

Journal of Electronic Research and Application

Editor-in-Chief

Joselito A. Dolot

Lyceum of the Philippines University-Batangas, Philippines

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD

(619 649 400)

Level 10

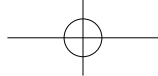
50 Clarence Street

SYDNEY NSW 2000

Copyright © 2026. Bio-Byword Scientific Publishing Pty Ltd.

Complimentary Copy





ISSN (ONLINE): 2208-3510

ISSN (PRINT): 2208-3502



Journal of Electronic Research and Application

Focus and Scope

Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication

About Publisher

Bio-Byword Scientific Publishing is a fast-growing, peer-reviewed and open access journal publisher, which is located in Sydney, Australia. As a dependable and credible corporation, it promotes and serves a broad range of subject areas for the benefit of humanity. By informing and educating a global community of scholars, practitioners, researchers and students, it endeavors to be the world's leading independent academic and professional publisher. To realize it, it keeps creative and innovative to meet the range of the authors' needs and publish the best of their work.

By cooperating with University of Sydney, University of New South Wales and other world-famous universities, Bio-Byword Scientific Publishing has established a huge publishing system based on hundreds of academic programs, and with a variety of journals in the subjects of medicine, construction, education and electronics.

Publisher Headquarter

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD

Level 10

50 Clarence Street

Sydney NSW 2000

Website: www.bbwpublisher.com

Email: info@bbwpublisher.com

Table of Contents

- 1 Design and Impact Resistance Analysis of Express Cushioning Materials Based on Biomimetic Gradient Structure**
Zixuan Xia
- 13 Research Progress on Liquid Explosive Dispersion**
Jinglong Ren, Mingkai Yue Qizhi Zhao, Yong Liu, Shuai Hou
- 20 Design of Optimized Recognition Algorithm for Degraded Facial Images in Wireless Channels**
Jiefu Yu
- 30 A Study on the Enhancement of PM_{2.5} Air Purification Capability in Cyclone Separators with Curved Structures Based on CFD Simulation**
Yitong Qiu
- 42 A Review of Reactive Power Optimization in Distribution Networks with High Penetration of Renewable Energy**
Changshuo Liu
- 51 Beyond Direct AI Assistance: The Mediating Role of AI Crafting in Converting Employee-AI Collaboration into Incremental and Radical Creativity**
Xiaolin Zhou, Dexin Zhao
- 56 Optimization of Artificial Immune Algorithms and Research on Key Issues**
Lu Hong, Su Li
- 63 Design of a High-Voltage Constant-Voltage/Constant-Current Power Supply Based on a PSFB Topology and an Automatic Frequency-Conversion Strategy**
Yuepu Sun
- 75 Research on the Construction of a Digital Credit Bank System for Vocational Education Students Based on Blockchain Technology**
Quanwei Zeng

- 81 Post-Grasp Pose Adjustment of a Tendon-Driven Underactuated Robotic Hand Using Finite-Action-Set Receding-Horizon Planning**
Lei Hu, Hongran Li
- 88 Research on High-P Precision Match Drilling Method for Reaming Mounting Holes When Replacing Sintering Machine Sprocket Tooth Plates**
Yun Ma, Zhifeng Chen, Wei Wei, Hanxiao Ma, Huan Jin
- 95 The Peak Regulation Role and Application Prospects of Pumped Storage in the Construction of a New Power System**
Boxuan Huang
- 101 Implementation and Project Management Strategy of Black Start Function in Utility-Scale Energy Storage Systems**
Ruo Jia
- 108 Polynomial Mitigation of the Fermionic Sign Problem via Topological Neural Flows and Generalized Complex Manifold Deformations**
Kaiyang Liu
- 116 Simulation Study on the Crashworthiness of a Bridge Crane Anti-Collision Mechanism**
Wenzhe Yan, Yufei Yan, Siqun Ma
- 126 A Risk-Adaptive TOPSIS Method for Electronic Decision Support in Vulnerable-Pedestrian Crossings**
Zhuorui Li
- 138 Data Privacy and Patient Consent in Smart Hospitals over the Past Decade**
Tongxing Zheng, Lin Zhang
- 146 Design and Implementation of a Dual-Layer Decoupled Gimbal Electronic Control System for Mobile Robots**
Xiaoyu Yang
- 156 Artificial Intelligence Technology and Innovation Path of Government Governance Capability**
Xiaorui Ding
- 163 The Application of an Intelligent Automatic Control System for Wastewater Treatment Containers**
Fengliang Tan, Yulin Zheng

Design and Impact Resistance Analysis of Express Cushioning Materials Based on Biomimetic Gradient Structure

Zixuan Xia*

Nanyang Model High School, 200032 Nanyang, China

*Corresponding author: Zixuan Xia, 376128053@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the rapid development of the logistics industry, the contradiction between green packaging and cargo safety has become increasingly prominent. Inspired by the lightweight, high-strength, and graded energy-absorbing microstructure of pomelo peel, this study proposes a novel biomimetic gradient porous cushioning structure. Using the ANSYS Explicit Dynamics module under a constant velocity crushing condition of 4.4 m/s, the impact resistance of three topologies, namely uniform circular holes (U-C), gradient circular holes (G-C), and gradient square holes (G-S), is compared. The simulation results indicate that, compared with the conventional uniform structure, the biomimetic gradient design successfully induces a controllable deformation mode characterized by layer-by-layer collapse, effectively reducing the initial impact peak. Among the three structures, G-C exhibits the best overall performance. Its peak contact reaction force is reduced by 16.2% compared with that of the uniform structure, and it avoids the mechanical instability due to stress concentration that occurs in the square-hole structure. Cross-checking the three simulation reports further indicates that the foam masses of the three models are all about 8.76g with only negligible differences. Therefore, the advantage of G-C should be interpreted as better overall cushioning performance under nearly equal-mass conditions, rather than as a direct proof of significantly reduced mass or the highest specific energy absorption. The study demonstrates that the biomimetic gradient circular hole structure has clear potential for impact protection, structural stability, and engineering feasibility in express packaging design.

Keywords: Biomimetic design; Gradient structure; Impact resistance; Finite element analysis; Lightweight

Online publication: Aug 11, 2026

1. Introduction

With the rapid development of China's e-commerce and express logistics, the consumption of packaging materials has become enormous. Foam plastics such as EPS and EPE, which are used for cushioning and protection, are difficult to degrade naturally after serving their short-term protective function, causing severe

white pollution and pressure on solid waste disposal. At the same time, packaging materials have faced two conflicting demands: they need to be lightweight to reduce transportation costs, yet they also require a high energy-absorption capacity to protect the contents. Traditional homogeneous foam often fails to satisfy both requirements^[1].

Nature inspires addressing this problem. The pomelo peel, with a gradient pore structure that is dense on the outside and loose on the inside, can dissipate impact energy through layer-by-layer and stable crushing deformation when the fruit falls, thereby protecting the pulp^[2]. The skeleton is a gradient structure of dense cortical bone and loose cancellous bone, which maintains sufficient support strength while minimizing weight^[3]. These biological prototypes demonstrate that gradient porous structures have significant potential in terms of lightness, high strength, and cushioning energy absorption.

In recent years, research on biological prototypes such as pomelo peel and macadamia nut shell has advanced (see **Figure 1**). Scholars have proposed a “low-high-low” porosity gradient model and, using techniques such as 3D printing, have shown that gradient designs enhance specific energy absorption, deformation stability, and impact resistance^[2,4–6]. However, the additive manufacturing processes used in most existing studies are too costly and too slow for the express packaging industry, where cost is critical, and volumes are huge^[2,4,7]. On the one hand, previous work has delved deeply into the biological porous structure. For example, the pomelo peel exhibits a typical gradient pattern along the thickness direction: low porosity in the endocarp, high porosity of the mesocarp, and low porosity of the exocarp^[2]. Meanwhile, the multilayered shell of the macadamia nut shows excellent compressive resistance under lateral impact^[6]. On the other hand, biomimetic designs have been extended to hierarchical honeycomb structures, partial stiffness enhancement, negative Poisson’s ratio structures, and TPMS structures^[8–11]. Under high strain rates, these designs exhibit stable energy absorption^[12]. The porous structures have also demonstrated their value in automotive crash safety and multi-impact protection^[13,14]. Meanwhile, research into design optimization methods for structural crashworthiness is advancing^[15]. Moreover, modern combinations of multifunctional ultralight porous materials now integrate structural support, energy absorption, and acoustic properties^[16]. In conjunction with the wide application of biomimetic honeycomb structures in various engineering fields, both achievements can provide useful information about cushioning materials^[17].

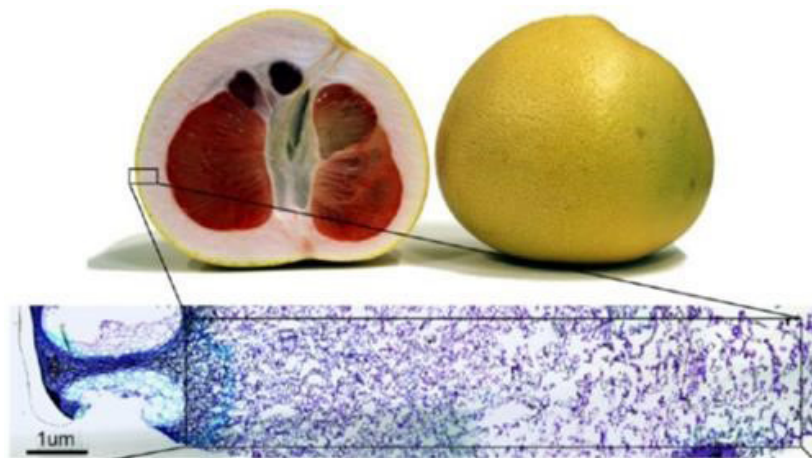


Figure 1. Cushioning energy absorption structure of pomelo peel.

Currently, several approaches are used to assess the properties of cushioning and energy-absorbing porous materials; namely, static and dynamic compression tests, drop impact hammer test, energy-absorbing efficiency and cushioning coefficient approaches, as well as numerical simulations^[1,6,18–20]. All these works prove the superiority of biomimetic gradient porous materials from a theoretical point of view. However, a lack of systematic design recommendations remains an obstacle for the development of affordable, industrially produced express packages.

Therefore, in this study, whether the performance advantages of the bioinspired gradient porous structure can be integrated into express cushioning material design without affecting their manufacturability and feasibility of mass production is explored. In this study, polyurethane foam is selected as the matrix material, and numerical comparisons among the three kinds of structures of uniform circle pores, gradient circle pores, and gradient square pores are performed. Specifically, the influence of gradient porosity and geometric shape on peak impact force, maximum compression displacement, total energy dissipation, specific energy dissipation, and deformation stability will be analyzed. Finally, through comprehensive analysis, structural guidelines for actual engineering applications will be put forward. This study intends to strike a balance among cushioning properties, lightness, stability, and manufacturability in cushioning materials, which provides the structure design basis for green logistics packaging products. On the other hand, the research results also provide theoretical references for further development work in the following aspects, such as material selection, parameter optimization, and low-cost mass production.

