2 Optimization model

Energy-saving speed profile optimization has elicited much attention in recent years. In the 1960s, Ichikawa (1968) proposed the speed optimization model to determine train trajectory for reducing energy consumption. In his work, a straight track without gradient was assumed in the entire inter-station. In the research of Howlett et al. (1994) and Liu and Golovitcher (2003), according to the analysis of the maximum principle, the optimal driving sequence with maximal acceleration, cruising, coasting, and maximal deceleration exhibits enhanced performance with regard to saving energy. Ke et al. (2009; 2012) considered speed optimization on moving and fixed block signaling systems, and the mathematical model was solved with a MAX–MIN ant system algorithm to decrease the computation time. Huang et al. (2018) designed a dynamic programming-based algorithm to optimize train speed profiles with on-board ESDs.

As indicated in the discussion above, a series of concerns, including speed and running time constraints and distance of the inter-station, are considered in the speed profile optimization of a single train. For a metro train, the mass of a train is simplified as a single particle. The energy-saving speed optimization model is formulated according to the train movement model to minimize the traction energy consumption $E(T)$, as shown below (Liu and Golovitcher, 2003).

$$\begin{cases} \min\ E(T) \\ \text{s.t.} \\ \qquad M\dfrac{\mathrm{d}v}{\mathrm{d}t} = u_f F(v) - u_b B(v) - F_{res}(s) \\ \qquad \dfrac{\mathrm{d}s}{\mathrm{d}t} = v \\ \qquad s(0) = S_0,\ s(T) = S_T \\ \qquad v(0) = V_0,\ v(S_T) = V_T \\ \qquad v \leqslant V(s) \\ \qquad u_f \in [0,1],\ u_b \in [0,1] \end{cases}, \quad (1)$$

where $v$ represents the current velocity; $s$ denotes the current site; $T$ indicates the total running time; $E(T)$ is the total energy consumption; $F(v)$ and $B(v)$ are maximum traction and braking forces, respectively; $u_f$ and $u_b$ represent traction and braking control ratios, respectively; $F_{res}(s)$ is the resistance of the train at site $s$; $M$ is the mass of train; $V(s)$ is the speed limit at site $s$; $S_0$ and $S_T$ are start and end sites of the train, respectively; and $V_0$ and $V_T$ are velocities at the start and end sites, respectively. The first two constraints provide a train movement model to calculate train speed. The boundary conditions of speed and position are described in the two subsequent constraints. The maximal speed limitation and decision variables are formulated in the last two constraints.

### 3.3 Optimization algorithm

Various algorithms have been proposed to solve the train speed profile problem. These algorithms can be classified into analytical, numerical, and intelligence algorithms. A detailed description is presented below.

**Analytical algorithms.** This type of algorithm uses the analytical methods to solve the optimization model and determine the global optimal solution. The maximum principle is a common method to generate the optimal speed profile. An example of the maximum principle is provided below to illustrate the solution pocess.

The entire inter-station can be divided into a set of small distance intervals by considering the gradient and speed limit of the track. As shown in Model (1), the objective function is reformulated as a Hamiltonian function, and the driving sequence is optimized to maximize the value of the Hamiltonian function in each distance interval. The detailed process is as follows (Howlett et al., 1994; Liu and Golovitcher, 2003). First, the train movement model can be rewritten as

$$\frac{\mathrm{d}v}{\mathrm{d}s} = \frac{u_f F(v) - u_b B(v) - F_{res}(s)}{M \cdot v}. \quad (2)$$

Aiming to minimize traction energy consumption, the objective function $J$ can be transformed by introducing Lagrange multiplier ($L > 0$), expressed as

$$J = \int_{S_0}^{S_T} (u_f F(v) + L/v)\mathrm{d}s.$$

With objective function $J$, Hamiltonian function $H$ is defined with conjugate function $P$ and complementary slackness conditions $M(s)$. Then, we obtain

$$H = -u_f F(v) - \frac{L}{v} + \frac{P}{M \cdot v}\big(u_f F(v) - u_b B(v) - F_{res}(s)\big),$$
$$\frac{\mathrm{d}P}{\mathrm{d}s} = -\frac{\partial H}{\partial v} + \frac{\mathrm{d}M}{\mathrm{d}s},$$
$$(v - V(s))\frac{\mathrm{d}M}{\mathrm{d}s} = 0,$$
$$\frac{\mathrm{d}M}{\mathrm{d}s} \geqslant 0.$$

Specifically, the optimal sequence can be obtained when $\frac{\partial H}{\partial s} = 0$ and $\frac{\mathrm{d}P}{\mathrm{d}s} = 0$ in each distance interval. The recommended speed profile can then be generated over the inter-station.

**Numerical algorithms.** This type of algorithm adopts different search methods, such as gradient method and sequential quadratic programming, to determine the local optimal solution. The algorithm parameters can be adjusted for different operational environments to make a trade-off between optimizaiton accuracy and computation time.

Sequential quadratic programming algorithm used by Gu et al. (2014) is presented here as an illustration to show how to find the near-optimal speed profile. In this algorithm, the searching process is seperated into two loops, i.e., outer loop and inner loop. In the outer loop, this algorithm approximate the nonlinear objective function into a quadratic function and the nonlinear constraints into their approximate linear formulations. Then, the inner loop employs the active set method to solve the quadratic programming subproblem. This pocess corresponds to a point to point search through constructing the searching direction by Hessian matrix and step length by linear search at each solution.

**Intelligence algorithms.** This type of algorithm is dependent on artificial intelligence and computer technology, and has been widely applied in complex environments to determine near-optimal solutions. Examples include genetic algorithm, colony optimization algorithm, and tabu search. The genetic algorithm is given as an example to illustrate train speed profile optimization (Chang and Sim, 1997; Ma et al., 2010; Yang et al., 2013).

