Research on "Yi-Xin-Europe" China-Europe Train to Improve China's Import Trade Level Based on Big Data Technology

Hongyu Li[1] and Ruchao Pang[2*]

[1,2*]Department of Economics and Management, Cangzhou
Normal University; Hebei Cangzhou 061000, China
Hongyu.Li666@outlook.com

Corresponding author: Ruchao Pang, pangruchao2023@163.com

Abstract: In order to explore the effect of "Yi-Xin-Europe" China-Europe train in improving China's import trade level, this paper combines big data technology to study the "Yi-Xin-Europe" China-Europe train to improve China's import trade level. Moreover, this paper proposes an intelligent processing algorithm for China-Europe trains and trade data, which introduces LDPC codes with short code lengths into frequency-hopping communications to achieve high-bandwidth-efficient error control. In addition, this paper evaluates the reduction in the amount of computation brought by the half-rate invertible linear block code to the system and theoretically analyzes the throughput gain brought by the double-bit error correction and multi-bit error detection CRC decoder. Finally, this paper constructs a research model of "Yi-Xin-Europe" China-Europe train to improve China's import trade level based on big data technology. Through the research, it can be seen that the research model of China-Europe trains to improve China's import trade level based on big data technology proposed in this paper has a good effect.

1   INTRODUCTION

Existing research on the impact of transport corridors on trade mainly includes research on international transport corridors and domestic transport corridors. One is the study of international transport channels. In the literature [13], the authors believe that the construction of transportation corridors can shorten the transportation distance in Eurasia, enhance Jilin Province and international exchanges, and accelerate the development of an export-oriented economy. Moreover, they used the shortest path algorithm to determine the goods trade transportation channel with the shortest transportation distance and the shortest freight cost. In the literature [16], the authors use the extended trade gravity model to compare the trade potential before and after the connection of the China-Pakistan-Turkey international transport corridor. They came to the same conclusion: the

construction of transportation corridors can shorten the time and space of trade, and the trade potential of China and the countries along the route can be significantly improved.

Reference [6] constructed a stochastic frontier gravitational model, and studied before and after the opening of the Arctic waterway. It was found that the shipping distance could be significantly shortened and the efficiency of China's import and export trade would be improved. Reference [14] analyses the economic benefits of the intermodal transport of the Changji-Tu International Corridor, qualitatively analyses the transport distance, time, cost, tools and transport links of the channel, and uses the specific data to calculate the transport cost and time, and compares it with the old channel to find a new channel. It can reduce transportation time and transportation costs, avoid the problem of limited railway transportation capacity, and promote the development of foreign trade in Jilin Province. Literature [10] believes that the construction of the China-Mongolia Great Corridor is closely related to the development of foreign trade in the cities along the route. Unblocking the international transportation corridor can improve the investment environment, and the development of the China-Mongolia Corridor can strengthen the trade exchanges between Northeast Asia and Mongolia countries; and the continuous development of international trade can bring strong support for the construction of the China-Mongolia Corridor. Literature [15] believes that the logistics channel is based on the comprehensive transportation network, and gradually becomes the main artery of the flow of materials, connecting the international and domestic markets, and eventually becomes a major trade channel and a vital carrier of the national economic corridor. And the inefficiency of the transport system adversely affects economic development, especially in landlocked and transit developing countries. Reference [9] uses the time-cost-distance method to evaluate and compare the multimodal transport corridors between Northeast Asia and Central Asia. The multimodal transport infrastructure and service level need to be improved and improved. Literature [1] believes that the China-Europe freight train can effectively use the bilateral and multilateral mechanisms between China and the countries along the "Belt and Road" to promote cooperation between China and the countries along the "Belt and Road" in railways, customs, inspection and quarantine, etc. The border crossing procedures are simplified, the level of trade facilitation is improved, and the trade potential is improved, and then the extended gravity model is used for quantitative verification.

As one of the international transportation channels, the New Eurasian Continental Bridge starts from Lianyungang, China in the east, passes through Kazakhstan, Russia, Germany and other countries, and ends at the Port of Rotterdam in the Netherlands in the west. The Pacific Ocean, the Baltic Sea, the North Sea and the Mediterranean Sea are closely connected with the inland areas of China and Central Asia and European countries; it has also become a new transportation channel for trade between Japan, South Korea, Southeast Asia and Eastern Europe, Western Europe, West Asia, Central Asia and South Asia. For transit transportation, the literature [12] believes that the New Eurasian Continental Bridge and the Siberian Continental Bridge have a great competitive relationship, and the freight volume of the original Siberian Continental Bridge is diverted by the New Eurasian Continental Bridge. However, although the New Eurasian Continental Bridge has a more favorable geographical location and climatic conditions than the Siberian Continental Bridge, it also has many advantages such as a short transportation distance and a wider range of countries and regions. It seems to be more attractive to the source of goods, but after its opening Since then, there are still many problems that weaken its competitiveness in actual transportation, and there is still a long way to go to become a real "golden transportation channel" across Asia and Europe. In operation, the New Eurasian Continental Bridge faces problems such as competition with the Siberian Continental Bridge in transit transportation, obstacles in the customs clearance process, obstacles in the release of transportation capacity, and lag in logistics operations and informatization construction. Countermeasures that can be taken include: effectively mobilizing the role of international organizations, strengthening coordination and cooperation among countries and regions along the New Eurasian Continental Bridge, continuously optimizing the customs clearance process, improving the public logistics information platform of the New Eurasian Continental Bridge, strengthening container multimodal transport organizations, Improve the management level and so