2. Principles of finite element simulation

The Finite Element Method (FEM) is one of the most significant numerical methods used in investigating the mechanical properties of cushioning materials. Continua with complex shapes can be broken down into finite elements, and the response of the structure will then be accomplished using global assembly and solving schemes. In designing cushioning structures, the FEM analysis can predict displacement, stress, reactive forces, and energy absorption during impact events. Thus, FEM analysis is an efficient way to bridge between biomimicry and engineering design. In experimental design, FEM analysis offers several advantages, including cost savings and design selection.

In practical research, finite element analysis proceeds through a set of sequential steps that cover discretization, element analysis, global assembly, application of load and constraint, solution, and post-processing. Explicit dynamic algorithms are especially well suited to tackling the transient impact problems explored in this work. With stable performance in handling nonlinear processes linked to high velocity, large deformation, and complex contact, these algorithms can capture the full mechanical response of polyurethane foam under compression. The material's behavioral changes across the elastic, plateau and densification stages can all be clearly reflected in the simulation outputs. Once the entire solving process finishes, researchers can extract core performance metrics from the results. These metrics include peak impact force, total energy absorption, specific energy absorption, and deformation patterns for further analysis and comparison.

Built for this research, the simulation framework integrates structural design and modeling, material model and solver selection, boundary and load condition setting, and result evaluation. In the modeling phase, aperture and gradient distributions are tuned while the overall structural dimensions remain unchanged. Key parameters are finalized before the formal solution run, including impactor velocity, contact properties, and fixed bottom constraints. Post-processing work centers on extracting total deformation and reaction force

data, which enables performance comparison across different topological configurations. The suitability of the selected material model and solver remains a core consideration throughout the entire workflow. Physical plausibility and numerical reliability are both verified through energy evolution checks and mesh sensitivity analysis.

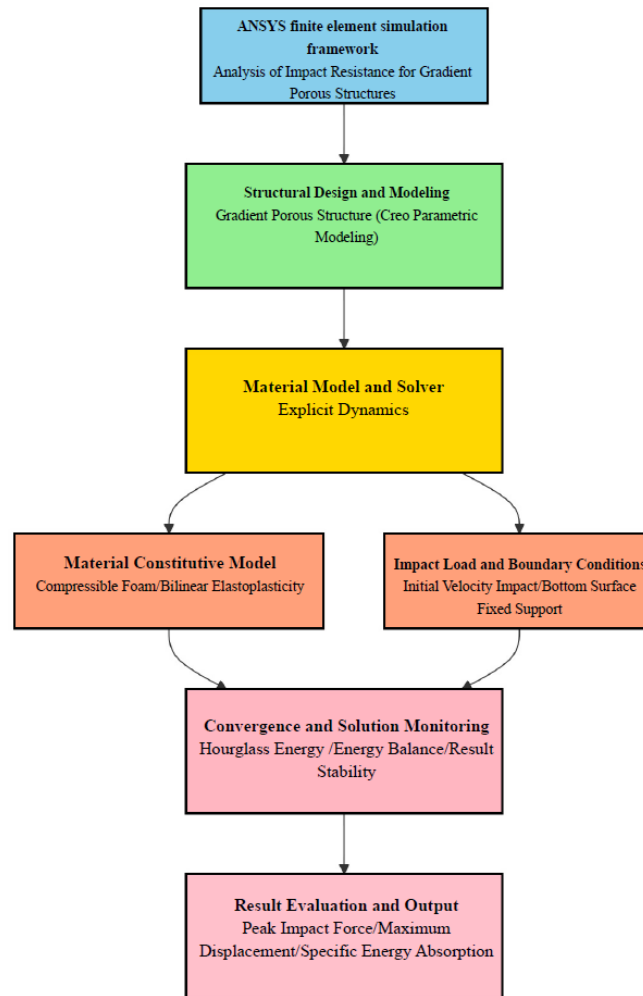


Figure 2. Finite element simulation framework.

3. Modeling and simulation schemes

Polyurethane (PU) foam is selected as the core cushioning matrix in this work. With inherent low density, high compressibility, and strong impact energy dissipation capacity, the material can be manufactured at low cost via processes including mold foaming, layer-pouring, and interlayer bonding. Such manufacturing flexibility supports its mass production and scalable adoption in express packaging. In the numerical simulation, a multilinear stress-strain relationship is applied to characterize full mechanical responses across the elastic, plateau crush, and densification stages. Considering that the emphasis of this study is on the performance comparison of various design configurations, a multilinear stress-strain curve of PU foam under compression

has been assumed. This model can later be modified using information from the material data sheet and compression tests to improve the consistency between the simulation results and the actual response.

The research object is a porous plate (see **Figure 3** to **5**). To consider the processing feasibility and modeling repeatability, the porous structure adopts a regular hole array and layers along the thickness direction to form a gradient. At the same time, a uniform structure and a gradient structure are set up to compare the two typical hole shapes of circular holes and square holes. Above the plate, a 30 mm × 30 mm × 20 mm rigid impact block is set to cover multiple hole units, reducing the randomness of the falling point and the interference of the boundary effect. All models maintain a consistent outer contour to enhance the feasibility of transforming the simulation conclusion to the actual express cushioning material design.

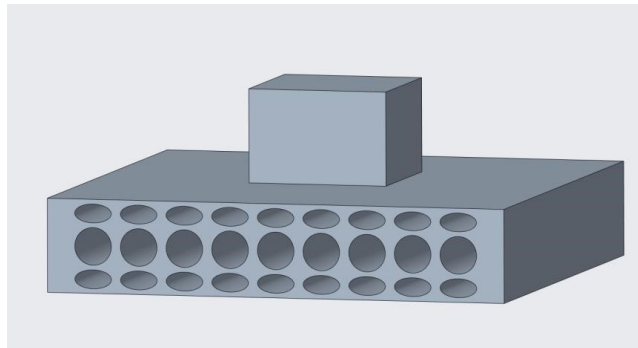


Figure 3. Gradient circular/elliptical holes.

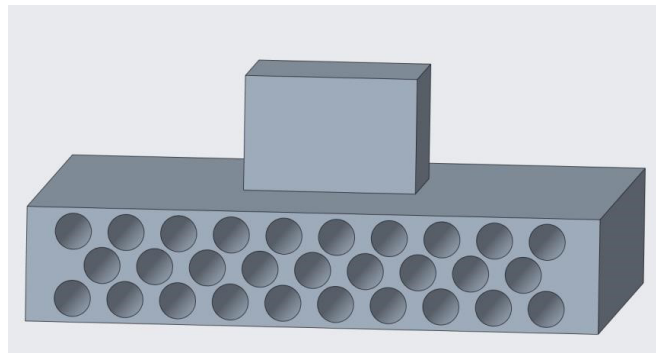


Figure 4. Uniform circular holes.

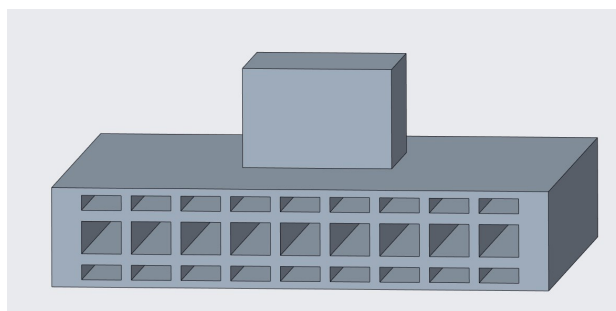


Figure 5. Gradient square holes.

To systematically examine the influence of gradient distribution on pore configuration, a controlled-vari-

able approach is adopted, with two sets of single-factor comparative experiments designed. The first set keeps hole shape, hole position, and total mass roughly consistent, and introduces structural variation solely through differences in pore distribution patterns. Comparisons are drawn between uniform circular holes (U-C) and three-layer gradient circular holes (G-C) under these constraints to identify mechanical differences between gradient and uniform structural designs. As for the second set, the three-layer gradient pattern remains unchanged, with gradient circular holes (G-C) and gradient square holes (G-S) selected for comparative analysis. This experimental setup supports exploration of how hole geometry and stress concentration affect structural stability, and effectively weakens interference from multi-factor coupling on final research conclusions.

PTC Creo Parametric is adopted for the construction of all parametric 3D geometric models, as the platform supports efficient generation of multiple model variants by defining parameters such as hole diameters in the upper, middle, and lower layers and interlayer spacing. After being imported into Ansys, the models undergo transient impact solution via the Explicit Dynamics module. The bottom surface of each porous plate is fully constrained, and an initial velocity of 4.4 m/s is applied to the impact block. The preset analysis time course covers the entire contact establishment process and the main stage of crushing energy dissipation. To guarantee comparability across structural comparisons, a unified meshing strategy is implemented for all models, with local mesh refinement deployed around hole edges and layer interfaces. Frictionless contact is uniformly specified between the impact block and the foam, and mesh sensitivity analysis on a representative model validates the rationality of these settings. During the post-processing phase, key performance indicators are extracted for further quantitative analysis, including peak impact force, maximum compression displacement, energy absorption, specific energy absorption, and crushing force efficiency. The uniformity and stability of the crushing process are then evaluated using displacement contours and deformation animations.

4. Result analysis

4.1. Simulation conditions and evaluation system

This chapter compares three models—uniform circular hole structure (U-C), gradient circular hole structure (G-C), and gradient square hole structure (G-S)—from three perspectives: microscopic deformation mode, macroscopic mechanical curves, and energy absorption efficiency.

To systematically evaluate the mechanical response characteristics of the biomimetic gradient structure under dynamic impact, this study develops a constant velocity crushing numerical model based on the ANSYS Explicit Dynamics module. The crushing rate is set at 4.4 m/s to characterize medium- to high-speed impact conditions; the compression termination time is 4.5 m/s, corresponding to a theoretical displacement of the impact block of approximately 19.8 mm, which covers the contact establishment and the main crushing energy dissipation stage. The bottom of the model is fixed to simulate ground support, while the top impact block is given a Y-direction velocity of -4.4 m/s to model the impact.

To fully characterize the structural response, monitors such as total deformation, reaction force, and energy probes are set up in the post-processing module of each model. Among them, total deformation characterizes the macroscopic crushing mode of the structure, reaction force is used to extract the contact interface load and evaluate the impact strength, and internal energy is used to quantify the energy absorption of the structure. The relevant results are statistically analyzed through the Ansys simulation report.

4.2. Structural deformation mode analysis

The structural deformation mode directly determines the stability of the energy absorption process. The three structures, as revealed by total deformation contours derived from Workbench post-processing, present distinctly different failure mechanisms. (Figure 6–9)

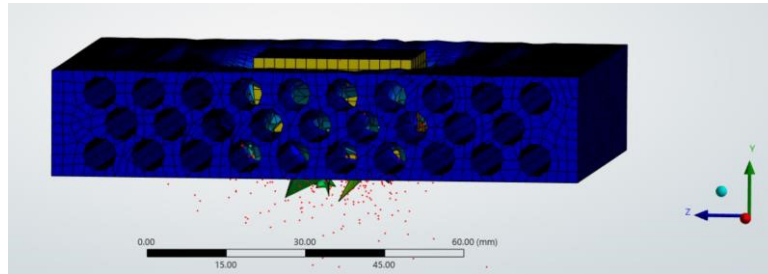


Figure 6. Deformation diagram of uniform circular hole structure (U-C).