In genetic algorithm, the decision variables should be effectively encoded as chromosomes in order to determine the optimal driving sequence. Then, a set of chromosomes are updated through selection, crossover, and mutation operators to determine the near-optimal solution. For enhanced performance, the fitness function of chromosomes and population size are designed with consideration of train operation constraints. An example is depicted in Fig. 2, where the considered inter-station is divided into

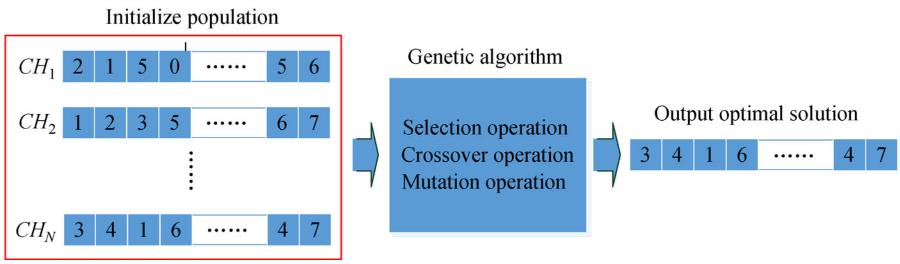


**Fig. 2** GA for optimization of train speed profile

distance intervals according to speed limitation and track gradient. In each distance interval, the train operation mode is represented as control level indexes 0–7 corresponding to the train velocity and recommended speed profile by the train movement model. The population is formed by a series of solutions, $CH_1$, $CH_2$, ..., $CH_N$, in which $N$ denotes population size. Feasible solutions are generated in each iteration through selection, crossover, and mutation operators. The search process is terminated when the termination condition is satisfied (e.g., the best objective value is not changed for a certain number of iterations or the search process reaches the maximum generation). Then, the best operation sequence is used to generate the best recommended speed profile.

## 4 Optimization of regenerative energy utilization in urban rail transit

Frequent braking during train operation generates considerable regenerative braking energy due to the short distance between subway stations. This part of regenerative braking energy can be fed back to the overhead contact line through the pantograph and immediately absorbed by nearby trains and auxiliary equipments. Therefore, improving the utilization ratio of regenerative braking energy is another important approach to reduce the overall system energy consumption. Moreover, this approach is also an important measure to improve energy efficiency and achieve energy conservation and environmental protection.

### 4.1 Utilization of regenerative braking energy

In urban rail transit, the regenerative braking energy generated by braking trains is preferentially used for the auxiliary systems (such as ventilation, air conditioning, and lighting) of the braking train itself. Then, a part of the remaining energy is directly transmitted to other accelerating trains in the same substation through the overhead contact line, and the remaining part is consumed by heating resistors to keep the voltage steady.

When a train is running on the metro line with less online trains, only a small portion of regenerative braking energy can be directly used because the auxiliary system of trains consumes relatively little energy. Most of the regenerative braking energy will be transformed into heat energy that cannot be fully utilized. This condition increases the burden on the subway ventilation system and further causes energy wastage. Therefore, how to efficiently recycle and reuse regenerative braking energy has become a key issue in urban rail transit, and it is of practical significance to reduce the overall system energy consumption.

With the progress of new technologies, three indirect utilization methods for regenerative braking energy have been gradually developed (Allegre et al., 2010; Zhao S, 2014); they are resistance energy-consuming device, inverter feedback device and energy storage device. The following text provides a specific discussion.

**Resistance energy-consuming device.** When more than one train brakes, a substantial amount of regenerative braking energy is fed back to the direct-current grid. This condition may cause the line voltage to exceed the voltage threshold and could damage the power supply system. At this time, a resistance energy-consuming device can consume the excess regenerative braking energy, and the voltage can be controlled within a reasonable range to stabilize the line voltage. Resistance energy-consuming devices are widely used due to their simplicity, strong reliability, and low cost. However, they also have shortcomings. Using heating resistors to consume excess regenerative braking energy may result in energy wastage and increased environment temperature. Therefore, there are some requirements for ventilation and temperature control devices, which further increase the energy consumption.

**Inverter feedback device.** The inverter feedback device is a tool to convert regenerative braking energy into alternating current through inverters and feed it back to the alternating-current grid for other trains and equipments. The inverter feedback device can not only improve the utilization ratio of regenerative braking energy but also reduce the power output from the traction power supply substation. Moreover, this device can prevent the rise of environment temperature and further decrease the burden on the ventilation and air-conditioning system. Therefore, the inverter feedback device is an energy-efficient means to recycle and reuse energy. With the latest technological development, urban rail transit mostly uses inverter feedback devices as the preferred method to absorb and reuse regenerative braking energy. This method has gradually become an important means of energy saving and emission reduction in urban rail transit.

**Energy storage device.** This type of device can store the regenerative braking energy and release it when the electricity is required for train traction. Thus, voltage can be effectively suppressed by employing this method, and the utilization ratio of regenerative braking energy can be efficiently improved. In addition, the damage on the power supply system caused by the feedback of regenerative energy can be avoided.

Different energy storage principles (Zhao S, 2014) indicate that energy storage methods can be divided into battery, flywheel, and super capacitor energy storage. Table 2 compares the advantages and disadvantages of these energy storage methods.

### 4.2 Calculation of regenerative braking energy

In urban rail transit systems, reasonable utilization of regenerative braking energy can further reduce the amount of traction energy obtained from the traction substation, thereby diminishing the total energy consumption of the subway system. The following discussion provides specific formulas to calculate regenerative braking energy.

During train operation, the actual generated regenerative braking energy $E_{br}$ can be expressed as the integral of the corresponding braking power over the time horizon. Braking power can be calculated as the product of braking force and speed.

$$E_{br} = \int_{t \in T} P_{br}(t)\mathrm{d}t = \int_{t \in T} F_{br}(v,t) \cdot v(t)\mathrm{d}t, \tag{3}$$

where $P_{br}(t)$ is the braking power of the train at time $t$, $F_{br}(v,t)$ is the braking force of the train at time $t$, and $v(t)$ is the velocity of the train at time $t$. Braking force can be easily calculated based on deceleration and total resistance, that is,

$$F_{br}(v,t) = M \cdot d(t) - F_{re}(v,t), \tag{4}$$

where $M$ refers to the mass of the train, $d(t)$ is the deceleration of the train at time $t$, and $F_{re}(v,t)$ is the total resistance of the train at time $t$.

The resistance of a train during operation varies according to the velocity of the train. Total resistance consists of various types of resistances, which can be divided into two categories, namely, basic and additional. Basic resistance exists in the entire running process when a train runs on the ideal line. Additional resistance is minimally affected by the train itself and mainly depends on line conditions; it consists of gradient, curve, and tunnel resistances. This study aims to provide specific formulas to compute different resistances (Rao, 2006).