on [3]. Reference [11] compared the Siberian Continental Bridge and pointed out that the main problems faced by the new Eurasian Continental Bridge are the lack of competitiveness in cost, the low degree of standardization of the whole transportation, and the need to strengthen international cooperation. Reference [18] believes that the transit capacity is limited, there are many checkpoints along the route, the customs clearance procedures are complicated, the return empty container rate is high, and the transportation time and freight rate are increased. These problems weaken the competitiveness of the New Eurasian Continental Bridge itself, and establish a unified Coordinating organization, improving reloading ability, and balancing round-trip containers are effective countermeasures to solve the problem. Reference [19] analyses the reasons for the obstruction of the operation of the New Eurasian Continental Bridge from the aspects of transportation costs, national gauges along the route, and logistics information services, and proposes to adjust the freight rate to reduce transportation costs and promote the implementation of the "standard gauge" of the New Eurasian Continental Bridge. and the construction of a public logistics information platform for the new Eurasian Continental Bridge. Literature [2] believes that transportation corridors will attract regional social and economic factors. With this corridor as the axis, and relying on closely attracting cities in the region, factors such as industry, population, resources, and towns will be concentrated, and further industrial belts will be formed. The role of infrastructure in facilitating trade and reducing trade costs is significant for China. Literature [8] studied the impact of the economic corridor transportation infrastructure construction proposed by the "Belt and Road" initiative on economic and trade development. The study found that transportation infrastructure construction will have a positive effect on the trade and economic development of China and major economic corridors. Literature [5] believes that the improvement of transportation infrastructure is beneficial to reduce the cost of regional trade, promote the expansion of market scale, promote the specialization of labor division, and improve the efficiency of regional trade through qualitative analysis; and build a gravity model with traffic density as the indicator of transportation infrastructure , the study found that the improvement of transportation infrastructure can significantly promote the development of regional trade. Reference [7] used the gravity model to represent the level of transportation infrastructure with the length index of highways in European countries. The study found that the difference in the quantity and quality of infrastructure resulted in different transportation costs, which had an impact on the trade of countries. Similarly, Gong Xinshu, literature [17] used the stochastic frontier gravity model to study the air freight volume and railway freight volume as indicators of the level of transportation infrastructure. The empirical results show that a developed transportation network and a sound transportation infrastructure Facilities can significantly reduce trade costs and improve trade efficiency. And poorer-quality, less-quantitative infrastructure hinders trade through increased transport costs. Every 1 percentage point increase in infrastructure will increase exports by about 0.6 percentage points and imports by about 0.3 percentage points. Reference [4] analyzes and estimates the transportation time before and after the "Belt and Road" initiative based on the latest data of the "Belt and Road" transportation project, and converts the change in transportation time into a country-sector level of ad valorem trade costs. change value. The survey results show that the construction of transportation infrastructure brought about by the Belt and Road Initiative will greatly reduce transportation time and trade costs.

This paper combines big data technology to study the "Yi-Xin-Europe" China-Europe train to improve China's import trade level, and promote the steady improvement of China's import trade level.

2   INTELLIGENT PROCESSING ALGORITHM OF CHINA-EUROPE TRAIN AND TRADE DATA

2.1   Relevant Principles and Technologies Used in the Program

The linear block codes of code rate $R = \frac{1}{2}$ are called "half rate" linear block codes, and the half rate BCH codes are reversible. In this section, its reversibility is extended to general linear block codes, and the conditions and construction methods for the reversibility of half-rate linear block codes are given.

$C_B(2k,k)$ represents a half-rate binary linear block code with a code length of 2k, and $\mathbb{F}_2$ is a binary Galois Field (GF). The information frame $x = [x_1, x_2, \cdots, x_k] \in \mathbb{F}_2^k$ is encoded by $C_B(2k,k)$ to obtain the systematic code $c = [x, p]$, where $p = [p_1, p_2, \cdots, p_k] \in \mathbb{F}_2^k$ is the check frame of x. If $H = [A_{k \times k}, B_{k \times k}]$ is the check matrix of $C_B(2k,k)$, then there is:

$$0_{k \times 1} = Hc^T$$
$$= [A, B] \begin{bmatrix} x^T \\ p^T \end{bmatrix} \qquad (1)$$
$$= Ax^T + Bp^T$$

From formula (1), the relationship between xp is obtained as:

$$p^T = B^{-1}Ax^T \qquad (2)$$

$$x^T = A^{-1}Bp^T \qquad (3)$$

The formula (2) is the coding formula of the $C_B(2k,k)$-system code, and the condition of this formula is that the matrix B is full rank. On the contrary, formula (3) is the inverse coding formula, that is, the information frame x is reversed from the known check frame p. This chapter calls this operation "reversal operation", and the condition for the inversion operation to be established is that the matrix A is full rank. Therefore, the necessary and sufficient condition for $C_B(2k,k)$ to be invertible is that both A and B in matrix $H = [A_{k \times k}, B_{k \times k}]$ are full rank.

"Single-bit error correction multi-bit error detection" CRC decoder can only achieve 1-bit error correction. In order to further enhance the error correction ability of CRC code, this section uses soft information and multi-bit error detection characteristics of CRC code to extend its error correction ability.

$C_R(k,r)$ represents the CRC code with the code length k and the check length r, $u = [u_1, u_2, \cdots, u_{k-r}]$ is the k-r bits of information to be encoded, and $v = [v_1, v_2, \cdots, v_r]$ is the CRC check of u. After the user sends $x = [u, v]$, the receiver receives and demodulates the likelihood

ratio information $y = [y_1, y_2, \cdots, y_k]$, and the hard decision of y is $\hat{x} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_k]$, where the absolute value of y_i indicates the reliability of $\hat{x}_i$.

2.2 Description of HARQ Scheme Based on Half-Rate Invertible Linear Block Code

This section presents the workflow of a high-performance second-class HARQ scheme based on half-rate invertible linear block codes. The data frame format in the scheme is shown in Figure 1. Among them, PN is a pseudo-random sequence with length L and controlled seed, which is not involved in coding and is mainly used for jamming detection. The seed is controlled to avoid sending the same information repeatedly in the data header of each frame, and the PN is known at the receiving end. Therefore, it can also be used by the demodulator as channel estimation and estimation of parameters such as carrier offset Δf, initial phase $\hat{\theta}$ and timing offset $\Delta \tau$.
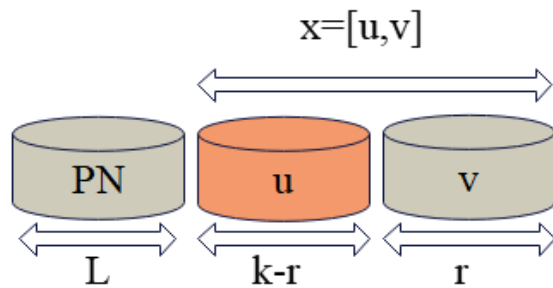
(a) Frame format of the information frame
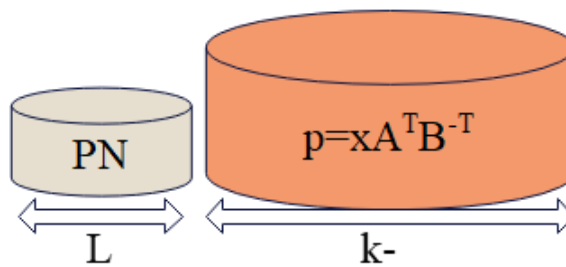
(b) Frame format of the verification frame

Figure 1: Data frame format and its relationship.

In Figure 1, the information frame $x = [u, v]$ of length k is the code word based on $C_R(k, r)$, p is the check frame obtained by x through $C_B(2k, k)$ system code encoding, and the length is also k. $C_R(k, r)$ is used in the scheme of this chapter for "double-bit error correction and multi-bit error detection", which can correct up to 2-bit errors.

1. Human interference detection

$$PN = [m_1, m_2, \cdots, m_L]$$ represents a pseudo-random sequence of length L, and PN is a known quantity at the receiving end. The pseudo-random sequence is modulated and transmitted, and the hard-decision sequence obtained by the receiver's demodulation is $PN = [\hat{m}_1, \hat{m}_2, \cdots, \hat{m}_L]$. The human interference detection algorithm is to calculate its Hamming distance $d = dist(PN, PN)$. If d is greater than a certain threshold, it is considered that there is human interference in the frame signal. Because the bit error rate of the frame data is close to $\frac{1}{2}$ when subjected to human interference, so d is also larger.