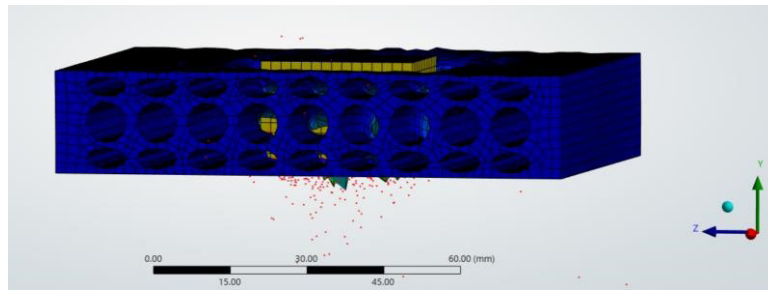


Figure 7. Gradient circular hole structure (G-C) deformation diagram.

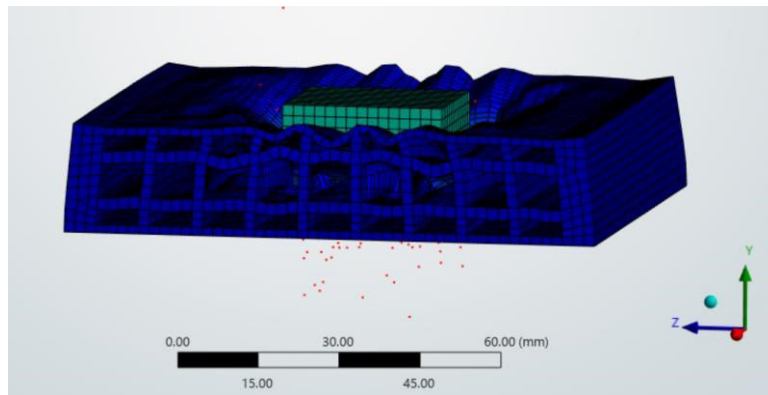


Figure 8. Gradient square hole structure (G-S) deformation.

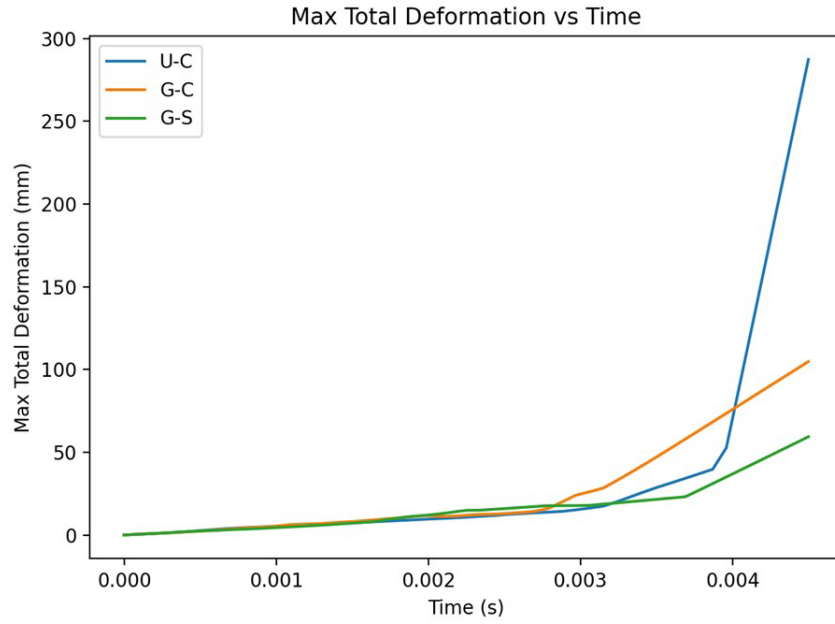


Figure 9. Time-history curves of maximum total deformation for the three structures.

In the comparison between the gradient structure and the uniform structure, the uniform circular pore structure (U-C) is more prone to random overall buckling or shear bands in weak areas at the initial stage of compression due to the uniform stiffness of the whole sample. This causes the stress to penetrate the sample and transfer to the substrate surface. Such deformation is non-hierarchical because the deformation movement of the pores can be easily exhausted, causing early local compaction. In contrast, the gradient circular pore structure (G-C) shows a mechanism of gradual collapse layer by layer. In the early stage of impact, the upper layer with high porosity first deforms and collapses; as the pressure increases, the middle and lower layers with higher density gradually bear the load and collapse in turn.

In the comparison of the hole geometric effects, although the gradient square hole structure (G-S) also adopts gradient design, there is pronounced stress concentration during its deformation process, mainly concentrated at the corners of the square holes. The vertical support ribs are also more prone to instantaneous buckling, resulting in large fluctuations in load-bearing capacity and indicating instability. In contrast, the gradient circular-hole structure (G-C) gradually evolves into an ellipse during compression, and the stress distribution along the hole wall is relatively uniform without significant concentration areas. Therefore, it can maintain good integrity under large deformation conditions, and the energy absorption process is also more stable.

4.3. Contact reaction force and impact protection performance

In express delivery logistics, the primary function of cushioning materials is to reduce the maximum impact force on the goods. This study extracts the Reaction Force at the contact interface between the impact block and the foam and uses it as a characterization parameter of the equivalent impact load on the goods during a drop.

According to the simulation results (**Table 1**), the peak reaction force (F_{max}) varies significantly among the three structures: 1487.30 N for U-C and 1246.50 N for G-C—a reduction of 16.2%. For G-S, the value drops further to 980.71 N, a reduction of 34.1%. A higher peak reaction force indicates that the U-C structure

generates a stiffer reaction at the moment of impact, which is detrimental to the protection of fragile goods. In contrast, G-C and G-S form a low reaction buffer through a low stiffness design at the top, effectively weakening the initial impact. The G-S has the lowest peak, mainly because the square hole structure is more prone to local instability and its overall stiffness is relatively lower. See **Figure 10**.

Table 1. Comparison of key impact resistance indicators for three structures

Model	Fmax (N)	Fmean (N)	CFE (%)	Dmax (mm)	EA(J)
U-C (uniform circular hole)	1487.30	356.32	23.96	287.260	1.4577
G-C (Gradient circular hole)	1246.50	285.43	22.90	104.750	1.2322
G-S (Gradient Square Hole)	980.71	314.56	32.07	59.343	1.0946

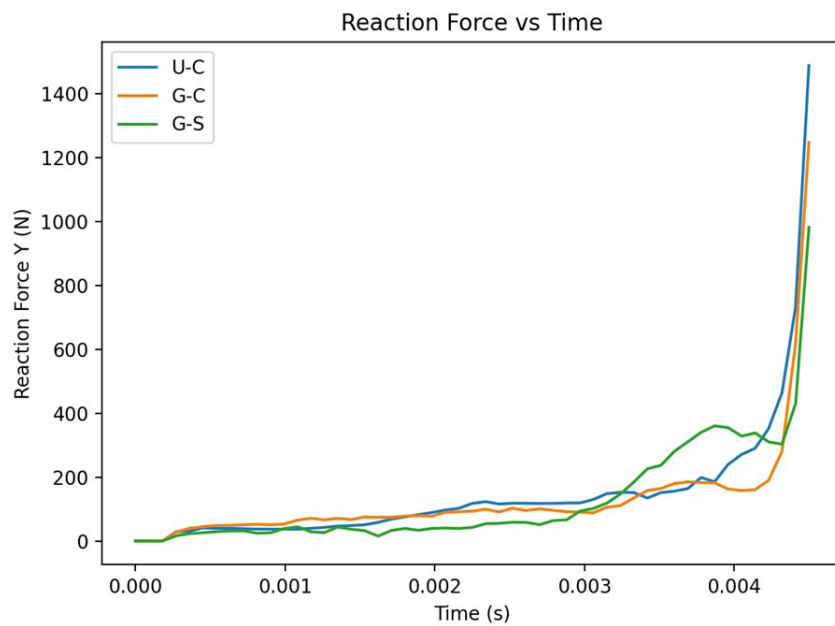


Figure 10. Time-history curves of reaction force for the three structures.

To further evaluate the stability of the cushioning process, this study introduces the crushing force efficiency (CFE, average force/peak force) indicator. The CFE of G-S is 32.07%, which appears to be the best. However, as shown in **Figure 4**, its force–displacement curve exhibits pronounced jagged fluctuations, indicating that the square-hole supporting units are prone to sudden instability during compression, which may in turn cause secondary vibration. The CFE of the G-C is 22.90%, slightly lower than that of the G-S, but its mechanical response curve is smoother with no abrupt changes, demonstrating better cushioning stability. The CFE of the U-C is about 23.96%, but its overall protection performance remains weak due to the high peak reaction force.

4.4. Energy absorption and lightweighting analysis

Another key performance indicator of the cushioning material is energy absorption capacity. In this study, the internal energy is extracted using an energy probe, as shown in **Table 1**.

The simulation results show that within the same compression duration (4.5 m/s), the U-C structure ab-

sorbs the highest internal energy 1.4577 J, but its maximum total deformation reaches 287.26 mm. It should be noted that this value is the result of cumulative mesh distortion, mainly reflecting the degree of local element failure rather than the actual physical deformation. In contrast, the absorbed internal energy of the G-C structure is 1.2322 J, and the maximum deformation is 104.75 mm; the G-S structure is 1.0946 J and 59.34 mm, respectively. Therefore, although U-C achieves relatively high total energy absorption, it does so at the expense of a higher peak force and more severe local damage. This is the classic high-stiffness behavior in impact energy absorption. In contrast, G-C and G-S engage in the loading through a gradient hierarchy, achieving energy dissipation while reducing the initial impact peak, which is more in line with the design requirements of express packaging that prioritize peak reduction while taking energy absorption into account.

The review of mass and specific energy absorption (SEA) indicates that the foam masses of the three models are 8.7618 g, 8.7571 g, and 8.7626 g, respectively, with a difference of less than 0.1%. Therefore, this group of results is more suitable for configuration comparison under nearly equal mass conditions and should not be interpreted as performance comparison under significant weight reduction conditions. In terms of EA/m, the SEA values for U-C, G-C and G-S are 166.37, 140.71 and 124.92 J/kg, respectively. Therefore, the advantage of G-C is that, under nearly equal mass conditions, the gradient circular hole design can effectively reduce the peak reaction force and improve the stability of the high-pressure breakdown process, thereby achieving better overall cushioning performance; the potential for lightweighting remains to be verified by adjusting porosity, introducing physical controls, or further parametric optimization.

4.5. Comprehensive evaluation and selection recommendations

By evaluating the three structures in terms of safety, stability, and manufacturability, the following conclusions can be drawn. The U-C hole structure is not the ideal choice due to several drawbacks. It generates the largest maximum reaction force at 1487 N, which will easily transmit a considerable load onto the protected goods in case of collision, thus increasing the likelihood of cargo damage. The deformation capacity of the U-C is relatively low, preventing full use of the shock-absorbing properties of the material. Despite generating the smallest maximum reaction force at 980 N among the considered types of holes, the G-S exhibits high stress concentrations and instability during crushing, resulting in a significant degree of irregularity and uncontrollability. In addition, the corners of the squares tend to tear off easily. This can be detrimental to the structural resistance. G-C, however, proves to be a good compromise between control over peak reaction and mechanical stability. Its total energy absorbed may be inferior to that of the U-C for this type of approximately equal mass models; nonetheless, its peak reaction is 16.2% lower than that of the U-C.

To conclude, in almost identical mass cases, the biomimetically designed gradient hole circular (G-C) structure provides a proper balance of low peak impact forces, consistent deformation behavior, and ease of manufacturing. Thus, it is considered better compared to others for express cushioning purposes.