**Basic resistance.** Basic resistance includes the frictional resistance between the axle and axletree, the rolling resistance between the wheel and rail, and the collision resistance from the rail joint to the wheel. In actual operation, basic resistance is difficult to calculate accurately using a theoretical formula due to the complex operating environment. On this basis, an empirical formula is generally adopted instead of a theoretical one to compute basic resistance. The Regulations on Train Traction Calculation indicate that the basic resistance of a train can be expressed as

$$w_{bre}(v,t) = a + b \cdot v(t) + c \cdot v^2(t),$$

where $w_{bre}(v,t)$ represents the unit basic resistance of the train (unit: N/KN) and $a$, $b$, and $c$ are coefficients determined by experiments.

**Gradient resistance.** Gradient resistance is generated by the component of gravity along a slope. This resistance changes due to positive or negative slope (where uphill is positive and downhill is negative). When a train goes uphill, the component of gravity along the slope is opposite to the running direction of the train. At this time, the gradient resistance hinders the train operation. On the contrary, when the train goes downhill, the component of gravity along the slope is similar to the train running direction, which promotes train operation. The Regulations on Train Traction Calculation indicate that gradient resistance is calculated as

$$w_{gre} = i,$$

where $w_{gre}$ is the unit gradient resistance of the train (unit: N/KN) and $i$ refers to the thousandth of the line slope (‰).

**Table 2** Comparison of three energy storage methods

| Storage methods | Advantages | Disadvantages |
|---|---|---|
| Battery energy storage | Good property in energy saving<br>High energy density | Expensive equipment cost<br>Short battery life<br>Environment pollution |
| Flywheel energy storage | Quick charge<br>Long life<br>Eco-friendly | Complex operation system<br>High requirement for working environment |
| Super capacitor energy storage | Quick charge and discharge<br>Long life<br>Good property in energy saving | Expensive equipment cost |

**Curve resistance.** When a train runs on a curved line, the centrifugal force increases the pressure between the rim and rail. Additional lateral and longitudinal sliding occur between the wheel and rail, thus creating an additional resistance to the running train. This resistance is denoted as curve resistance. The Regulations on Train Traction Calculation indicate that curve resistance can be calculated as

$$w_{cre} = c/R,$$

where $w_{cre}$ denotes the unit curve resistance of the train (unit: N/KN), $c$ indicates the empirical coefficient, and $R$ is the radius of curvature (unit: m).

**Tunnel resistance.** Tunnel resistance mainly refers to the air resistance in the tunnel. When a train runs in the tunnel, the air in the tunnel is squeezed by the train, creating a pressure difference between the head and rear of the train. Tunnel resistance is related to many factors, such as train speed, train length, windward area, tunnel length and cross-sectional area, and surface roughness of the train and tunnel, which are generally calculated using the following empirical formula.

$$w_{tre} = \begin{cases} L_s \cdot v^2(t)/10^7, \text{with limited gradient} \\ 0.00013 \cdot L_s, \text{without limited gradient}, \end{cases}$$

where $w_{tre}$ denotes the unit tunnel resistance of the train (unit: N/KN) and $L_s$ is the tunnel length (unit: m).

In summary, train resistance is the sum of basic and additional resistances. The specific calculation equation is

$$F_{re}(v,t) = (w_{bre} + w_{gre} + w_{cre} + w_{tre}) \times g \times M/1000, \quad (5)$$

where $g$ denotes the gravitational acceleration.

These analyses indicate that the regenerative braking energy generated by braking trains can be calculated with Eqs. (3), (4), and (5). Train resistance is closely related to the running state, and the time-dependent running state significantly increases the calculation difficulty of the resistance. Therefore, regenerative braking energy is usually calculated using simulation methods.

### 4.3 Optimization of regenerative braking energy

During the operation of urban rail transit, the actual energy consumption of the system is the difference between the traction energy consumption and the utilization of regenerative braking energy. The following content introduces the utilization of regenerative braking energy from two aspects.

#### 4.3.1 Optimize the direct utilization process

When a train brakes, it can convert 20%–60% of its own kinetic energy into electrical energy (regenerative braking energy), which is then fed back to the contact line. In this case, when a train brakes and other trains that belongs to the same power supply substation accelerate, the regenerative braking energy generated by the braking train is directly transmitted to the accelerating trains. On the contrary, when no train accelerates in this substation, the generated braking energy is converted into heat energy by heating resistors. Therefore, the entering and leaving times of multiple trains should be reasonably scheduled on each segment, and the switching time of traction, cruising, and braking operations should be arranged to effectively improve the direct utilization of regenerative braking energy. Accordingly, the overlapping time between accelerating and braking trains that belong to the same power supply substation can be improved. For clarity, Fig. 3 illustrates an example to describe the train operation scenario.

As shown in Fig. 3(a), train $i$ brakes to enter station $n+1$ while adjacent trains $i+1$ and $i-1$ accelerate. Suppose that stations $n$ and $n+1$ belong to the same power supply substation. Then, the regenerative braking energy generated by train $i$ can be directly fed back to trains $i-1$ and $i+1$ through the contact line.

Meanwhile, Fig. 3(b) describes the overlapping time among braking train $i$ and accelerating trains $i-1$ and $i+1$ to clearly introduce the utilization process. Assume that the leaving time of train $i-1$ at station $n$+1 is $t_d^{i-1,n+1}$, the leaving time of train $i+1$ at station $n$ is $t_d^{i+1,n}$, and the entering time of train $i$ at station $n+1$ is $t_a^{i,n+1}$. These notations indicate that the regenerative braking energy produced by train $i$ in time interval $[t_d^{i-1,n+1}, t_d^{i+1,n}]$ is only transmitted to train $i-1$. By contrast, the regenerative braking energy in time interval $[t_d^{i+1,n}, t_a^{i,n+1}]$ is simultaneously transmitted to the accelerating trains (i.e., $i-1$ and $i+1$).