2. Equal gain package merge

The receiving end receives a frame of data and considers that there is no interference after interference detection, and the frame of data is called an effective transmission frame. $\tilde{y}$ represents the soft information obtained by combining the equal-gain packets after x has undergone i-1 effective transmissions (retransmissions), and y represents the soft information received during the i-th effective transmission (retransmission). Then, the new soft information obtained after equal-gain packet merging is:

$$\tilde{y} \leftarrow \frac{1}{i}[(i-1)\tilde{y} + y]$$

(4)

Equal-gain packet merging is simple and effective, and effectively utilizes the performance gain brought by multiple retransmissions at low signal-to-noise ratios.

The workflow of the second type of HARQ scheme based on the half-rate reversible linear block code is shown in Figure 2.
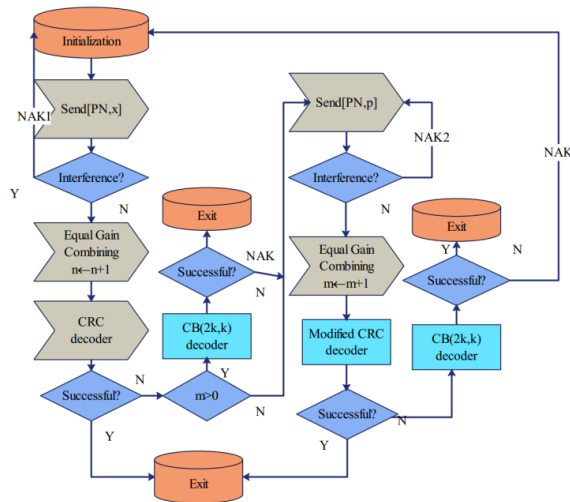
Figure 2: Flowchart of the second type of HARQ based on half-rate reversible linear block codes.

In Figure 2, the process of processing information frames in the left dashed box is defined as state I (s_1), the process of processing check frames in the middle dashed box is defined as state II (s_2), and the decoding process of linear block code is defined as state III (s_3). s_e means exit status. According to Algorithm 2, when the probability of human interference is $\epsilon = 0$, the state transition diagram and transition probability of the second type of HARQ working process of the HARQ scheme based on the half-rate reversible linear block code are shown in Figure 3. In the figure, $p_f^x(n)$ and $p_f^p(m)$ respectively represent the frame error rate of the n-th effective transmission of the information frame and the m-th effective transmission of the check frame, and $p_{Be}(k)$ represents the probability that the linear block code decoder fails in the k-th decoding.
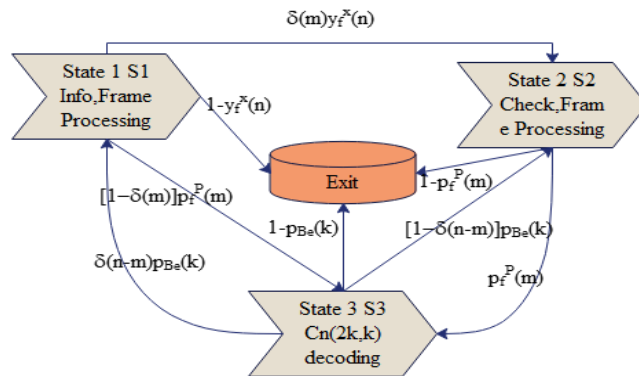
Figure 3: State transition diagram and transition probability in Algorithm 2 when the probability of human interference is $\varepsilon = 0$.

When the maximum number of retransmissions is ∞ and the system achieves successful transmission, the average number of valid frames (including information frames and check frames) received by the receiver can be expressed as:

$$N_{eff} = \underbrace{\sum_{n=1}^{\infty}(2n-1)q_1(n)p(s_1 \to s_e) + \sum_{n=2}^{\infty}(2n-1)q_1(n)p(s_1 \to s_3)p(s_3 \to s_e)}_{s_1 Go\ to\ success\ in\ status}$$

$$+ \underbrace{\sum_{m=1}^{\infty}2m\left[q_2(m)p(s_2 \to s_e)+q_2(m)p(s_2 \to s_3)p(s_3 \to s_e)\right]}_{s_2 Go\ to\ success\ in\ status} \tag{5}$$

$$q_1(n+1)=\begin{cases}1, n=0 & (6)\\ q_2(n)p(s_2 \to s_3)p(s_3 \to s_1), n \geq 1 & (7)\end{cases}$$

And

$$q_2(m+1) = \begin{cases} q_1(1)p(s_1 \to s_2), m = 0 & (8) \\ q_1(m)p(s_1 \to s_3)p(s_3 \to s_2), m \geq 1 & (9) \end{cases}$$

After calculating and obtaining N_{eff} according to the above formula, when the probability of human interference is ϵ, the average number of frames sent by the sender is:

$$\bar{N} = \frac{1}{1-\epsilon} N_{eff} \qquad (10)$$

When the receiver receives a valid frame, the sender has (in the average sense) transmitted $\dfrac{1}{1-\epsilon}$ times, of which $\dfrac{\epsilon}{1-\epsilon}$ is discarded by the receiver due to artificial interference.

The physical meaning of $\bar{N}$ is: When the receiving end receives a frame of information X without errors, the transmitting end needs to transmit $\bar{N}$ frames of data on average (including information frames and check frames). Therefore, when the bandwidth occupied by ACK and NAK transmission is ignored, the average bandwidth efficiency η of the scheme is obtained as:

$$\eta = \frac{1}{\bar{N}} \frac{k-r}{k} = \frac{1}{N_{eff}} \left(1 - \frac{r}{k}\right)(1-\epsilon) \qquad (11)$$

Among them, $1 - \dfrac{r}{k}$ is the coding efficiency of the CRC code.

At medium and high SNR, most frames can be successfully decoded by CRC decoder or successfully corrected by linear block code. That is, reliable transmission is achieved with a high probability through the following situations:

a) As long as X is transmitted once, the CRC decoding is successful, and $s_1 \to s_e$ is realized, and its probability is $q_1(1)p(s_1 \to s_e)$;

b) If the CRC decoding in a) fails, then $s_1 \to s_2$. The sender transmits P once, and the receiver performs the reverse operation $s_2 \to s_e$ and CRC decoding. If the decoding is successful, the algorithm exits and D is realized, and its probability is $q_2(1)p(s_2 \to s_e)$;

c) If CRC decoding fails after p is reversed in b), then $s_2 \rightarrow s_3$. The receiving end performs linear block code decoding, and if the decoding is successful, there is $s_3 \rightarrow s_e$, and its probability is $q_2(1)p(s_2 \rightarrow s_3)p(s_3 \rightarrow s_e)$;

d) If the linear block code decoding in c) fails, then $s_3 \rightarrow s_1$. The system needs to retransmit x times, and the receiving end does equal gain packet combination after receiving x. It assumes that the combined soft information is decoded by CRC and linear block code to achieve successful error correction with probability 1 and exit the algorithm. Successful error correction is implemented and the algorithm exits. That is, $p(s_1 \rightarrow s_e) + p(s_1 \rightarrow s_3)p(s_3 \rightarrow s_e) = 1$, and at moderate to high signal-to-noise ratios, the hypothesis holds with a high probability. The probability of this case is $q_1(2)\big(p(s_1 \rightarrow s_e) + p(s_1 \rightarrow s_3)p(s_3 \rightarrow s_e)\big) = q_1(2)$.