5. Conclusions

In this paper, numerical simulations have been performed to analyze the effect of pore distribution and geometry on the mechanical behavior of cushioning materials. The key outcomes have been outlined hereafter.

- (1) The results from the numerical simulation demonstrate that structures characterized by homogeneous distribution of pores tend to undergo global buckling, accompanied by fast stress transmission. The

bionic gradient structure realizes the layer-by-layer collapse mode of the soft layer yielding first and the hard layer yielding later through the spatial gradient of stiffness, thus prolonging the energy absorption process and improving the impact release effect.

- (2) With regard to the reaction force, its magnitude can be decreased thanks to the introduction of gradient distribution. The G-C structure is characterized by a peak reaction force value of 1246.50 N; thus, the reduction in comparison with the U-C structure equals 16.2%. Therefore, the probability of damaging delicate cargo during transportation is greatly reduced.
- (3) The hole shape is one of the factors that greatly affects the cushioning response of the material. Although the gradient square hole structure ensures the minimum peak reaction force (980.71 N), it causes high stress concentration at corners and leads to the instability of supporting layers and obvious fluctuation of mechanical response. In contrast, the stress distribution of the gradient circular hole structure is more uniform, the crushing process is more stable, and the engineering applicability is stronger.
- (4) From the review of the mass and energy data, we can see that the mass of the foam for all three models is about 8.76 g, with slight differences between them. Thus, conclusions that much less mass of G-C or maximum specific energy absorption cannot be drawn. A more reasonable expression is that under the condition of approximately equal mass, G-C has both lower peak reaction force, a more stable crushing response, and better energy absorption performance, and the comprehensive cushioning performance is the best.

In summary, the biomimetic gradient circular hole structure can better balance safety, stability, and engineering manufacturability under nearly equal-mass conditions, demonstrating its potential as a cushioning packaging structure for express delivery.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhang W, 2007, Review of Evaluation Methods for Cushioning Performance of Energy Absorbing Materials. *Packaging Engineering*, 28(5): 1–5.
- [2] Yang B, 2022, Construction, Performance and Applications of Pomelo Peel-Inspired Gradient Porous Materials, thesis, Wuhan University.
- [3] Gibson L, Ashby M, 1997, *Cellular Solids: Structure and Properties* (2nd ed.). Cambridge: Cambridge University Press.
- [4] Wang M, 2023, Study on Impact Behaviors of Hierarchical Graded Cellular Materials and Sandwich Structures, thesis, Southeast University.
- [5] Lin X, 2023, Shock Resistance and Energy Absorption Performance of Tube Reinforced Hyperelastic Gradient Porous Structure, thesis, Harbin Engineering University.
- [6] Zhou J, et al., 2025, Study on Impact Resistance of Biomimetic Structures with Different Pore Shapes and Densities Under Lateral Impact. *Mechanical and Electrical Engineering*, 42(5): 945–954.
- [7] Li J, 2021, Research on Impact Resistance and Interface Performance of Composite Structure Based on 3D Printing Method, thesis, Dalian University of Technology.

- [8] Zhang S, 2022, Design of Novel Cellular Structures with Impact Resistance Properties, thesis, Harbin Engineering University.
- [9] Yan J, Li X, Song X, et al., 2025, Deformation Modes and Energy-Absorbing Properties of Porous Structures of Local Stiffness-Reinforced Composites. *Journal of Composite Materials*: 1–16.
- [10] Lian J, 2022, Research on Impact Resistance Behavior of Honeycomb Porous Structures, thesis, Harbin Engineering University.
- [11] Li X, Song W, Chen J, 2019, Numerical Simulation of Mechanical Properties of 3D Printing TPMS Cellular Materials. *Journal of Taiyuan University of Technology*, 50(3): 386–393.
- [12] Zhang W, 2023, Research Progress on Energy Absorption Properties and Vibration of Periodic Lightweight Porous Structures. *Journal of Vibration and Shock*, 42(4): 1–19.
- [13] Peng Y, Xin Y, 2009, The Application of Cellular Materials in Automobile Crash-Safety. *Journal of Mechanical Strength*, 31(2): 325–329.
- [14] Zhang J, et al., 2005, Preparation of Elastic Porous Composite Material and Its Energy-Absorption Characteristics. *Packaging Engineering*, 26(5): 13–15.
- [15] Baroutaji A, Sankar A, Gilani N, et al., 2017, On Design Optimization for Structural Crashworthiness and Its State of the Art. *International Journal of Crashworthiness*, 22(5): 455–477.
- [16] Lu T, 2009, Multifunctional Ultralight Porous Materials with Simultaneous Load-Bearing, Energy Absorbing and Sound Absorbing Capabilities. *Science China: Technical Sciences*, 39(9): 1477–1491.
- [17] Zhang Q, Yang X, Li P, et al., 2015, Bioinspired Engineering of Honeycomb Structure—Using Nature to Inspire Human Innovation. *Progress in Materials Science*, 74: 332–400.
- [18] Liu Q, 2016, Numerical Study on Dynamic Compression of Porous Materials, thesis, Beijing Institute of Technology.
- [19] Wang D, 2006, Characterization Method of Buffering Energy Absorption Characteristics of Porous Materials. *Packaging Engineering*, 27(1): 37–39.
- [20] Zhang H, He B, 2011, Finite Element Analysis. Beijing: Machinery Industry Press.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research Progress on Liquid Explosive Dispersion

Jinglong Ren¹, Mingkai Yue^{1*}, Qizhi Zhao², Yong Liu², Shuai Hou²

¹School of Equipment Engineering of Shenyang Ligong University, Shenyang 110159, Liaoning, China

²Unit 61699, Yichang 443200, Hubei, China

**Corresponding author:* Mingkai Yue, 13032486996@163.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Liquid explosive dispersal refers to the process in which the instantaneous high pressure generated by explosive detonation drives the liquid within the device. This paper reviews domestic and international research on liquid explosive dispersal, categorizing it into core directions including dispersal mechanism, influencing factors, and numerical simulation, and systematically summarizes the research progress of key parameters during the liquid explosive dispersal process.

Keywords: Simulant; Explosive dispersion; Numerical simulation

Online publication: Aug 11, 2026

1. Introduction

When an explosive detonates, it creates an instant high pressure that drives the liquid inside a device, this is the core of liquid explosive dispersion. Depending on the liquid's properties, it can serve specific functions; it's widely used in engineering and agricultural fields for tasks like fire suppression, explosion inhibition, pest elimination, and artificial rainfall ^[1-3]. Once the central explosive charge initiates, detonation products drive and spread the surrounding medium, kickstarting the dispersion process. This process is super-fast, dynamic, and involves messy, hard-to-track interactions between liquid, solid, and gas phases. It covers physicochemical phenomena such as shock wave and reflected wave interactions, causing structural failure, liquid interface disturbance, fragmentation, and phase change. A whole range of factors shape this process; areas based in explosion mechanics, solid mechanics, fluid dynamics, thermodynamics, chemical kinetics, and dynamic testing techniques seem to contribute to how it unfolds ^[4]. For this paper, I'm going to break down liquid explosive dispersion from four distinct perspectives.

Back in 1990, Gardner and Glass from Sandia National Laboratories split the liquid explosive dispersion process into 'near-field' and 'far-field' stages, sorting them based on which forces have the biggest effect

on the liquid ^[5,6]. Detonation forces dominate the ‘near-field’ stage; aerodynamic drag forces take over in the ‘far-field’ stage. They used hydrodynamic methods to run numerical simulations and analyzed key characteristics of both stages; this was the first full look at the entire liquid explosive dispersion process. This relatively simple classification seems to have been widely adopted in later research. Research in the area dates back to the 1960s in the USA, though. Poppoff et al. divided the explosive dispersion process into four stages, basing their work on phenomena they actually observed: detonation, shock, separation, and jetting ^[7]. Samirant et al. used flash X-ray photography to capture images of the liquid explosive dispersion process, and these shots revealed a liquid ring formed when expanding detonation product gases pushed the liquid fill outward ^[8]. Also, Zabelka et al. took photos that showed another dispersion pattern, jets spraying radially outward from the explosion center ^[9].

2. Method and design

2.1. Research on liquid explosive dispersion mechanisms

2.1.1. Primary fragmentation

Primary fragmentation is the first stage when dispersed liquid starts breaking apart; it marks the beginning of the whole atomization process. It decides the motion range of the liquid and gives secondary fragmentation its initial conditions. When we want to study this initial stage of fragmentation, Linear instability theory seems to be the primary theoretical approach we rely on.

When the interface between liquid and gas gets unstable and keeps getting worse, it will eventually make the liquid break apart, this is the initial stage of liquid explosive dispersion that Yue Zhongwen calls primary fragmentation. This is how the explosive starts to spread; it may happen quicker than we expect ^[10]. Then, common instabilities include ^[11].

Let me kick off with Rayleigh-Taylor (R-T) instability; it arises from interface instability between two fluids of different densities when gravity and inertial forces act on them.

Kelvin-Helmholtz (K-H) instability is the second phenomenon I talked about. It comes up when two fluids have a tangential velocity difference; interface perturbation starts to develop as a direct outcome of that side-to-side speed difference between them.

Richtmyer-Meshkov (R-M) instability is the phenomenon we’re talking about. When a shock wave transmits through a fluid interface, it imparts a finite acceleration to the interface; this then sets off perturbation growth.

Yang Lei et al. did experimental research on axisymmetric dispersion of glycerol; they used a modified vertical shock tube to carry out their tests ^[12]. When they went through all the data collected from these experiments, they found some key results that may help us understand how glycerol disperses in that axisymmetric setup.

We ran experiments with transparent glycerol; we could actually watch how R-T and R-M instabilities develop step by step throughout the entire process of axisymmetric dispersion. We captured clear images of bubble and spike growth. By going through these images, we can tell that this growth in the mixing region may lead to primary fragmentation.

So when we ramp up driving force and shock Mach number, it seems to affect the unstable interface in two ways. The initial perturbation wavelength gets shorter; the wavenumber goes up.

We cranked up the driving force, and spike growth picked up really fast. Location of primary fragmentation didn't shift at all. K-H instability and surface tension seemed to have a stronger influence on the spikes; this made spike fragmentation far more intense than before.

2.1.2. Secondary fragmentation

Secondary fragmentation of droplets is a crucial stage in liquid dispersion. In this stage, droplets formed during primary fragmentation undergo further breakup under aerodynamic forces. The main research methods include phenomenological analysis and linear instability theory.

Hinze's research showed that the mode of secondary fragmentation is determined by two dimensionless numbers composed of three forces: the Weber number (We) and the Ohnesorge number (Oh) ^[13].

The Weber number (We) is the ratio of inertial force (drag) to surface tension force:

$$We = \frac{\rho_e d_l v^2}{\sigma_l} \quad (1)$$

Where ρ_e is the ambient gas density, d_l is the droplet diameter, v is the droplet velocity, and σ_l is the surface tension coefficient.

The Ohnesorge number (Oh) is the ratio of viscous force to surface tension force:

$$Oh = \frac{\mu_l}{(\rho_l d_l \sigma_l)^{\frac{1}{2}}} \quad (2)$$

Where μ_l is the dynamic viscosity coefficient of the liquid.