The regenerative energy generated by train $i$ in time interval $[t_d^{i-1,n+1}, t_d^{i+1,n}]$ can be expressed as $E_{br}^1 = \int_{t_d^{i-1,n+1}}^{t_d^{i+1,n}} P_{br}^i(t)\mathrm{d}t$ according to the calculation formula of regenerative braking energy. The required traction energy of train $i-1$ in the same time interval is $E_{tr}^1 = \int_{t_d^{i-1,n+1}}^{t_d^{i+1,n}} P_{tr}^{i-1}(t)\mathrm{d}t$, where $P_{br}^i(t)$ and $P_{tr}^{i-1}(t)$ represent the braking power of train $i$ and the traction power of train $i-1$ at time $t$, respectively. In view of the short distance in each section, we do not consider the transmission loss of the regenerative braking energy. Therefore, the total utilized regenerative braking energy in time interval $[t_d^{i-1,n+1}, t_d^{i+1,n}]$ is $U_1 = \min\{E_{br}^1, E_{tr}^1\}$, and the utilization ratio is $(U_1/E_{br}^1) \cdot 100\%$. Similarly, the generated braking energy by train $i$ in time interval $[t_d^{i+1,n}, t_a^{i,n+1}]$ is $E_{br}^2 = \int_{t_d^{i+1,n}}^{t_a^{i,n+1}} P_{br}^i(t)\mathrm{d}t$, and the total required traction energy

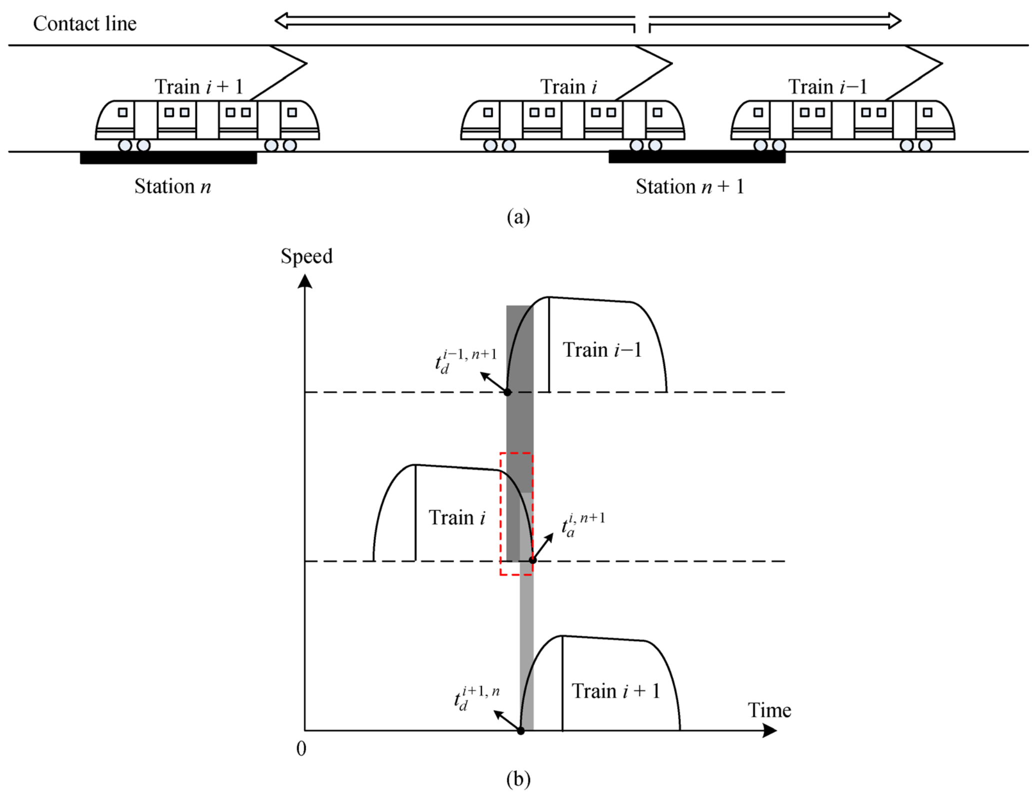


**Fig. 3** Schematic of the utilization of regenerative braking energy

of trains $i-1$ and $i+1$ is $E_{tr}^2 = \int_{t_d^{i+1,n}}^{t_a^{i,n+1}} [P_{tr}^{i-1}(t) + P_{tr}^{i+1}(t)]\mathrm{d}t$, in which $P_{tr}^{i+1}(t)$ is the traction power of train $i+1$ at time $t$. Hence, the total utilized regenerative braking energy is $U_2 = \min\{E_{br}^2, E_{tr}^2\}$, and the utilization ratio is $(U_2/E_{br}^2)\cdot 100\%$.

In summary, the direct utilization process of regenerative braking energy is closely associated with the overlapping time between accelerating and braking trains that belong to the same power supply substation. In reality, with the given recommended speed profile, the entering and leaving times of multiple trains at stations or on the segments can be optimized to maximize the overlapping time, further improve the utilization ratio of regenerative braking energy, and reduce the total energy consumption. The related research was conducted by Yang et al. (2013), Li and Lo (2014), Zhao L (2014) and Yang (2016).

### 4.3.2 Optimize the indirect utilization process

In practice, ESDs can be adopted to further improve the utilization ratio of regenerative braking energy. For example, wayside ESDs can be installed on a metro line or station to capture recovery energy that is not used immediately and provide this energy to accelerating trains, thus achieving the indirect utilization of regenerative braking energy. In this way, the indirect utilization process is realized (Liu et al., 2018).

Figure 4 shows a scenario of the utilization of regenerative braking energy. Train 1 stops at a station, and at this time, train 2 brakes to enter this station. Considering that train 1 cannot directly utilize the regenerative braking energy generated by train 2, such energy is stored in the wayside ESD. Suppose that train 1 starts and runs toward the next station and train 2 stops at the station. The stored regenerative braking energy is released and provided to train 1 for its traction. Evidently, this process includes the storage and release of regenerative braking energy, resulting in a reduction in energy consumption in the urban rail transit system.

During the utilization of regenerative energy, the direct utilization process is first considered, followed by the indirect one. On this basis, the regenerative braking energy that is directly used should be calculated, and the remaining energy is stored in the ESDs for further reuse. The following content introduces a simulation method for the calculation of energy consumption in an urban rail transit system with a wayside ESD. The detailed procedure is summarized below (Liu et al., 2018).

Step 1: Initialize start time $t$, time step $\delta$, total simulation time $T$, and energy $E_n(t)$ in ESD.

Step 2: Calculate traction energy $E_{tr}^i(t)$, auxiliary energy $E_{au}^i(t)$, and regenerative braking energy $E_{br}^i(t)$ according to the running state of train $i$ at time $t$.