According to the above four situations, formula (5) can be simplified as:

$$N_{eff} \approx \underbrace{q_1(1)p(s_1 \rightarrow s_e)}_{Case\ a} + \underbrace{2q_2(1)p(s_2 \rightarrow s_e)}_{Case\ b} + \underbrace{2q_2(1)p(s_2 \rightarrow s_3)p(s_3 \rightarrow s_e)}_{Case\ c} + \underbrace{3q_1(2)}_{Case\ d} \qquad (12)$$

$$\begin{aligned} &= p(s_1 \rightarrow s_e) + 2p(s_1 \rightarrow s_2)\big[p(s_2 \rightarrow s_e) + p(s_2 \rightarrow s_3)p(s_3 \rightarrow s_e)\big] \\ &\quad + 3p(s_1 \rightarrow s_2)p(s_2 \rightarrow s_3)p(s_3 \rightarrow s_1) \end{aligned} \qquad (13)$$

Under the same channel conditions, the probability that x and p contain un-correctable errors of the CRC code (respectively p_f^x and p_f^p) are the same and denoted by p_e, that is,

$$p(s_1 \rightarrow s_2) = \delta(m)p_e, p(s_2 \rightarrow s_3) = p_e \qquad (14)$$

$$p(s_1 \rightarrow s_e) = p(s_2 \rightarrow s_e) = 1 - p_e \qquad (15)$$

The probability $p(s_3 \rightarrow s_1) = p_{\text{Be}}(1)$ is related to the linear block code and decoding algorithm used. Generally, there is no closed expression, but its experience curve can be obtained through simulation.

By substituting formulas (14) and (15) into formula (13), we get:

$$\begin{aligned} N_{eff} &\approx 1 - p_e + 2p_e\big[(1 - p_e) + p_e \cdot p(s_3 \rightarrow s_e)\big] + 3p_e^2\big(1 - p(s_3 \rightarrow s_e)\big) \\ &= 1 + p_e + p_e^2 p(s_3 \rightarrow s_1) \end{aligned} \qquad (16)$$

The maximum error correction capability of the double-bit error-correcting multi-bit error-detecting CRC decoder is 2, so there is no exact expression for p_e. After the empirical values of $p(s_3 \rightarrow s_1)$

and p_e are obtained through simulation, the throughput of the HARQ protocol in this chapter under the medium and high signal-to-noise ratio can be obtained through formulas (16) and (11).

Under the same channel conditions, the residual misdiagnosis rates of x and p after CRC decoding are both p_e. Then, the probability that there is no need to start the linear block code decoding algorithm is:

$$P_u = p_e \left(1 - p_e\right) \tag{17}$$

If the reversible code is not used, as long as the information frame contains uncorrectable errors (regardless of whether the check frame is correctly transmitted or not), the decoding algorithm of the linear block code needs to be started, and the probability is:

$$P_v = p_e \tag{18}$$

Then the percentage of the saved operation amount to the original required operation amount is:

$$\rho = \frac{P_u}{P_v} = 1 - p_e \tag{19}$$

Since the maximum error correction capability of the double-bit error correction multi-bit error detection CRC decoder is 2 bits, there is no exact expression for p_e, and its upper and lower bounds under the $AWGN$ channel (its bit error probability is p_b) are:
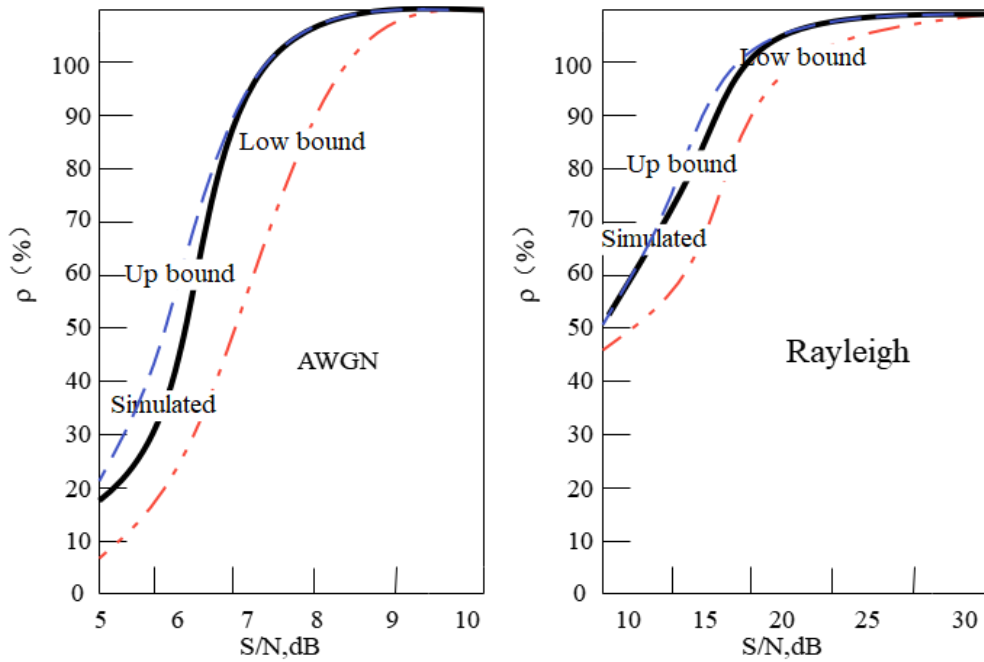
$$p_2 \leq p_e \leq p_1 \tag{20}$$

$$p_2 = p(|e| \geq 3) = 1 - \left(1 - p_b\right)^k - k p_b \left(1 - p_b\right)^{k-1} - \binom{k}{2} p_b^2 \left(1 - p_b\right)^{k-2} \tag{21}$$

$$p_1 = p(|e| \geq 2) = 1 - \left(1 - p_b\right)^k - k p_b \left(1 - p_b\right)^{k-1} \tag{22}$$

|e| represents the number of error bits. The upper and lower bounds of ρ under the AWGN channel are:

$$\left(1 - p_b\right)^k + k p_b \left(1 - p_b\right)^{k-1} \leq \rho \leq \left(1 - p_b\right)^k + k p_b \left(1 - p_b\right)^{k-1} + \binom{k}{2} p_b^2 \left(1 - p_b\right)^{k-2} \tag{23}$$

Although there is no exact expression of p_1, p_2 under Rayleigh channel, it can still be obtained by simulation.