Yang Wei et al. used the CLSVOF multiphase flow model coupled with adaptive mesh refinement to study the transition mechanism between different droplet breakup modes ^[14]. Their research indicated that R-T instability plays a significant role in the transition of droplet breakup modes. The ratio of the droplet's maximum diameter to the maximum R-T unstable wavelength directly influences the droplet breakup mode.

Wang Chao et al. conducted experimental and numerical simulation studies on the breakup process of droplets under incident shock waves ^[15]. They analyzed droplet deformation and breakup using We and Oh numbers. Analysis of experimental and simulation results showed that increasing We promotes droplet breakup. Beyond a certain We value, droplet breakup transitions from shear breakup to catastrophic breakup.

2.2. Numerical simulation research on liquid explosive dispersion

Shi Yina et al. used a dimensionally split Eulerian program coupled with the Youngs' interface treatment method to simulate the liquid stratification phenomenon in the flow field of glycerol and water media driven by a central explosive charge ^[16]. Combined with experimental results, they proposed three coexisting mechanisms during droplet formation: fragmentation of the outer jet, R-T instability in the inner layer, and "cavitation" fragmentation in the intermediate liquid layer.

Measuring the concentration of explosive materials post-dispersion is challenging. Numerical simulation offers a solution; however, existing models for calculating particle concentration often neglect measuring initial particle conditions, leading to inaccuracies in describing concentration distributions. Chen et al. proposed a model to predict the concentration distribution of dispersed liquid and particulate materials after an explosion, which was validated and showed good agreement with experimental data ^[17].

2.3. Research on factors influencing liquid explosive dispersion

2.3.1. Influence of explosive energy

(1) Effect of charge mass

Wu Deyi et al. conducted dispersion experiments with water and alcohol under different specific charge masses (charge mass per unit liquid mass) under identical conditions ^[18]. Results showed that increasing the charge mass significantly increases the initial dispersion velocity of the liquid. Simultaneously, the velocity attenuation coefficient increases, the dispersion duration decreases, and the dispersion radius increases. Yan Shilong et al. concluded from water explosion experiments with varying specific charge masses that larger specific charge masses result in more uniform water fragmentation, better continuity of the water mist, and a relatively uniform overall mist concentration ^[19]. Tang Hui, based on energy conservation and detonation principles and introducing a correction coefficient 'k' and 'ω', proposed a method for calculating dispersion parameters:

initial liquid velocity:

$$v_1 = \sqrt{\frac{4m_e Q_v \left[1 - \left(\frac{r_0}{r_1} \right)^4 \right]}{k^2 m_e + 2m_1 + 2k^2 m_s}} \quad (3)$$

Where: r_0 is the initial radius of the shell; r_1 is the rupture radius of the shell; m_1 is the mass of the liquid filling; m_s is the mass of the warhead shell; m_e is the mass of the central explosive tube; Q_v is the explosive heat of the explosive; k is the correction coefficient (taken as 0.3 according to the results of previous tests).

Throwing radius:

$$R = r_0 + \frac{v_1}{\omega} \quad (4)$$

Where: ω is a correction coefficient (fitted from static explosion results, taken as 32)

Comparison of calculated and measured values showed maximum relative errors of 18.8% for initial velocity and 17.6% for dispersion radius, both meeting error control requirements. Experiments also showed that both the initial velocity and dispersion radius of liquid fire suppressant are proportional to the central explosive charge mass.

(2) Effect of energy distribution

Experiments showed significant differences in fuel dispersion velocity at different positions for Structure A, with velocity at the small end notably higher than at the large end. Structure B formed an approximately cylindrical cloud initially, with no significant difference in fuel diameter between large and small ends. Dispersion velocity differences across positions in Structure B were also smaller, with the maximum velocity at the small end slightly higher than at the large end. It can be concluded that the fuel dispersion velocity at different positions of the truncated cone device is related to the charge mass per cross-section, with initial velocity increasing with charge mass. Based on extensive experiments and analysis with different liquids, scholars concluded that for liquid explosive dispersion, the initial cloud velocity and dispersion radius are linearly related to the charge mass, but this relationship changes beyond a certain point.

2.3.2. Influence of device structure

Lu Xiaoxia et al. conducted liquid explosive dispersion experiments with different casing materials (PMMA, PVC, Steel, Film - **Figure 1**)^[20]. They found that under the action of reflected rarefaction waves, liquid cavitation occurs. Higher casing strength delays cavitation onset and shifts its location closer to the explosion center. Furthermore, mixing between detonation product gases and the liquid occurs earlier near their interface with stronger casings.

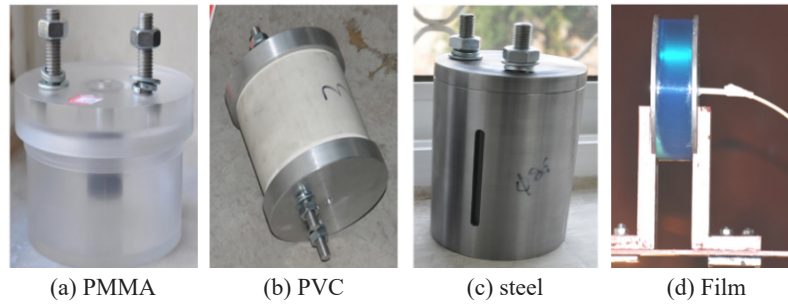


Figure 1. Explosive dispersion devices with different casing materials.

2.4. Droplet size measurement

The Sauter Mean Diameter (SMD or D32) is a widely used parameter in multiphase systems to describe the average particle size of droplets or particulate groups. Cai Qingjun et al. utilized an optical method based on Dobbins' laser scattering principle to measure the spatial and temporal distribution of the SMD of droplets formed during the secondary fragmentation of axisymmetric liquid dispersion^[21]. The experimental setup is shown in **Figure 2**.

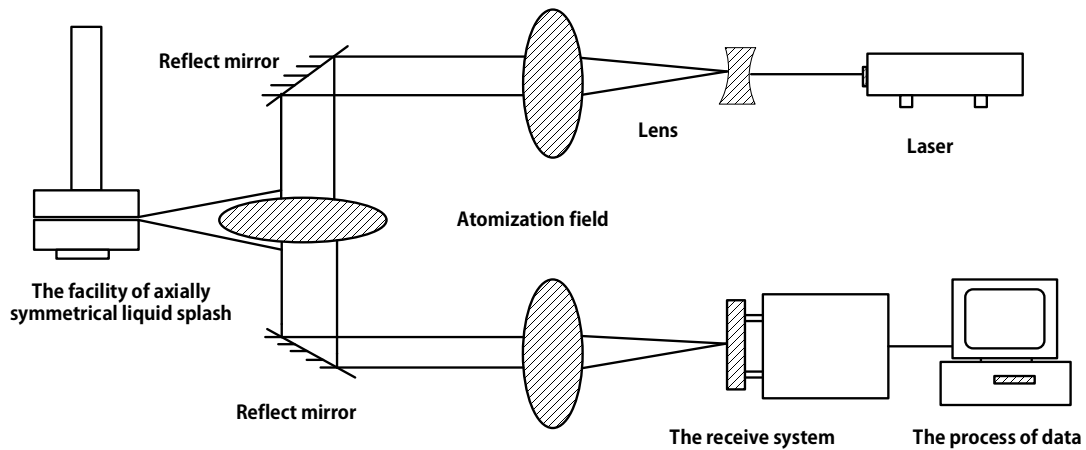


Figure 2. Optical path for measuring particle size from secondary fragmentation in axisymmetric horizontal liquid dispersion.

As shown in **Figure 2**, the setup uses a U-shaped container and a diaphragmless shock tube to achieve axisymmetric liquid ring dispersion and atomization. A laser beam is generated, reflected twice, and focused onto a receiver. A computer processes the acquired data to obtain the SMD of secondary fragmented droplets. System calibration confirmed the reliability and accuracy of this optical measurement system for determining the size of aerosol clouds generated by axisymmetric liquid dispersion.

3. Results and discussion

This paper discusses the dispersion of liquid explosions from three perspectives: the dispersion mechanism of liquid explosions, the factors influencing the dispersion of liquid explosions, and numerical simulation of the dispersion of liquid explosions.

Research on the dispersion and fragmentation mechanism of liquid explosions primarily employs linear instability theory as the main research method, dividing the entire process of liquid explosion dispersion into initial fragmentation and secondary fragmentation. Therefore, in subsequent research, the classification of these two dispersion stages can be further refined or alternative classification methods can be explored.

Regarding the numerical simulation of explosion dispersion, scholars both domestically and internationally typically employ a staged approach, simulating the near-field and far-field stages using different models. To more accurately simulate the various stages of liquid explosion dispersion, future simulations could explore alternative models and compare them with existing models to identify the optimal model under different conditions.

Research on the influencing factors of explosion dispersion is conducted by varying the explosion energy and dispersion devices. Explosion energy includes the effects of charge quantity and energy distribution on dispersion efficiency. Scholars have conducted experiments with different explosion energies and dispersion devices, concluding that under appropriate conditions, a larger charge quantity leads to better dispersion efficiency; different dispersion devices yield different dispersion efficiencies, so appropriate devices should be selected based on specific scenarios in practical applications.

4. Conclusions

Right now, when we focus on multiphase coupling models and high-precision measurements, current research still has some noticeable shortcomings; it may not cover all the little details we need to get accurate results.

Let me start with model limitations: current staged models can't describe the multiphase coupling effects of turbulence-fracturing-phase transition. They may only handle each process on its own. To some extent, they focus on single-phase behaviors, completely missing how turbulence, fracturing, and phase transition interact in a multiphase system.

We face a technical bottleneck right now. In situ measurements of nanoscale droplets don't have enough precision; we need to work on improving that to get more accurate, usable data for our thesis.

Next, for future developments, we can focus on intelligent optimization algorithms based on machine learning; we'll pair these with microfluidic experiments to reveal fragmentation mechanisms at submillimeter scale, and this may drive dispersion devices toward better controllability and higher efficiency.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Ding J, Liu J, Peng J, 1999, One-dimensional Numerical Simulation of the Initial Stage of Liquid Fuel Explosive Dispersion. *Journal of Ballistics*, 1999(03): 29–33.