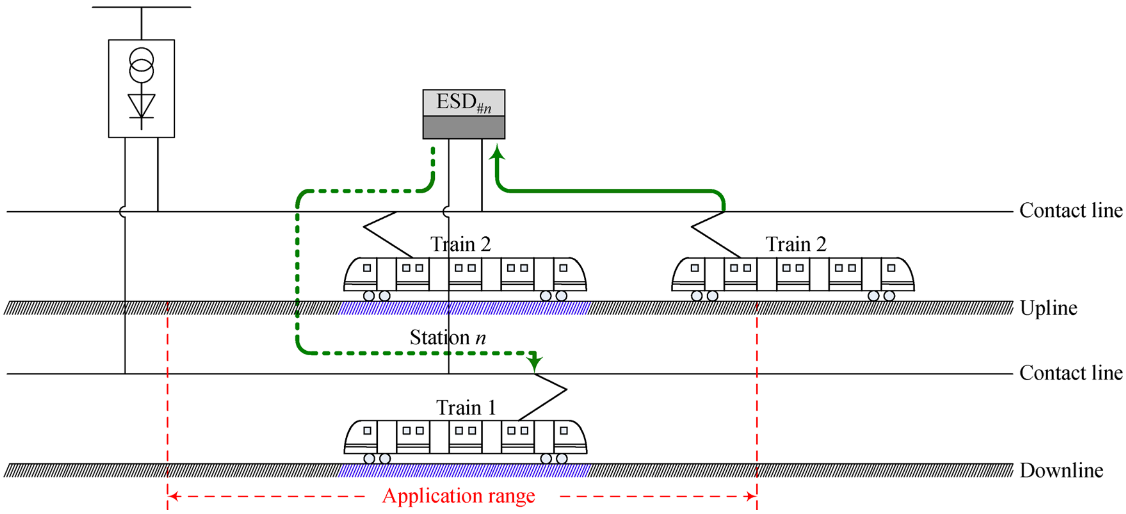


**Fig. 4** Storage and utilization of regenerative braking energy

Step 3: Determine whether a train exists in $S_n$ at time $t$ on the basis of the application range $S_n$ of ESD.

Step 4: If no train exists in $S_n$ at time $t$, update $E_n(t)$ according to the energy $E_n(t-\delta)$ in ESD.

Step 5: If one or more trains exist in $S_n$ at time $t$, calculate the total consumed energy $\sum_i(E^i_{tr}(t)+E^i_{au}(t))$ of all trains in $S_n$ and the total generated braking energy $\sum_i E^i_{br}(t)$ by all trains in $S_n$ according to step 2.

Step 6: Compute the directly utilized regenerative braking energy $\overline{E}^n_{br}(t)$ in $S_n$ at time $t$ according to the calculation results of step 5. Then, calculate the indirectly utilized regenerative braking energy $\hat{E}^n_{br}(t)$ in $S_n$ and actual consumed energy $E^n_{ac}(t)$ according to the energy $E_n(t-\delta)$ in ESD. Subsequently, update $E_n(t)$.

Step 7: Compute the total consumed energy $E'_{ac}(t)$ of all trains that are not in $S_n$ at time $t$.

Step 8: Compute the actual total consumed energy $E^t_{ac} = E'_{ac}(t)+\sum_n E^n_{ac}(t)$ of all trains at time $t$ and the total of directly and indirectly utilized regenerative braking energy $\overline{E}^t_{br} = \sum_n \overline{E}^n_{br}(t)$ according to the calculation results of steps 6 and 7. Then, update $t=t+\delta$. Repeat the above-mentioned steps until $t=T$.

The above-mentioned analyses indicate that the train timetable must be optimized to improve the overlapping time between accelerating and braking trains. When a train cannot operate according to the timetable due to disturbance, the direct utilization process of regenerative braking energy is affected. In this condition, the utilization ratio of regenerative braking energy can be effectively increased if we employ ESDs to absorb and store the regenerative braking energy. Therefore, comprehensively considering the direct and indirect utilization of regenerative braking energy is practically important for energy saving of urban rail transit systems.

## 5 Integrated optimization of train timetable and speed profile

Currently, the train scheduling approach in the urban rail transit includes the following approaches. The first is manual plotting method that is based on manual experience using Excel and AutoCAD software. In fact, this kind of method often takes a lot of time to plot a train timetable and the workload is huge, which is also difficult to meet the requirements of urban rail managers in adjusting with varied passenger demand, complex signaling equipment and transportation organizations. Besides, automatic train scheduling process can be fulfilled with the aid of computers. The advanced train control system of rail transit is embedded with the train timetable generation module, and it only needs to input specific line information, vehicle and passenger demand data to automatically generate a corresponding train timetable.

However, the current train scheduling method only allocates rail transit vehicle resources to the operating line according to experience or certain rules, which does not fully consider the dynamic characteristics of passenger demand and energy consumption of trains. In addition, the train timetabling method has not been combined with the ATO speed curve design and the utilization of regenerative energy during train operation. Thus, this part of regenerative energy cannot be fully utilized, resulting in unnecessary waste of electric energy. This section focuses on the collaborative optimization of train timetable and speed profile to deal with this situation. The purpose is to simultaneously achieve service quality improvement and energy consumption reduction (as shown in Fig. 5).

### 5.1 Energy consumption and train timetable

In practice, the train running time between two stations has a direct relationship with traction energy consumption. Specifically, reducing the running speed between any two stations increases train running and cyclical times but saves train energy consumption. Therefore, in the process of train scheduling and designing the ATO recommended speed profile, the train dwelling time and running time between stations can be reasonably adjusted according to the dynamic demand of passenger flow. Moreover, the energy consumption can be further diminished with the guaranteed passenger service level. This kind of method is discussed as follows.

The maximum line capacity in an urban rail transit is determined by train departure frequency. Specifically, a short train departure headway corresponds to a large line capacity. This relationship can be explicitly expressed as follows:

$$C = \frac{3600}{h},$$

where $C$ is the maximum line capacity per hour and $h$ is the train departure headway (unit: s). As shown in the equation, the departure headway is inversely related to system capacity. Given the huge passenger demand during peak hours, the train running interval should be compressed as much as possible to increase the system capacity. By contrast, the passenger demand is low during off-peak hours. So it is not necessary to guarantee the maximum capacity of the line, and the train departure headway can be appropriately increased. In this manner, the number of trains can be reduced, thereby saving a substantial amount of energy.