Figure 4: Contribution curve of reversible codes to the reduction of decoding complexity of linear block codes.

Through simulation, it is obtained that under the $AWGN$ -channel and Rayleigh-fading channel, the reduction of the computation amount brought by the reversible code to the system accounts for the percentage of the original required computation amount (formula (19)) as shown in Figure 4. The upper and lower bounds in formula (23) are also drawn under the AWGN channel, and the upper and lower bounds are also obtained through simulation under the Rayleigh channel. Although P_v and P_u are very small at high signal-to-noise ratio, the curve in the figure shows that the reduction of computation amount brought by the use of reversible codes accounts for a considerable percentage of the originally required computation amount. The curve in the figure also shows that ρ is very close to its upper bound under high signal-to-noise ratio. Therefore, combining formulas (19) and (23) and Figure 4, replacing p_e with p_2 in the theoretical analysis can also obtain accurate conclusions.

The average number of frames to be transmitted is:

$$N_{eff} = 1 - p_e + 2p_e = 1 + p_e$$

(24)

1: CRC is only used for error detection throughput.

At this time there is $p_e = p(|e| \geq 1) = 1 - p(|e| = 0)$, so:

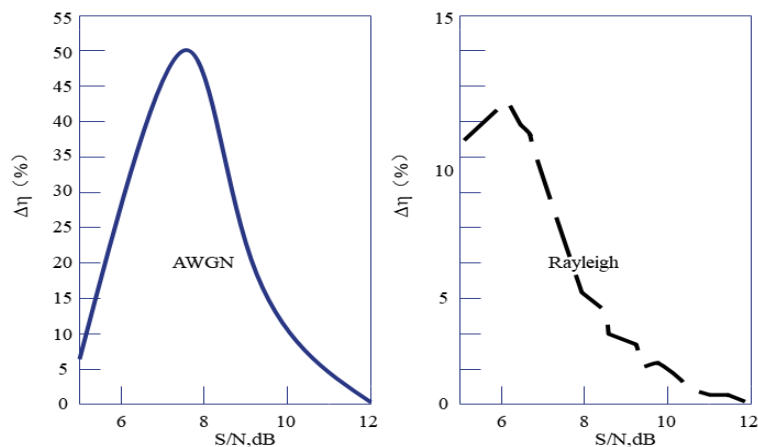$$N_{eff}^{1} = 1 + 1 - p(|e| = 0) = 2 - p(|e| = 0)$$

(25)

2: Throughput with double-bit error correction CRC decoding

We replace p_e with p_2 , which is $p_e = p(|e| \geq 3)$. At this time there is $p_e = p(|e| \geq 3) = 1 - p(|e| \leq 2)$, so:

$$N_{eff}^{2} = 1 + 1 - p(|e| \leq 2) = 2 - p(|e| \leq 2)$$

(26)

The ratio of the throughput increase value of scheme 2 relative to square 1 to square 1 is:

$$\Delta\eta = \frac{\eta_2 - \eta_1}{\eta_1} = \frac{N_{eff}^{1}}{N_{eff}^{2}} - 1$$

$$= \frac{2 - p(|e| = 0)}{2 - p(|e| \leq 2)} - 1$$

(27)

Figure 5: The throughput gain of the dual-bit CRC decoder for the HARQ scheme in this chapter at medium and high SNR.

Under the AWGN channel, if $p_0 = p(/e/=0) = 1 - (1-p_b)^k$, there is $\Delta\eta = \dfrac{2-p_0}{2-p_2} - 1$. p_0 and p_2 under the Rayleigh channel can be obtained by simulation. Figure 5 shows the simulation curve of $\Delta\eta$, which shows that the double-bit CRC decoder can effectively improve the throughput of the system under high signal-to-noise ratio. With the further increase of the signal-to-noise ratio, $\Delta\eta$ decreases. The reason for this is that both p_0 and p_2 gradually converge to 0.

2.3 Examples of Specific Schemes Using Different Linear Block Codes

$C_B(2k,k)$ adopts the half-rate de-reversible punctured BCH code to obtain a hybrid ARQ error control system based on BCH code. The main design points of the system are as follows:

We determine the codeword space $GF(q)$ where BCH is located, and set N=q-1. According to the system performance index, the error correction capability t of the BCH code is selected, and the $BCH(N,K)$ mother code with the error correction capability t is designed. Puncturing the $BCH(N,K)$ mother code to get the $BCH(N-\ell,K-\ell)$ shortened code with the code rate $\dfrac{1}{2}$, where $\ell = 2K - N$ and $K - \ell = k$.

g(x) represents the generator polynomial of the BCH code, m(x) and c(x) represent the information polynomial and the codeword polynomial, respectively. b(x) is the check polynomial of m(x), and $\rho(x)$ represents the cofactor polynomial obtained by dividing x^k by g(x), namely:

$$b(x) = x^k m(x) \bmod g(x) \tag{28}$$

$$\rho(x) = x^k \bmod g(x) \tag{29}$$

$$m(x) = b(x)\rho(x) \bmod g(x) \tag{30}$$

The reverse operation of BCH code can be replaced by formula (30), thus avoiding the matrix operation in equation (3).

x is the information frame, p represents the check frame of x based on LDPC code, x is also the CRC code word based on $C_B(k,r)$, and $\mathcal{S}(\hat{x})$ represents the syndrome of $\hat{x}$. e_i represents a single-bit error vector. That is, only the i-th position in e_i is 1 and other positions are 0. $A^{-1}B$ in

the formula (3) is represented by D. If it is assumed that p contains a single-bit error, and $\hat{x} = D(p + e_i)$, then:

$$\begin{aligned}
\mathcal{S}(\hat{x}) &= \mathcal{S}(D(p + e_i)) \\
&= \mathcal{S}(x + De_i) \\
&= \mathcal{S}(De_i)
\end{aligned}$$

(31)

Therefore, when performing single-bit error correction and multi-bit error detection CRC decoding on the check frame p, TAB records the one-to-one mapping relationship between $\mathcal{S}(De_i)$ and e_i. However, when the information frame x is CRC decoded, the TAB records the mapping relationship between $\mathcal{S}(e_i)$ and e_i.

The LDPC code used in the simulation is an irregular code, and the code length parameters are shown in Table 2 (see the next page). The distribution of the code's metric pair (λ, ρ) is (edge description):

$$\lambda(x) = 0.262287x + 0.214265x^2 + 0.0956022x^3 + 0.427846x^{10}$$
$$\rho(x) = 0.4301x^6 + 0.5699x^7$$

According to (λ, ρ), the corresponding check matrix is generated by PEG algorithm, and the check matrix is permuted to make it satisfy the full rank condition. The decoder of the LDPC code in the simulation adopts the SPA algorithm, and the maximum number of iterations in the algorithm is set to 50 times.