- [2] Zuo X, 2020, Experimental Study on Combustion, Explosion and Atomization Characteristics of Liquid Fuel, thesis, Nanjing University of Science and Technology.
- [3] Hou S, Zhao Q, Bai H, et al., 2023, Numerical Simulation and Experimental Study on Typical Liquid Explosive Dispersion. *Chemical Defense Research*, 2(1): 56–64.
- [4] Xue T, Xu G, Huang Q, et al., 2015, Research Progress on the Explosive Dispersion Process. *Science Technology and Engineering*, 15(21): 60–67.
- [5] Gardner D, 1990, Near-field Dispersal Modeling for Liquid Fuel-Air Explosives. In: *Proceedings of the 13th International Symposium on Military Applications of Blast Simulation*.
- [6] Glass M, 1990, Far-field Dispersal Modeling for Fuel-Air-Explosive Devices. Sandia National Laboratories Report SAND90-8239.
- [7] Poppoff I, Thuman W, 1967, Research Studies on the Dissemination of Solid and Liquid Agents. Final Report.
- [8] Samirant M, Smeets G, Baras C, et al., 1989, Dynamic Measurements in Combustible and Detonable Aerosols. *Propellants, Explosives, Pyrotechnics*, 14: 47–56.
- [9] Zabelka R, Smith L, 1969, Explosively Dispersed Liquids. Part 1. Dispersion Model. Report.
- [10] Yue Z, 2006, Study on Water Atomization by Explosive Dispersion, thesis, Anhui University of Science and Technology.
- [11] Ding J, 2001, Theoretical Model and Full Process Numerical Simulation of Liquid Explosive Dispersion, thesis, Dalian University of Technology.
- [12] Yang L, Yang X, Huang Z, 2016, Experimental Study on Shock Wave Driven Axisymmetric Dispersion of Liquid. *Journal of Experiments in Fluid Mechanics*, 30(6): 32–36.
- [13] Hinze J, 1955, Fundamentals of the Hydrodynamic Mechanism of Splitting in Dispersion Processes. *AIChE Journal*, 1(3): 289–295.
- [14] Yang W, Jia M, Sun K, et al., 2017, Study on the Transition of Droplet Deformation-Bag-Multimode Breakup. *Journal of Engineering Thermophysics*, 38(2): 416–420.
- [15] Wang C, Wu Y, Shi H, et al., 2016, Breakup Process of Droplet Under Shock Wave Impact. *Explosion and Shock Waves*, 36(1): 129–134.
- [16] Shi Y, Hong T, Yao W, et al., 2014, Simulation of Explosion-Driven Liquid Stratification and Droplet Size. *Chinese Journal of Computational Mechanics*, 31(6): 781–786.
- [17] Chen X, Wang Z, Yang E, et al., 2021, Predictive Model for Concentration Distribution of Explosive Dispersal. *ACS Omega*, 6(3): 2085–2099.
- [18] Wu D, Yang X, 2003, Experimental Study on Liquid Dispersion Under Strong Shock Wave. *Explosion and Shock Waves*, 2003(1): 91–95.
- [19] Yan S, Liu F, Yue Z, et al., 2007, Effect of Specific Charge on the Motion Characteristics of Water Explosive Dispersion and Atomization. *Journal of Harbin Institute of Technology*, 2007(11): 1825–1828.
- [20] Lu X, Li L, Zhao S, et al., 2016, Effect of Casing Constraint on Flow Field Characteristics of Liquid Explosive Dispersion. *Explosion and Shock Waves*, 36(6): 803–810.
- [21] Cai Q, Han Z, Wan Q, et al., 1999, Measurement of Cloud Particle Size Formed by Secondary Fragmentation of Liquid Ring and Calibration of Testing System. *Explosion and Shock Waves*, 1999(2): 56–62.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Design of Optimized Recognition Algorithm for Degraded Facial Images in Wireless Channels

Jiefu Yu

Yunnan Open University (Yunnan Technical College of Industry), Kunming 650500, Yunnan, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The Design of the wireless channel-degraded face image optimization recognition algorithm aims to solve the problem of reduced image quality due to wireless transmission and thus reduce the drop in recognition accuracy. A new optimization recognition algorithm has been put forward, and at the same time, a way to preprocess the degraded image, enhance its features, and improve the clarity of the image has also been introduced. Many attributes and a combination of local and global features will be employed to increase the accuracy and robustness of the recognition. Based on the experiment, it can be seen that this algorithm has improved the recognition rate of all kinds of degraded facial image and has good application prospects; thus, it offers a new solution for face image recognition in wireless communication.

Keywords: Wireless channel; Degraded facial image; Optimized recognition algorithm

Online publication: Aug 13, 2026

1. Introduction

With the development of wireless communication technology, facial image can now be transmitted wirelessly. Due to the instability of the wireless channel, there may be blurring and noise in the face image; thus, it will affect the accuracy of face image recognition. There are still some deficiencies in the current face recognition algorithms for degraded images, and they do not meet the requirements of actual applications. Therefore, we need to develop an algorithm that can improve the recognition accuracy of degraded facial image in wireless channels. The main goal of this paper is to introduce a new algorithm that can improve the recognition performance of degraded facial image and promote the development of face image recognition technology in wireless communication environments.

2. Algorithm design foundation

2.1. Analysis of wireless channel characteristics

Wireless channels are open electromagnetic wave transmission media and are not as stable as the environment for wired channels. The two main characteristics are large-scale transmission loss and small-scale channel

fading. The two main reasons are path loss and shadow fading. During the transmission of facial image, there are many obstacles such as walls and vegetation that will cause electromagnetic wave energy to be continuously absorbed; as a result, the overall brightness and effective pixel amplitude of the received image will decrease. The features of the small scale are several path effects, Doppler frequency shift, and random electromagnetic interference in the channel. Multihop is a type of effect that changes the phase of pixels in the face image and causes blurring of the image edges. Most of the Doppler frequency shifts occur in the case of mobile acquisition terminals, and there will be spectral distortion. The Wireless Channel is time-varying and random. The Signal-to-Noise Ratio and transmission bandwidth of the channel will vary according to the spatial environment. Instability is the root cause of the decline in face image transmission, and it is also the primary constraint condition for the design of the following algorithm adaptation. Ground radio frequency communication and dedicated wireless frequency bands for the Internet of Things may also have interference between frequency bands. The main visual transmission frequency bands of 2.4 GHz and 5.8 GHz are easily interfered with by surrounding civilian radio frequency equipment, and as a result, there are fluctuations in the time-domain and frequency-domain of the channel.

2.2. Research on degradation factors of facial image

Based on the wireless channel transmission link, the three main reasons for the degradation of facial image can be divided into transmission layer degradation, inherent degradation at the acquisition end, and link coupling degradation. Degradation in the transmission layer is caused by channel noise, bit errors during transmission, and loss of data packets; it will appear as salt-and-pepper noise, color block mosaic, and absence of local feature pixels in the face image. Degradation at the acquisition end is due to optical distortion and exposure deviation of the camera equipment, and it will be further amplified during wireless transmission. Link coupling degradation is caused by the change in the wireless channel over time and image encoding compression. In a low Signal-to-Noise Ratio (SNR) channel, the high-frequency part of the image features will be filtered out, and fine differentiating features, such as the area around the eyes and nasolabial folds, will be completely lost. The different causes are related. A single image defect is the result of multiple channels and imaging parameters combined, and it is more difficult to restore face features or recognize them. Most of the face image transmissions are based on the H.264 and H.265 video encoding standards. In a low-bandwidth channel, the encoder will compress the high-frequency texture information of the face to a certain extent, and it will be lost permanently.

2.3. Advantages and disadvantages of existing recognition algorithms

Most general face recognition algorithms are divided into two types. The first is the traditional machine learning recognition algorithm, and the second is the end-to-end deep learning recognition algorithm. Traditionally, the algorithms that have been employed are LBP, HOG artificial feature matching, and SVM classification. The structure of the model is simple, the volume of inference calculation is small, and it can meet the requirements of low-bandwidth wireless transmission terminals. However, they are not robust to noise and geometric deformation. If the channel is poor, there will be a reduction in the success rate of feature matching. CNN and Transformer backbone networks are used in the recognition algorithm of deep learning, and they have good feature extraction abilities. In the case of normal image quality, the recognition accuracy is very high; however, the model has a large number of parameters and requires high integrity of the input image pixels. It cannot change with changes in the adverse environment of the wireless channel,

has too many functions, or cannot detect small samples of degraded images. The two general algorithms do not consider the characteristics of the channel and do not have a logic that has been optimized for wireless transmission distortion. Most of the lightweight deep learning algorithms that have been used so far have been trained on high-definition indoor static datasets and cannot solve the distortion caused by wireless conditions, such as radio-frequency interference and loss of transmission packets ^[1].

2.4. Goals of the optimization algorithm

Due to the operating difficulties of a wireless channel, an optimization algorithm has been added at all links of the preprocessing, feature extraction, and classification/recognition stages to solve the problems in all these links. Finally, it is hoped that fine restoration of multi-type channel-degraded facial image can be achieved, complete noise suppression, geometric distortion correction, and reconstruction of facial detail features can be performed, and the peak signal-to-noise ratio of the degraded image will be raised to over 32 dB. The second reason is to carry out anti-interference dual-domain feature extraction; that is to say, combine image pixel features with channel state correlation features to select high-discrimination face identification features and eliminate redundant interference features in the channel. First, the aim is to achieve high accuracy and low latency in the process of recognizing complex and changing channel conditions; among all the other main algorithms, this one can improve the recognition accuracy of degraded samples by 8–12%, maintain an inference delay of 20ms or less, be deployed lightly on embedded wireless terminals, and exhibit good generalization performance in adverse conditions. At the same time, due to engineering constraints, the total number of parameters in the algorithm must be kept below 8 million, and it should meet the storage and computing power requirements of mid-to-high-end embedded edge terminals; the amount of floating-point operations in the whole chain is also limited to meet the frame-rate demands for real-time video stream face recognition service.

3. Image preprocessing

3.1. Selection of noise reduction method

A hierarchical noise reduction method based on adaptive mixed filtering is used to solve the problem of Gaussian electromagnetic noise and salt-and-pepper impulse noise in the wireless channel. For general Gaussian channel noise abroad, an improved bilateral filter is used to smooth; that is to say, both the distance in space and how close the grey level is to the background noise are considered for smoothing, but sharp edges and uncovered areas of the facial contour are preserved. To solve the problem of local salt-and-pepper noise caused by data loss, an adaptive median filter algorithm will be used, and at the same time, according to different circumstances, the size of the filter window will also be changed dynamically to detect and correct isolated abnormal noise pixels. To solve the problem of having only one filtering algorithm, lightweight bilateral filtering is employed in the high signal-to-noise ratio situation to automatically select one of the two filtering logics according to the threshold of the channel signal-to-noise ratio; at a low signal-to-noise ratio degradation level, the two-filtering process is carried out in series to eliminate channel noise without losing key biological feature pixels in the face. Fix the logic error in the completion of boundary pixel filling for filtering, address the issue of excessive smoothing of facial contour edge pixels by traditional filtering algorithms, and do not lose boundary information ^[2].

3.2. Image enhancement technology

To solve the problems of low image contrast, loss of high-frequency details, and uneven brightness distribution caused by wireless transmission, a local adaptive multi-scale Retinex image enhancement algorithm is employed to improve the image. By separating the illumination component and the face reflection feature part in the algorithm, the global illumination distortion caused by deviation in wireless channel transmission is removed, and then non-linear grey mapping enhancement is applied to the face region of interest. In regions where there is a small amount of data and high discrimination, such as the eyes and contours of the face, increase the weight; at the same time, reduce the gain of pixels in background invalid areas to avoid overexposure and deformation of facial features due to global enhancement. The gamma correction module can dynamically adjust the dynamic range of different levels in a grayscale image according to how many grayscale levels are emitted by the output of different bandwidth wireless channels to correct damage to facial texture details caused by data transmission loss and improve the distinguishing ability of pixels with different classes. A constraint branch is added to partition the foreground region of the face and the complex monitoring background in the face mask branch accurately so that it is not interfered with by background noise under strong outdoor light and backlighting conditions, and to prevent disrupting the facial enhancement effect. A grey-scale saturation threshold is set to avoid too much enhancement that would distort the skin color features of the face, and thus the feature extraction module can accurately identify the biological characteristics of the face in the next step.