The running time between two stations and train traction energy consumption are also interrelated. For instance, the minimum traction energy consumption of a train typically has an inverse relationship with segment running time. In particular, with long train running time, the train running speed is low, and the traction energy consumption is diminished. Conversely, with short running time, the train running speed is high, and the traction energy consumption is increased. Therefore, the running time between stations at off-peak hours can be appropriately increased. In addition, the traction energy consumption of the train in the entire system can be reduced by sacrificing the "cyclic time."

From the perspective of train timetabling, the utilization of regenerative energy can also be improved by adjusting the departure time of trains in up and down directions and the switching points of train operation states (e.g., accelerating, cruising, braking and coasting). Figure 6 shows the relationship between the spatiotemporal position and energy consumption of a pair of trains under two scenarios. When a braking train generates regenerative energy, the regenerative energy increases the line voltage when no train is in traction operations in the same power supply substation (see scenario A in Fig. 6). When the line voltage rises to a certain value, the system initiates overvoltage protection, and the regenerative energy is consumed by the heating resistor. When the braking and accelerating trains are successfully matched and these two are in the same power supply substation, the regenerative energy is absorbed and utilized, thereby saving substantial system energy (see scenario B in Fig. 6).

### 5.2 Integrated optimization by a space–time network presentation method

The space–time network is a commonly used mathematical tool in the modeling of dynamic traffic management. This method plays a vital role in designing train timetables and speed profiles. This section briefly introduces the integrated optimization of train timetables and speed profiles

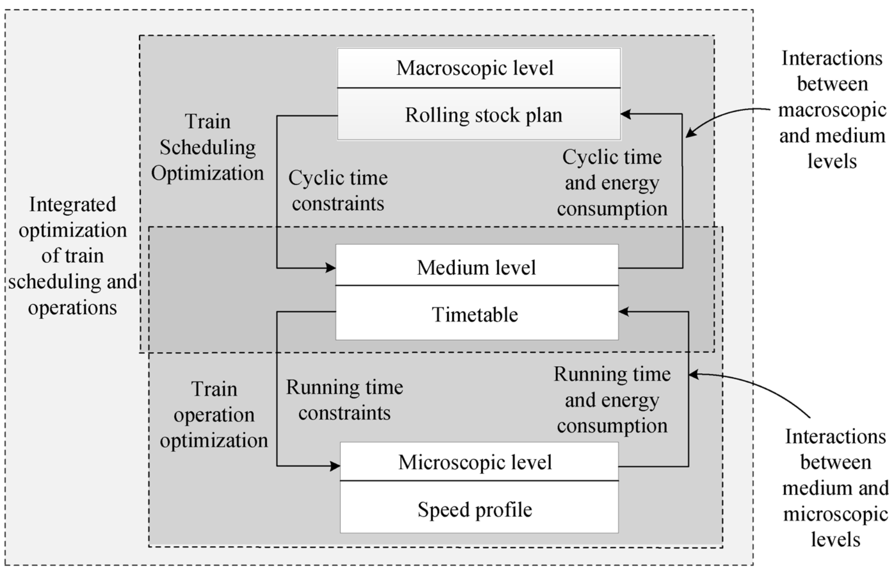


**Fig. 5** Integrated optimization of train scheduling and operations

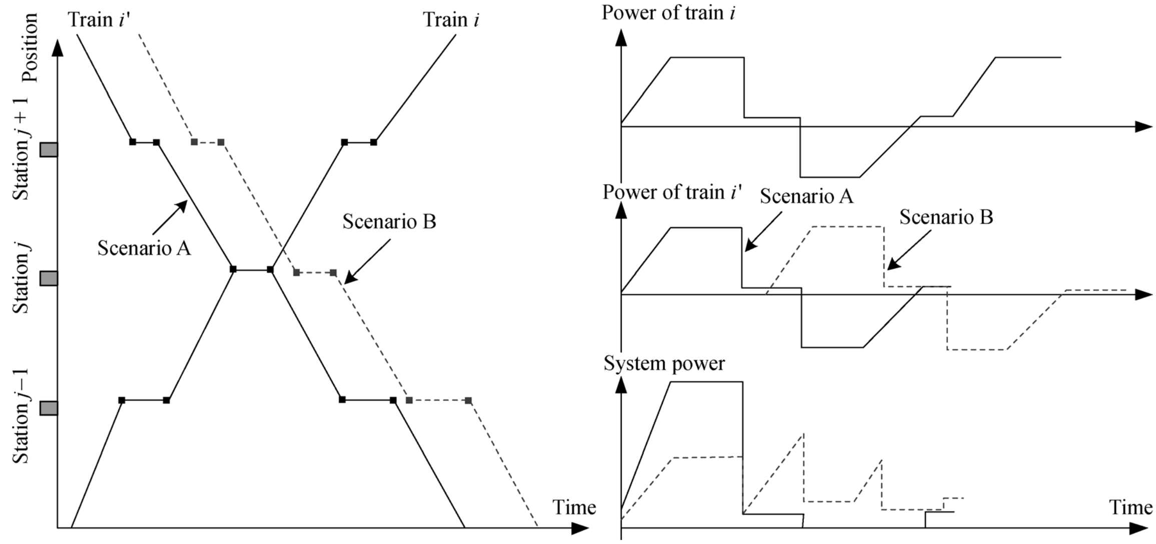


**Fig. 6** Improving the utilization of regenerative energy by matching the departure times of trains

with the aid of space–time networks.

In train control systems of urban rail transit, ATS and onboard ATO systems can generate several train speed profiles with different levels, which correspond to different trip times on each segment. The aim is to improve flexibility under different scenarios. For example, ATO speed profiles contain five levels, as shown in Fig. 7, where the operation switch points are different. If a train is delayed, the system can use a higher speed (i.e., level 1) to achieve punctuality. Otherwise, the system can use a lower speed (e.g., level 4 or 5) to slow down the train and save energy.

On the basis of this description, space–time network formulation for train scheduling is detailed as follows. The space–time network is an extension of the physical network in the time dimension. The concept of time axis is added based on the physical network, and a space–time network is generated. In detail, the time horizon is discretized into a set of equidistant time points to describe the network development process, and its corresponding physical nodes are generalized as space–time nodes. Each space–time node can describe two kinds of attributes, namely, space and time attributes of the node. The connection of the space–time nodes is called a space–time arc. Under multiple train running levels, a space–time network can be constructed to describe feasible train paths, as shown in Fig. 8.