Figure 6 shows the numerical simulation curves and theoretical analysis results of the throughput of the two example schemes under different channels when the jamming probability is $\epsilon = 0$. The theoretical analysis results (Identified in the figure as BCH theoretic analysis and LDPC theoretic analysis) of the two schemes in different channels in the figure are obtained by first obtaining the empirical values of p_e and $p(s_3 \rightarrow s_1)$ through simulation, and then calculating η according to equations (11) and (16) and drawing the obtained curve.

The curve in Figure 6 shows that the simulation curve is very close to the theoretical analysis results under the medium and high SNR, which fully verifies the correctness of the theoretical analysis. Compared with the two simulation curves of LDPC-HARQ and BCH-HARQ in Figure 6, their performances are very close at high signal-to-noise ratio. Under high signal-to-noise ratio, error correction can be achieved only by a double-bit error correction CRC decoder in most cases, and a linear block code decoder is rarely used. Therefore, its performance is mainly determined by the double-bit error correction CRC decoding algorithm. The LDPC-HARQ scheme has a performance gain of 1~2dB at low SNR, which shows the strong error correction ability of LDPC codes.
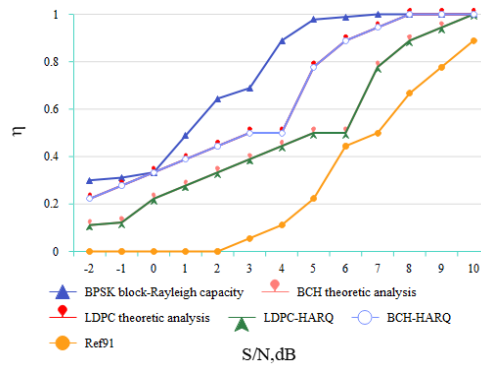
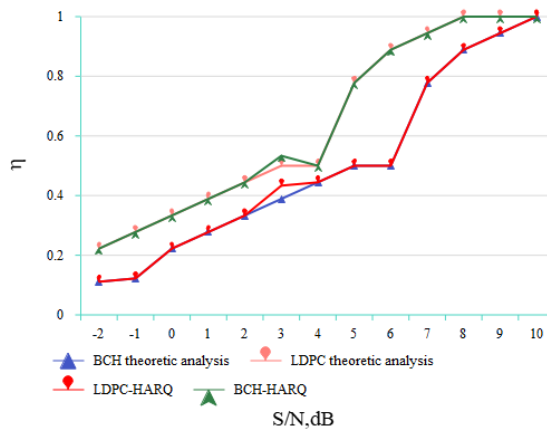(a) Simulation results and theoretical analysis curve under the AWGN channel
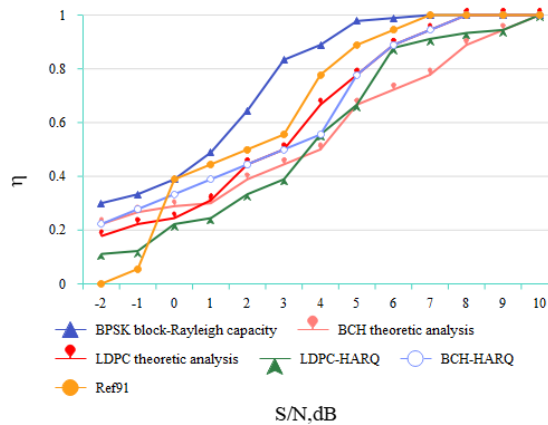
(b) Simulation results and theoretical analysis curve under the Rayleigh block fading channel
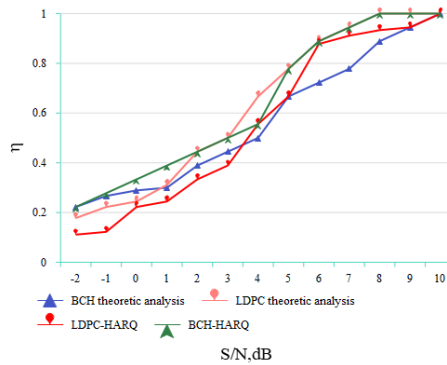
Figure 6: Throughput of BCH-HARQ and LDPC-HARQ under different channels when $\varepsilon = 0$.

Figure 6 also gives the Shannon limit of the BPSK signal under the AWGN channel and the Rayleigh fading channel (indicated as BPSK-AWGN capacity and BPSK block-Rayleigh capacity in the figure). For comparability, this limit also takes into account the coding efficiency of the CRC code. The LDPC-HARQ scheme is very close to the Shannon limit in both channels, especially in the Rayleigh block fading channel, its spectral efficiency is very close to the theoretical limit, which shows the high efficiency of the scheme in this chapter.

In addition, Figures 7 and 8 show the throughput simulation curves and theoretical analysis results when the human interference probabilities are $\in = 0.1$ and $\in = 0.2$, respectively. These two figures also fully illustrate the effectiveness of the scheme in this chapter against human interference and the feasibility of reliable data transmission in harsh communication environments.
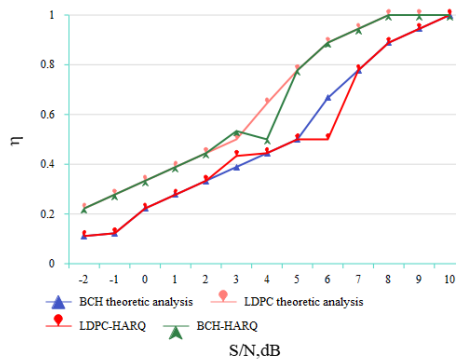
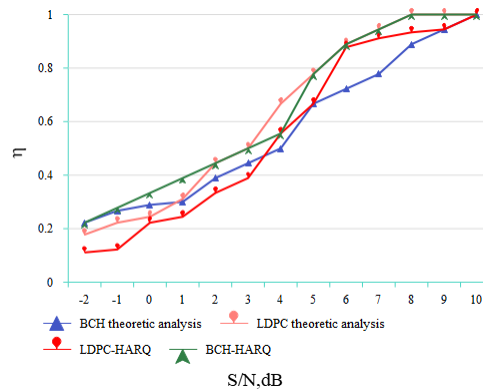(a) Simulation results and theoretical analysis curve under the AWGN channel

(b) Simulation results and theoretical analysis curve under the Rayleigh block fading channel

Figure 7: Throughput of BCH-HARQ and LDPC-HARQ under different channels when $\in$ =0.1

(a) Simulation results and theoretical analysis curve under the AWGN channel

(b) Simulation results and theoretical analysis curve under the Rayleigh block fading channel

Figure 8: Throughput of BCH-HARQ and LDPC-HARQ under different channels when $\in =0.2$.

The Shannon limits of BPSK signals under different channels are calculated as follows. $\mathcal{X}=\{-1,+1\}$ represents the set of BPSK symbols, and $\mathcal{Y}$ represents the set of channel outputs.