3.3. Geometric correction strategy

According to the distortion parameters of the wireless channel transmission and the matching of facial feature points, a sub-pixel closed-loop geometric correction method is used. Based on the standardized anchor point features of the extracted facial features, a standard frontal face geometric coordinate template is established, and then the translation, rotation, and perspective distortion offsets of the facial feature points in the degraded image are calculated. According to the phase distortion parameters of the wireless transmission link, a distortion transformation matrix is built to carry out a general pixel geometric inverse transformation and correct the stretching, tilting, and perspective distortion of the face image caused by multipath. A local micro-correction branch is added in the area of non-rigid minor deformation of the face, and local pixel fine-tuning is employed to correct local geometric distortion caused by electromagnetic wave scattering. Therefore, special hardware calibration equipment does not need to be employed, and it can also dynamically update the correction matrix in real time according to the channel transmission parameters to meet the demand for distortion correction of dynamic mobile wireless acquisition terminals. Due to large-angle deflection and severe perspective distortion of the sample, it is difficult to correct, so only the deformation range based on previous knowledge of the facial structure can be used. At the same time, normal physiological non-rigid deformation of the face should not be corrected during the correction process, and geometric standardization correction should not be carried out to avoid artificial geometric tampering that would lead to drift in identity feature samples ^[3].

4. Feature extraction and fusion

4.1. Local feature extraction

The goal of this paper is to have stable biological recognition of facial local regions, and a modified

SIFT algorithm is used to extract strong local features from the channel. The four areas of the face that are relatively immune to interference are selected: the outline of the eyebrows and eyes, the ridge lines of the nose, the texture of the lips, and the moles on the face. In order to reduce the range of coverage and computing power for feature extraction and computation, it may be that there is interference due to background noise in the wireless channel. According to the requirements of medium-resolution degraded facial image after preprocessing, the hierarchy of the Gaussian difference pyramid has been optimized to reduce the rate of missed detection of feature key points due to image blur. A local feature confidence screening mechanism is added to remove the pseudo feature key points that are caused by channel distortion, and then local feature descriptors with geometric invariance and gray-scale anti-interference are obtained. Generally speaking, the above type of features is related to the microbial characteristics of the face and has good natural fault tolerance to small channel deformation. It can help address the problem of missing global features in the recognition of degraded samples. It reduces the sampling density of key points in weak feature areas and increases the weight of retained high-stability facial turning point features to reduce the deviation of feature description vectors caused by noise residuals.

4.2. Global feature extraction

A light convolutional neural network branch is used to extract the general characteristics of degraded facial image. It extracts the general contour of the face and the topological relationship between different parts of the face in the entire image for semantics. It reduces the number of redundant convolutional layers in the traditional convolutional backbone network and adds a channel feature normalization module to solve the problem of global pixel amplitude offset caused by different signal-to-noise ratios of wireless channels. Pooling is employed to merge the sampled multi-scale global semantic features and keep the spatial structure information of the face; thus, local feature information will not be lost in recognition. For the samples in the local face area that have been affected by severe packet loss, the algorithm will reconstruct the missing semantic information based on global topological features; and it will not be impossible to identify the sample because only one local feature is required. Both the global feature branch and the local feature branch are used to align the output sequences of multi-scale features. Depth-separable convolution is chosen to replace the standard convolution in order to reduce the number of parameters and computation of the global feature extraction branch and meet the upper limit of computing power of the wireless terminal ^[4].

4.3. Feature fusion method

A weighted adaptive feature fusion algorithm based on the attention mechanism is used to connect and merge the local texture features of the face with the global topological features. A channel attention module is added to distribute the fusion weight of the two feature branches automatically, and at the same time, this weight ratio can be dynamically changed according to the actual wireless channel conditions. If there is a poor channel, more weight will be given to high-stability local features; when the channel is good, the weight of global semantic features will be relatively high. The feature layer combines the cross-domain features and normalizes them to the same dimension as the other feature vector to solve the problem of different data distributions in the two feature branches. Redundant channel interference feature vectors are not employed in the fusion process, and a combined face feature vector of microscopic biological texture and macroscopic spatial structure is obtained. Now the fusion mode does not have the scene constraints of the fixed-weight

fusion algorithm and is more in line with the dynamic and time-varying transmission conditions of wireless channels. The embedded channel state encoding vector is used as the bias term for the attention weight, and the fusion logic of channel parameters and features is more concise.

4.4. Feature selection strategy

To find a good set of features for the wireless degraded scenario, a simplification and optimization method for the high-dimensional fused feature vector will be used in this paper, which is the Relief-F feature selection algorithm. Calculate the category discrimination weight for all fused feature vectors, and then select the main feature vector that is sensitive to the face identity label and insensitive to channel noise. It does not have the bad noise of channel distortion either. Set a threshold for the feature dimension to limit the scale of the output feature vector, reduce redundant coupled features, and thereby decrease the inference computational load of the back-end classification model. Keep the principal components of the main features for face identity recognition and discard the secondary feature components that are related to the background, light, channel fluctuations, etc., to obtain a low-dimensional, highly robust, and easily identifiable final face feature dataset. It can reduce the number of redundant feature parameters by more than 35% without losing the main identity identification information. In order to reduce the number of collinear feature vectors and avoid overfitting caused by redundant features in the classification model, an orthogonal feature check step has been added. A dynamic dimension threshold is set, and if there is a considerable decrease in performance, the number of features is reduced reasonably to make up for the loss of some necessary information due to image deformation and maintain an equilibrium between feature compression and recognition accuracy.

5. Classification and recognition model

5.1. Model type selection

In order to meet the demands of a lightweight deployment of wireless edge terminals, an improved lightweight capsule network is chosen as the main face classification and recognition model. The capsule network is more suitable for the problem of inputting incomplete and distorted features of degraded facial image than traditional convolutional networks because it can keep the positional correlation relationship in the feature space and will not lose the spatial topological information of facial features due to convolutional pooling operations. Due to the weak computing power of wireless terminals, the number of neurons in the primary capsule layer of the capsule network has been reduced, and the number of dynamic routing iterations has also been decreased to reduce the quantity of floating-point operations for the model. An interface for the channel state embedding is added to the input end of the model, and the real-time wireless channel signal-to-noise ratio and packet loss rate are employed as auxiliary input parameters to improve the adaptive classification ability of the model in the presence of various levels of sample degradation. The model is also relatively small and fast, meets the requirements of an edge device, and is suitable for wireless transmission of face recognition engineering, etc. The improved capsule network has increased the activation of low-quality feature input, so the invalid secondary routing path does not need to be used, and the serial computation time at the edge terminal has been reduced. The model can be used in the main inference framework of embedded devices (TensorFlow Lite); INT8 quantization deployment is supported, and it can be used with wireless monitoring hardware motherboards that have ARM architecture.

5.2. Model training process

The whole recognition model is trained by means of a phase transfer learning method to address the problem that there is a small number of wireless degraded face sample datasets. The first step is to pre-train the model with the public high-definition face benchmark dataset to learn the general distribution of facial features and basic biological texture characteristics, and the parameters of the basic feature extraction layer at the bottom of the network are frozen. The second stage is trained on synthetic channel-degraded face datasets to train the model further; thus, it can learn the distribution laws of all types of channel-distortion features and adapt to the characteristics of the input distribution of degraded features. In the third stage, a small number of real wireless transmission face samples are employed for closed-loop fine-tuning to update the parameters of the dynamic routing classification layer. A way of batch training with mixed noise samples is added to the training period to improve the generalization ability of the model under all kinds of channel degradation samples, and an early stopping function is used to avoid overfitting. To ensure that the weights of the classification layer in the latter stages of the model converge smoothly and do not get stuck in a local optimum, cosine annealing learning rate decay is employed. The batch size for the identity samples is set so that there are not too many similar samples in a single batch, which would cause feature bias in the model and reduce the generalization ability of the cross-scenario wireless samples ^[5].

5.3. Model parameter optimization

A modified Particle Swarm Optimization algorithm will be used to optimize all the parameters and weights of the model generally. The three objectives of the multi-objective optimization are the accuracy rate of degraded face recognition, model inference delay, and edge-terminal memory consumption; the other four main hyperparameters of the capsule network routing that are optimized are the number of iterations, learning rate, regularization coefficient, and capsule neuron dimension. To solve the problem of local optima in the conventional particle swarm optimization, an adaptive inertia weight is used; then, convergence is carried out to obtain the optimal model parameters for the wireless environment. L2-regularisation has the problem of reducing the magnitude of the network's weights to prevent overfitting in the face of channel noise. Pruning and Quantisation will not reduce the recognition accuracy much, and at the same time, given the limited computing power of the embedded wireless monitoring terminal, the memory required by the model will be reduced. Add a convergence threshold judgment condition to limit the number of iterations and prevent reaching the engineering performance indicator; otherwise, it will take too long to optimize the large number of parameters. Set an upper and lower bound for the sensitive hyperparameter; otherwise, it will be too large or too small, and the model will not run stably on the edge device.

5.4. Model performance evaluation

A multi-dimensional index system is employed to complete the overall performance assessment of the recognition model, and according to the different parts of the steady-state channel and fluctuating channel test scenarios, it is divided. The evaluation indicators for the recognition accuracy of the model are the recognition accuracy rate, recall rate, and F1 score; they can quantify the classification results of various degraded-level face samples, and the evaluation indicators for engineering deployment performance are the model's floating-point operation quantity, single-frame inference delay, and device memory usage. Strengthen the direction of robustness evaluation to determine how much the model's performance will decrease in different circumstances of signal-to-noise ratio and all sorts of degrees of data packet loss in

the channel. Under good channel conditions in steady state, the recognition accuracy is 98.7%, and it is still more than 91% even in the presence of severe fluctuation and degradation. All of the above indicators are better than those of the traditional face recognition model at the same scale, and it can meet the real-time recognition engineering performance requirements in wireless environments. Long-term and long-cycle tests will be carried out to check for the stability of the model, and after 72 hours of continuous streaming input of wireless-transmitted video face samples, the drift in model performance will be statistically evaluated.

6. Algorithm verification and application

6.1. Construction of the experimental dataset

An experimental dataset of the mixture of exclusive wireless channel degradation was created, and it contained both actual collected samples and simulated synthetic samples. Actual samples were collected wirelessly in a typical wireless environment, for instance, indoors where there was occlusion, outside during movement, and subjected to electromagnetic interference; at the same time, multiple-identity facial image were taken, and synthetic samples were generated based on the standard public face dataset. All kinds of large-scale fading and multipath effects were simulated by adding various amounts of Gaussian noise, salt-and-pepper noise, and geometric distortion. Based on gender, age, collection scene, and channel degradation level, the dataset was divided and classified. Then a training set, a validation set, and a test set were separated with a ratio of 4:1:1. Fuzzy invalid and abnormal samples were eliminated. The number of positive and negative samples, as well as samples with different degrees of degradation, were balanced to prevent bias in the experimental results due to data skewness and to guarantee the objectivity of the experimental results. Labels for the channel operating parameters corresponding to them were added to the labeled samples to link the input of image features with channel parameters for algorithm training.