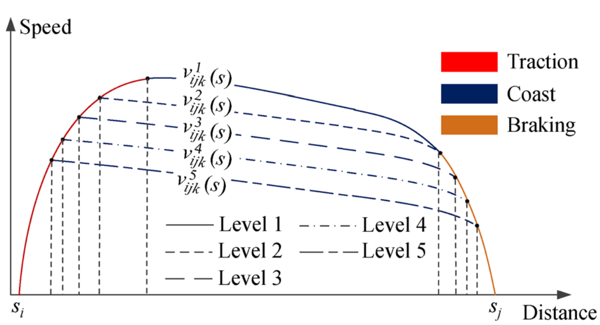


**Fig. 7** Speed profiles under different levels

Then, a space–time network can be developed to describe feasible space–time trajectories for the trains. The space–time network contains two types of space–time arcs, namely, train waiting and traveling arcs. An arc can be denoted by $(i,j,t,t')$, where $i$ and $j$ respectively represent the physical origin and destination of the arc, and $t$ and $t'$ respectively represent the time origin and destination of the arc. The space–time arc not only represents the connection between nodes $i$ and $j$ but also specifies the traveling time on this arc. Different space–time arcs starting from the same node can represent various running curves because distinct levels correspond to diverse travel times. In Fig. 8, the black solid, red dot, and blue dot lines denote the train dwelling arc, traveling arc of level 1, and traveling arc of level 2 in the space–time network, respectively. In addition, all of the traveling trajectories can be represented by these types of arcs in the space–time network. If the space–time trajectories of trains are plotted in the space–time network under safety constraints, a feasible train schedule can be generated. This schedule considers not only train arrival and departure times at each station but also the detailed speed profiles on each segment. In this way, the integrated train scheduling and operation problem can be transformed into a space–time trajectory optimization problem.

Based on the above description, we can denote a set of variables to represent the total energy consumption and regenerative energy in case that the train timetable is determined, given as follows.

$$F_a(u,t) = \sum_{k \in K} \sum_{(i,j,\tau,\tau') \in A, \tau \leqslant t, \tau' > t}$$

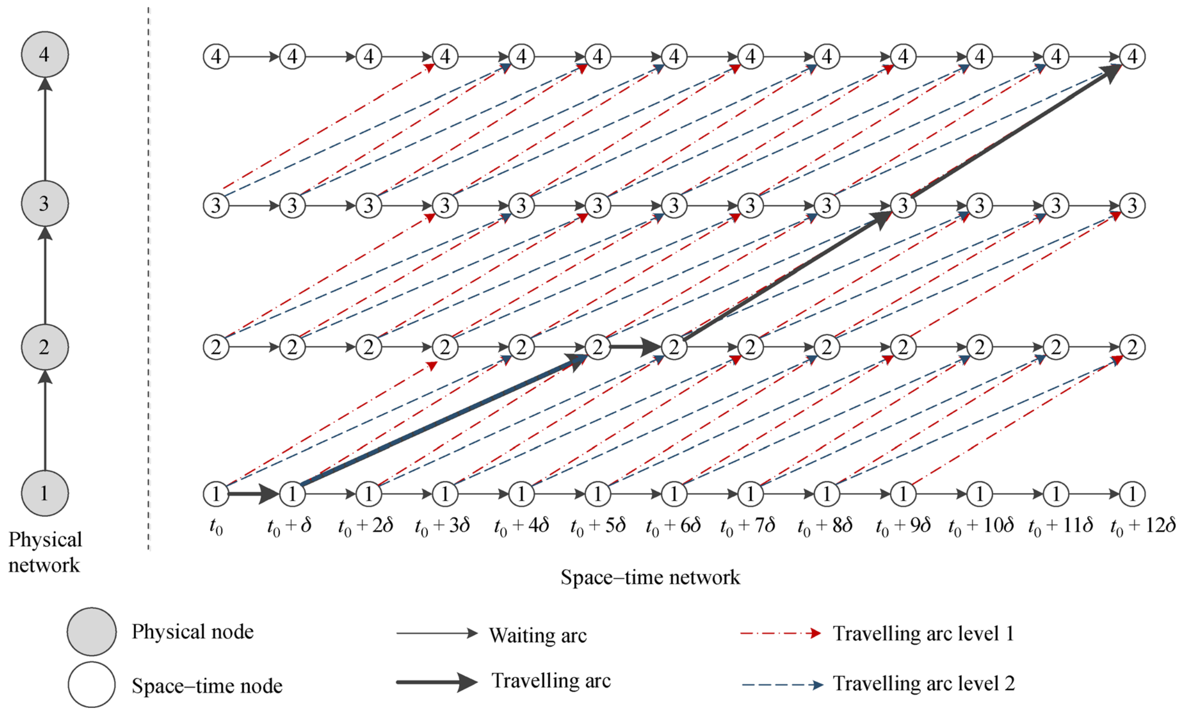


**Fig. 8** Illustration of the space–time network

$$\left[m_k\phi(i,j,u)x^k_{ij\tau\tau'}\int_t^{t+\delta}F^k_{ij\tau\tau'}(y)v^k_{ij\tau\tau'}(y)\mathrm{d}y\right],$$

$$R_b(u,t)=\sum_{k\in K}\sum_{(i,j,\tau,\tau')\in A,\tau\leqslant t,\tau'>t}$$

$$\left[m_k\phi(i,j,u)x^k_{ij\tau\tau'}\int_t^{t+\delta}B^k_{ij\tau\tau'}(y)v^k_{ij\tau\tau'}(y)\mathrm{d}y\right],$$

where $F_a(u,t)$ and $R_b(u,t)$ represent the total traction energy consumption and regenerative energy of all the trains under power supply substation $u$ in time interval $[t,t+\delta]$, respectively. In addition, $K$ and $A$ represent the sets of trains and space–time arcs, respectively. $m_k$ is the train weight, and $\phi(i,j,u)$ is a binary indicator showing whether section $(i,j)$ is in power supply substation $u$. $x^k_{ij\tau\tau'}$ is the binary decision variable indicating whether train $k$ chooses space–time arc $(i,j,\tau,\tau')$. $F^k_{ij\tau\tau'}(y)$ and $B^k_{ij\tau\tau'}(y)$ represent the traction and braking forces of train $k$ on the space–time arc, respectively. $v^k_{ij\tau\tau'}(y)$ represents the travelling velocity of train $k$ on this arc. In this manner, we can calculate the utilized regenerative energy in power supply substation $u$ and time unit $t$, which is the minimum value between regenerative energy and required traction energy, that is,