When the BPSK symbols satisfy the uniform distribution $\left(p(x=-1)=p(x=1)=\dfrac{1}{2} \right)$, the mutual information of the discrete binary input continuous output channel is:

$$I(\mathcal{X};\mathcal{Y})=\frac{1}{2}\sum_{x\in\mathcal{X}}\int_{\mathcal{Y}} p(y/x)log\frac{p(y/x)}{p(y)}dy \tag{32}$$

For AWGN channels, there are:

$$I(\sigma)=-\int_{-\infty}^{\infty}\phi_{\sigma}(y)log\,\phi_{\sigma}(y)dy-\frac{1}{2}log\left(2\pi e\sigma^{2}\right) \tag{33}$$

Among them,

$$\phi_{\sigma}(y)=\frac{1}{\sqrt{8\pi\sigma^{2}}}\left(e^{-\frac{(y-1)^{2}}{2\sigma^{2}}} + e^{-\frac{(y+1)^{2}}{2\sigma^{2}}} \right) \tag{34}$$

According to formula (33), the information rate (Shannon limit) $C=I(\sigma)$ corresponding to different signal-to-noise ratios under the AWGN channel can be obtained. Similarly, the noise power required to achieve the information rate can be obtained from the information rate C:

$$\sigma^{2}=\left(I^{-1}(C)\right)^{2} \tag{35}$$

Among them, $I^{-1}(C)$ is the inverse function of $I(\sigma)$, which can be obtained by numerical method. Under Rayleigh fading channel, if the frame length of the physical layer is k, the relationship between input and output is:

$$y_1^k = hx_1^k + n_1^k \qquad (36)$$

$n_1^k = [n_1, n_2, \cdots, n_k]$ is a Gaussian vector of length k, where n_i and $n_j(i \neq j)$ are independent and identically distributed, and both satisfy $\mathcal{N}(0, \sigma^2)$. The fading coefficient h remains unchanged within a frame, and is different from frame to frame, and satisfies the Rayleigh distribution:

$$p(h) = he^{-\frac{h^2}{2}} \qquad (37)$$

$\mathcal{X}^k$ represents the value space of $x = [x_1, x_2, \cdots, x_k]$, namely $x \in \mathcal{X}^k$. Then, under the Rayleigh fading channel, formula (32) can be written as:

$$I(\mathcal{X}; \mathcal{Y}) = \frac{1}{2^k} \sum_{x \in \mathcal{X}^k} \int_0^\infty p(h) \int p(y/x, h) \log \frac{p(y/x, h)}{p(y/h)} dy dh \qquad (38)$$

Like the AWGN channel, the Rayleigh channel is a symmetrical channel, that is, the bit error rate of the channel has nothing to do with the input. It can be assumed that the input is all "1" signals, and formula (38) can be simplified as:

$$I(\mathcal{X}; \mathcal{Y}) = \int_0^\infty p(h) \int p(y/1, h) \log \frac{p(y/1, h)}{p(y/h)} dy dh \qquad (39)$$

We define:

$$\alpha(h, y) = \log \frac{p(y/1, h)}{p(y/h)} \qquad (40)$$

Then, by means of simulation, N vector pairs $[h_j, y_j]$ can be generated according to the distribution pair $[p(h), p(y/1, h)]$, and the s $\alpha(h_j, y_j)$ of each vector pair can be calculated. As long as N is large enough, there is:

$$I(\mathcal{X}; \mathcal{Y}) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \alpha(h_j, y_j) \qquad (41)$$

According to formula (41), the Shannon limit under Rayleigh block fading channel is obtained by simulation.

3 RESEARCH ON "YI-XIN-EUROPE" CHINA-EUROPE TRAIN TO IMPROVE CHINA'S IMPORT TRADE LEVEL BASED ON BIG DATA TECHNOLOGY

Under the category of transportation infrastructure, the mechanism for analyzing the impact of the new Eurasian Continental Bridge on trade between China and Central Asia. The New Eurasian Continental Bridge specifically affects the trade between China and Europe in two ways, as shown in Figure 9. One is that with the reduction of transportation costs and the shortening of transportation time, it can directly promote trade growth by reducing transportation costs and improving transportation efficiency. The second is to indirectly promote trade growth by promoting regional economic agglomeration and diffusion.
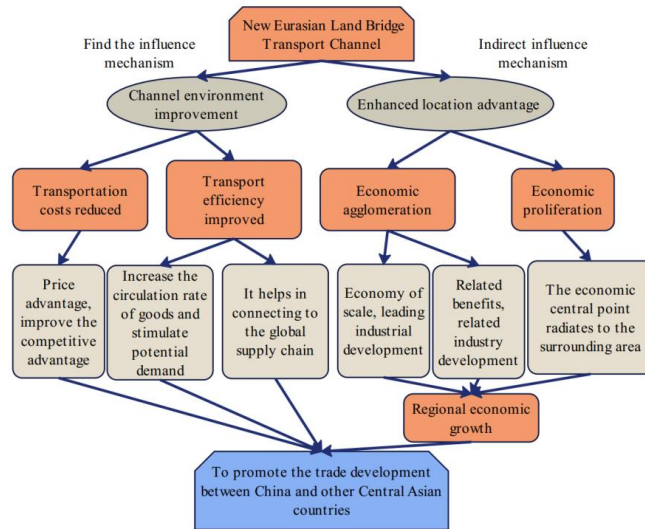
Figure 9: Research model of China-Europe trains to improve China's import trade level based on big data technology.

The top five import commodity structures of countries along the route from Yiwu to "Yi-Xin-Europe" China-Europe train are machinery and appliances, miscellaneous products, beverages and cigarettes, metal products, chemical products and related products, mainly production materials, daily consumer goods, and the main way of import trade is general trade. The goods imported through the "Yi-Xin-Europe" China-Europe train mainly include alcohol, edible oil, maternal and child products, high-end kitchenware, auto parts, etc., as well as cross-border e-commerce goods.

On the basis of the above model, the effect of the research model proposed in this paper to improve China's import trade level based on big data technology is evaluated. This paper takes the improvement of China's import trade level by the "Yi-Xin-Europe" China-Europe train as the research object, and analyses multiple sets of data to obtain the results shown in Figure 10.

Through the above research, we can see that the model of China-Europe trains to improve China's import trade level based on big data technology proposed in this paper has a good effect.

4 CONCLUSION

Compared with road transportation, which is also a land transportation method, railway transportation has irreplaceable characteristics such as economical environmental protection and large single transportation volume.
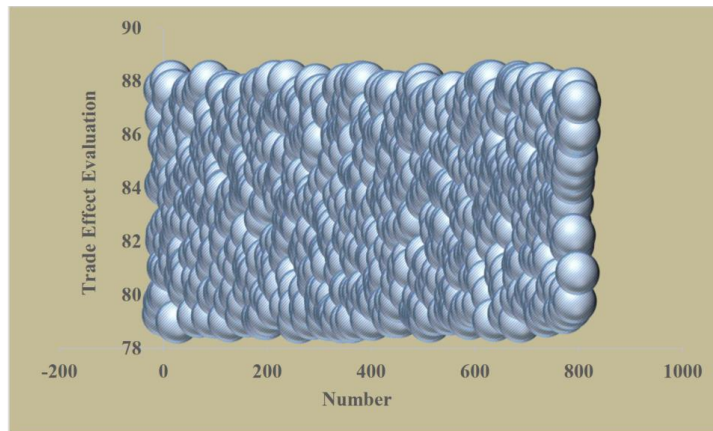
Figure 10: Evaluation of the effect of the research model for China-Europe freight trains to improve China's import trade level based on big data technology.

Therefore, for a long time, railways have played an important role in China's land transportation and even the entire transportation system. At the same time, the railway is also an important channel for China's land trade. Its "One Belt, One Road" initiative aimed at enabling more neighbouring countries to achieve more prosperous bilateral trade. This paper combines big data technology to study the "Yi-Xin-Europe" China-Europe train to improve China's import trade level, and promote the steady improvement of China's import trade level. Through the research, it can be seen that the research model of China-Europe trains to improve China's import trade level based on big data technology proposed in this paper has a good effect.

Hongyu Li, https://orcid.org/0009-0003-8691-1226
Ruchao Pang, https://orcid.org/0009-0004-2613-4717

REFERENCES

[1] Choi, S.; Chung, H.; Kim, J.; Ahn, J.; Kim, I.: Mobile Hotspot Network System for High-Speed Railway Communications Using Millimeter Waves, ETRI Journal, 38(6), 2016,1052-1063. https://doi.org/10.4218/etrij.16.2716.0018

[2] Gao, M.; Ai, B.; Niu, Y.; Wu, W.; Yang, P.; Lyu, F.; Shen, X.: Efficient hybrid beamforming with anti-blockage design for high-speed railway communications, IEEE Transactions on Vehicular Technology, 69(9), 2020, 9643-9655. https://doi.org/10.1109/TVT.2020.3000757

[3] He, R.; Zhong, Z.; Ai, B.; Oestges, C.: Shadow fading correlation in high-speed railway environments, IEEE Transactions on Vehicular Technology, 64(7), 2014, 2762-2772. https://doi.org/10.1109/TVT.2014.2351579

[4] He, R.; Zhong, Z.; Ai, B.; Ding, J.; Yang, Y.; Molisch, A. F.: Short-term fading behavior in high-speed railway cutting scenario: Measurements, analysis, and statistical models, IEEE Transactions on Antennas and Propagation, 61(4), 2012, 2209-2222. https://doi.org/10.1109/TAP.2012.2235399

[5] He, R.; Zhong, Z.; Ai, B.; Guan, K.: Reducing the cost of high-speed railway communications: From the propagation channel view, IEEE Transactions on Intelligent TransportationSystems, 16(4), 2015, 2050-2060. https://doi.org/10.1109/TITS.2015.2390614

[6] He, R.; Zhong, Z.; Ai, B.; Wang, G.; Ding, J.; Molisch, A. F.: Measurements and analysis of propagation channels in high-speed railway viaducts, IEEE Transactions on Wireless Communications, 12(2), 2012, 794-805. https://doi.org/10.1109/TWC.2012.120412.120268

[7]   Kobayashi, K.; Okumura, M.: The growth of city systems with high-speed railway systems, The annals of regional science, 31(1), 1997, 39-56. https://doi.org/10.1007/s001680050038

[8]   Lei, L.; Lu, J.; Jiang, Y.; Shen, X. S.; Li, Y.; Zhong, Z.; Lin, C.: Stochastic delay analysis for train control services in next-generation high-speed railway communications system, IEEE Transactions on Intelligent Transportation Systems, 17(1), 2015, 48-64. https://doi.org/10.1109/TITS.2015.2450751

[9]   Lin, S. H.; Xu, Y.; & Wang, J. Y.: Coverage analysis and optimization for high-speed railway communication systems with narrow-strip-shaped cells, IEEE Transactions on Vehicular Technology, 69(10), 2020, 11544-11556. https://doi.org/10.1109/TVT.2020.3013341

[10] Liu, L.; Tao, C.; Qiu, J.; Chen, H.; Yu, L.; Dong, W.; Yuan, Y.: Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz, IEEE Journal on Selected Areas in Communications, 30(4), 2012, 834-845. https://doi.org/10.1109/JSAC.2012.120516

[11] Lu, J.; Xiong, K.; Chen, X.; Fan, P.: Toward traffic patterns in high-speed railway communication systems: Power allocation and access selection, IEEE Transactions on Vehicular Technology, 67(12), 2018, 12273-12287. https://doi.org/10.1109/TVT.2018.2875817

[12] Luo, W.; Fang, X.; Cheng, M.; Zhao, Y.: Efficient multiple-group multiple-antenna (MGMA) scheme for high-speed railway viaducts, IEEE Transactions on Vehicular Technology, 62(6), 2013, 2558-2569. https://doi.org/10.1109/TVT.2013.2244106

[13] Luo, W.; Zhang, R.; Fang, X.: A CoMP soft handover scheme for LTE systems in high-speed railway, EURASIP Journal on wireless Communications and Networking, 2012(1), 2012, 1-9. https://doi.org/10.1186/1687-1499-2012-196

[14] Song, H.; Fang, X.; Fang, Y.; Millimeter-wave network architectures for future high-speed railway communications: Challenges and solutions, IEEE Wireless Communications, 23(6), 2016,114-122. https://doi.org/10.1109/MWC.2016.1500255WC

[15] Song, H.; Fang, X.; Yan, L: Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway, IEEE Transactions on Vehicular Technology, 63(9), 2014, 4633-4646. https://doi.org/10.1109/TVT.2014.2315231

[16] Wu, K.; Ni, W.; Su, T.; Liu, R. P.; Guo, Y. J.: Recent breakthroughs on angle-of-arrival estimation for millimeter-wave high-speed railway communication, IEEE Communications Magazine, 57(9), 2019, 57-63. https://doi.org/10.1109/MCOM.001.1800966

[17] Yuzhe, Z.; Bo, A.: Quality of service improvement for high-speed railway communications, China Communications, 11(11), 2014, 156-167. https://doi.org/10.1109/CC.2014.7004533

[18] Zhou, T.; Tao, C.; Salous, S.; Liu, L.;Tan, Z.: Implementation of an LTE-based channel measurement method for high-speed railway scenarios, IEEE Transactions on Instrumentation and Measurement, 65(1), 2015, 25-36. https://doi.org/10.1109/TIM.2015.2477166

[19] Zhou, T.; Tao, C.; Salous, S.; Liu, L.: Measurements and analysis of angular characteristics and spatial correlation for high-speed railway channels, IEEE Transactions on Intelligent Transportation Systems, 19(2), 2017, 357-367. https://doi.org/10.1109/TITS.2017.2681112