$$G_u(t)=\min\{F_a(u,t),\ c_aR_b(u,t)\},$$

where $G_u(t)$ represents the utilized regenerative energy on power supply substation $u$ and time unit $t$ and $c_a$ is the utilization rate of regenerative energy. Suppose that traction energy is $E_k(i,j,t,t')$ for train $k$ over space–time arc $(i,j,t,t')$. We can denote the objective function that minimizes the total energy consumption given the feasible region of decision variable $x^k_{ijtt'}$, that is,

$$Z=\sum_{k\in K}\sum_{(i,j,t,t')\in A}E_k(i,j,t,t')x^k_{ijtt'}-\sum_{t\in T}\sum_{u\in U}G_u(t),$$

where $T$ and $U$ represent the sets of time units and power supply substations, respectively. The first term in this objective function represents the total traction energy consumption of all the trains, and the second one denotes the utilized regenerative energy. This objective function can be formulated into a mixed integer linear programming model that can be successfully solved by commercial solvers, such as CPLEX and LINGO. In addition, several other constraints, including network flow balance, train headway, train dwelling time, rolling stock, and train capacity, are required to obtain a feasible solution for practical applications. Yin et al. (2017) provided further details on this subject.

## 6 Future research

With the continuous growth of passenger demands in large cities of China, urban rail transit managers usually improve the service frequency of lines by compressing the train headway in order to enhance the line capacity. However, if train schedule is not rigorously optimized, the overall line capacity, electricity power, rolling stock and crew resources will be evidently affected. Therefore, more research should be conducted to analyze the intrinsic characteristics of the metro line, the network origin-destination passenger flow, and the coupling relationship between trains, passenger flow and energy flow based on the actual operational data of the rail transit network. Moreover, demand-oriented timetable optimization in a whole rail transit network should be promising to improve the operational service level while reducing operating energy consumption and providing in depth theoretical support for safety, efficiency and energy-saving operations. Specifically, further research can include the following aspects:

(1) In practice, the uncertain passenger demand could affect train weight and accelerating and braking rates, leading to stochastic variations in traction energy consumption and regenerative energy. Thus, the dynamic and uncertain characteristics in operations should be further considered, and a robust train scheduling approach combined with speed profiles should be formulated. This could evidently reduce the decision risks in the uncertain operation environment.

(2) At present, train scheduling research that involves regenerative ESDs remains in the initial stage. Thus, it is meaningful to consider the utilization and storage of the regenerative energy, explicitly analyze the mechanisms of transmission, storage and utilization, and further apply these new technologies into the optimization of train services for energy savings.

(3) Most existing literature focused on energy-efficient train operation on a single segment or line. Only a few studies addressed the train scheduling problem in an entire network. Therefore, combining train scheduling with flexible passenger flow control and energy management at the network level is another important research direction for realizing collaborative optimization of trains, passengers, and energy flows.



## References

Allegre A L, Bouscayrol A, Delarue P, Barrade P, Chattot E, El-Fassi S (2010). Energy storage system with supercapacitor for an innovative subway. IEEE Transactions on Industrial Electronics, 57(12): 4001–4012

Chang C S, Sim S S (1997). Optimizing train movements through coast control using genetic algorithms. IEE Proceedings-Electric Power Applications, 144(1): 65–73

Gu Q, Tang T, Cao F, Song Y D (2014). Energy-efficient train operation in urban rail transit using real-time traffic information. IEEE Transactions on Intelligent Transportation Systems, 15(3): 1216–1233

Howlett P G, Milroy I P, Pudney P J (1994). Energy-efficient train control. Control Engineering Practice, 2(2): 193–200

Huang Y R, Yang L X, Tang T, Gao Z Y, Cao F, Li K P (2018). Train speed profile optimization with on-board energy storage devices: A dynamic programming based approach. Computers & Industrial Engineering, 126: 149–164

Ichikawa K (1968). Application of optimization theory for bounded state variable problems to the operation of train. Bulletin of the JSME, 11 (47): 857–865

Ke B R, Chen M C, Lin C L (2009). Block-layout design using MAX–MIN ant system for saving energy on mass rapid transit systems. IEEE Transactions on Intelligent Transportation Systems, 10(2): 226–235

Ke B R, Lin C L, Yang C C (2012). Optimization of train energy-efficient operation for mass rapid transit systems. IET Intelligent Transport Systems, 6(1): 58–66

Li X, Lo H (2014). An energy-efficient scheduling and speed control approach for metro rail operations. Transportation Research Part B: Methodological, 64: 73–89

Liu P, Yang L X, Gao Z Y, Huang Y R, Li S K, Gao Y (2018). Energy-efficient train timetable optimization in the subway system with energy storage devices. IEEE Transactions on Intelligent Transportation Systems, 19(12): 3947–3963

Liu R, Golovitcher I M (2003). Energy-efficient operation of rail vehicles. Transportation Research Part A: Policy and Practice, 37 (10): 917–932

Ma C Y, Ding Y, Du P, Mao B H (2010). Study on coast control of train movement for saving energy based-on genetic algorithm. Railway Computer Application, 19(6): 4–8

Rao Z (2006). Train Traction Calculation. Beijing: China Railway Press

Wang Y (2016). Calculation of Train Operations in Urban Metro Systems. Beijing: Science Press

Yang L X, Li K P, Gao Z Y, Li X (2012). Optimizing trains movement on a railway network. Omega, 40(5): 619–633

Yang X (2016). Research on train timetable optimization for energy-saving operations in urban rail transit. Dissertation for the Doctoral Degree. Beijing: Beijing Jiaotong University

Yang X, Li X, Gao Z Y, Wang H W, Tang T (2013). A cooperative scheduling model for timetable optimization in subway systems. IEEE Transactions on Intelligent Transportation Systems, 14(1): 438–447

Yin J T, Yang L X, Tang T, Gao Z Y, Ran B (2017). Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. Transportation Research Part B: Methodological, 97: 182–213

Zhao L (2014). Research on metro timetable optimization model and algorithm based on regenerative braking. Dissertation for the Master Degree. Beijing: Beijing Jiaotong University

Zhao S (2014). Research and simulation of urban rail transit super capacitor energy storage system. Dissertation for the Master Degree. Changsha: Central South University of Forestry and Technology