6.2. Construction of experimental environment

A closed-loop experiment environment for hardware and software has been created to simulate the whole process of wireless transmission and recognition of facial image in reality. At the hardware level, a software radio device was used to construct adjustable-parameter wireless channels, and at this time, an embedded edge computing terminal, high-definition face-capture camera, network transmission module, etc., were also set up; the main channel parameters could be precisely adjusted, such as but not limited to channel signal-to-noise ratio, transmission bandwidth, data packet loss rate, and Doppler frequency shift. PyTorch was used for the implementation of the algorithm code at the software level for deep learning, and the same version of the image processing dependency library and model compilation environment was set up. The power parameters of the unified experimental hardware and the running hyperparameters of the software were set to avoid errors due to the hardware device system. All the comparison algorithms were run on the same software and hardware to avoid any influence from experimental environment factors on the evaluation results of the algorithms. A log collection module was built to record the parameters of the channel, the computation power occupancy rate, and the algorithm's run delay at various times during the experiment, and also to store trace data of the whole experiment.

6.3. Algorithm comparison experiment

Some sets of horizontal comparison experiments have been carried out to verify the effect of the optimized

algorithm. The three control models are the old LBP-SVM algorithm, the native CNN face recognition algorithm, and the standard capsule network algorithm. According to the three levels of wireless channel degradation (low, medium, and high), synchronous tests were carried out to determine the recognition accuracy, single-frame processing delay, and robustness of the four algorithms in a harsh environment. Based on the data from the horizontal comparison experiment, under the conditions of moderate and severe channel degradation, the recognition accuracy of the optimized algorithm is much higher than that of the three control algorithms, and the total link single-frame processing time also meets the requirements for real-time recognition. Based on the ablation study, the pre-processing module, the attention feature fusion module, and the channel parameter embedding branch have all improved the performance of degraded face recognition; therefore, it can be concluded that each optimization module is beneficial. There are no invalid redundant algorithm structures. A test of statistical significance is carried out on the standard deviations of many repetitions of the experiment to verify that the improvement in algorithm performance is statistically reliable.

6.4. Analysis of practical application

The optimized algorithm can be used in the three general cases of wireless transmission and face recognition engineering, and it is small enough for the business needs of the edge. The first type is the outdoor wireless security access control scenario; it has the problem of face recognition failure due to electromagnetic interference and shadow fading in the open-air wireless monitoring link of the park. The second type is the mobile terminal remote identity verification scenario, which is suitable for the remote face comparison business of mobile phones under fluctuating channel conditions of cellular wireless networks. The third type is the industrial wireless robot visual identity control scenario; it solves the problem of face distortion in close-range face recognition in complex electromagnetic environments in industrial workshops. It is a small algorithm that can be carried out at the edge for inference without the need for large computing power in the cloud. It can be added to the existing hardware system of the wireless vision directly, and only the channel adaptation parameters need to be slightly modified before deployment in the scenario. The algorithm is in the private transmission protocol of the main security device, so it does not need to modify the existing hardware transmission link and will reduce the budget for engineering construction and renovation.

7. Conclusion

Optimization and Recognition of Degraded Wireless Channel Images are the problems that algorithms in this field need to solve. Based on the foundation of algorithm design, this paper will present an all-encompassing optimization recognition algorithm covering image pre-processing, feature extraction and fusion, classification recognition model, etc. Based on the experiment, the above algorithm can improve the recognition rate of degraded facial image. In the future, more adaptive research will be conducted on the algorithm in all kinds of complex wireless channel environments, and its scope of application will also be expanded. At the same time, with the help of new technologies such as deep learning, continuous improvements will also be made to enhance the performance of the algorithm and provide stronger support for face image recognition in the field of wireless communication.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Shen D, 2026, Design of Smart Park Access Control System Based on Wireless Communication Technology. *China Broadband*, 22(1): 140–142.
- [2] He F, 2025, Research on Secure Communication Protocol for Intelligent Lock Based on Wireless Communication Technology. *China Broadband*, 21(10): 85–87.
- [3] Zhang J, Ning B, 2024, Integration Application of LTE Wireless Communication Technology and Internet of Things Technology. *Satellite Television and Broadband Multimedia*, 21(23): 13–15.
- [4] Cai Y, Sun Z, Cao S, et al., 2024, Train Network Architecture and Secure Transmission Technology Based on Wireless Communication. *China Railway*, 2024(10): 106–113.
- [5] Liu X, 2024, Research on Video Monitoring and Face Recognition Technology Based on Wireless Communication. *Science and Technology Horizons*, 14(25): 49–51.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Study on the Enhancement of PM_{2.5} Air Purification Capability in Cyclone Separators with Curved Structures Based on CFD Simulation

Yitong Qiu*

High School Affiliated To JiaoTong University JiaDing Campus, Shanghai 201821 China

*Corresponding author: Yitong Qiu, 969073704@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Driven by the demand for purifying indoor air from particulate matter, this study has carried out systematic numerical research on the effect of cyclone separator geometry on the performance of PM_{2.5} filtration. A comprehensive CFD analysis system is constructed based on three cone segments generatrices: straight, concave, and convex. The continuous phase in the simulation adopts the Reynolds Stress Model (RSM) due to its high ability to predict the strong anisotropic turbulent flow inside the cyclone separator. The particle phase in the present study is simulated through the Discrete Phase Model (DPM) by the Lagrangian method and considers the effects of drag force, centrifugal force, and turbulence diffusion to predict the migration and deposition process of particles with various diameters. From the results, the convex generatrix shows great superiority in capturing fine particles, obtaining a high efficiency of PM_{2.5} separation up to 82.4% and a small cut-off diameter down to 2.1 μm . In comparison, the concave generatrix outperforms in reducing the pressure drop and optimizing the energy efficiency ratio, while the straight generatrix ranks in between. The results show that the reasonable control of geometric curvature can significantly improve the flow field stability and particle distribution characteristics.

Keywords: Cyclone separator; PM_{2.5}; CFD; Grade efficiency

Online publication: Aug 11, 2026

1. Introduction

As industrialization and urbanization advance rapidly, air quality safety has become a key issue for society. Taking Shanghai as an example, indoor concentration of PM_{2.5} changes periodically due to such reasons as population, traffic, and cooking^[1]. However, traditional ways of purifying air, including mechanical filters, electrostatic precipitation and photocatalysis, are inefficient in many aspects: filters require frequent change of consumable components, consume significant amounts of energy and produce considerable noise; electrostatic precipitation creates undesirable products, requires complicated maintenance procedures,

and photocatalysis depends on light sources and can hardly provide stable results for a long time period [2–5]. Aiming at indoor scenes such as home and office, this paper constructs an air purification system that integrates the structural design of curve cyclone separator and CFD numerical simulation, regulates the geometric parameters and internal flow field of cyclone separator, accurately predicts the separation efficiency and pressure drop characteristics of PM2.5, and explores the best combination of structural parameters, so as to provide support for improving the environment, reducing risks and promoting urban green development.

Traditional PM2.5 air purification methods primarily rely on physical or chemical technologies such as mechanical filters, electrostatic precipitation, and photocatalysis. Although these approaches can reduce indoor PM2.5 concentrations in the short term, they are limited by efficiency decay, high energy consumption, secondary pollution, and high maintenance costs [6]. For instance, HEPA filters are prone to clogging, electrostatic purifiers may generate by-products and suffer from efficiency decline, and photocatalytic methods exhibit insufficient stability [7]. By contrast, cyclone separators are simple in structure, durable and reusable, and thus have application potential. However, their low capture efficiency for fine particles $\leq 2.5 \mu\text{m}$ limits their widespread adoption [8].

As a result of the advancements in CFD technology, structural optimization of cyclone separators has gained great attention recently. Parameters such as the design of the inlet, cylinder geometry, and cone angle affect the flow field inside the separator and the trajectory of the particles [9]. The Reynolds Stress Model (RSM) and Large Eddy Simulation (LES) have been employed to simulate the turbulence phenomenon. RSM models vortices, whereas LES increases the accuracy of the simulation for the separation of sub-micron particles. As far as structural optimization is concerned, a curved conical section is one of the most important methods for enhancing efficiency [10]. This is because a curved conical section can affect the vortex intensity and particle residence time. The research shows that the design of a symmetrical double inlet and variable diameter cone section can reduce the pressure drop and improve the PM2.5 separation efficiency.

Overall, research on cyclone separators for indoor PM2.5 purification has shifted from empirical design to precise CFD-based optimization, exhibiting trends toward miniaturization, low energy consumption, and intelligent operation. In the future, by integrating multi-scale simulation, experimental validation, and intelligent optimization algorithms, novel cyclone-based air purification solutions are expected to be realized.

This project will establish a method for cyclone separator air purification based on CFD numerical simulation and geometric structure optimizations. The proposed approach is capable of rapidly and accurately predicting the separation performance of cyclone separators under different structural parameters and operating conditions (PM2.5 capture efficiency, cut-off particle size, pressure drop, etc.). Firstly, a parameterizable simulation model of the cyclone separator will be constructed, enabling visual analysis of internal flow field characteristics and particle motion patterns. Subsequently, the separation performance of different conical sections, including straight, concave, and convex profiles, will be compared to identify key geometric variables. Finally, low-energy air purification design solutions suitable for domestic and office environments will be proposed, providing theoretical and technical guidance for indoor PM2.5 control.

2. Design of cyclone separator

2.1. Principle of the cyclone separator

The cone section of the traditional cyclone separator is designed in a straight line, and the fluid flow is prone

to turbulence disturbance, boundary layer separation, and other problems, resulting in limited separation efficiency, such as insufficient classification accuracy of fine particles and high energy consumption. In this design, the spline curve is introduced to reconstruct the internal cone section profile of the cyclone separator to improve the fluid flow characteristics.

The mechanical aspects of the radial forces that act on particles within the hydrodynamic behavior of the hydro cyclone with the curved profile are as follows: Centrifugal force (FC), radial pressure gradient force (FB), Stokes drag force (FD), and Magnus lift (FM). Of all these mechanical forces, the centrifugal force and the radial pressure gradient force form the most significant forces in the radial motion of the particles, which together determine the radial motion trajectory of particles. See **Table 1**.

Table 1. Radial forces acting on particles in a hydro cyclone

Force name	Formula	Number
F_c	$F_c = m \frac{v_t^2}{r} = \frac{\pi \rho_p d_p^2 v_t^2}{6r}$	(2)
F_B	$F_B = v_p \rho_f \frac{v_t^2}{r} = \frac{\pi D_p^3 v_t^2}{6r} \rho_f$	(3)
F_D	$F_d = 3\pi D_p \mu v_r$	(4)
F_M	$F_m = k \rho_f D_p^3 \omega v_r$	(5)

2.2. Design of structural parameters for cyclone separators

In the geometric design of the cyclone separator, the cylindrical-cone generatrix is usually based on a linear structure, that is, a constant cone rate is formed by a linear connection between the inlet cylindrical radius and the outlet radius. This straight generatrix stands as the simplest geometric construction and contributes to preserving gas flow continuity through the cylinder–cone transition. However, a straight conical section alone often fails to balance separation efficiency and pressure drop when finer regulation of the flow field characteristics is sought. Spline curves are therefore introduced to adjust the curvature of the straight generatrix. Bending the middle section of the curve inward produces a concave generatrix, which shortens the radial migration path of particles and, in turn, lowers the pressure drop. When the curve extends outward in the middle section, forming an outwardly convex bus bar, it helps to enhance the centrifugal force and the particle retention time, thereby improving the capture efficiency of fine particles. Because the three profiles keep the same parameters at both ends, the transition from the straight profile to the concave or convex generatrix can be realized continuously by changing the sign of the curvature parameter, thus yielding a set of geometric schemes that can be systematically compared. See **Table 2**.

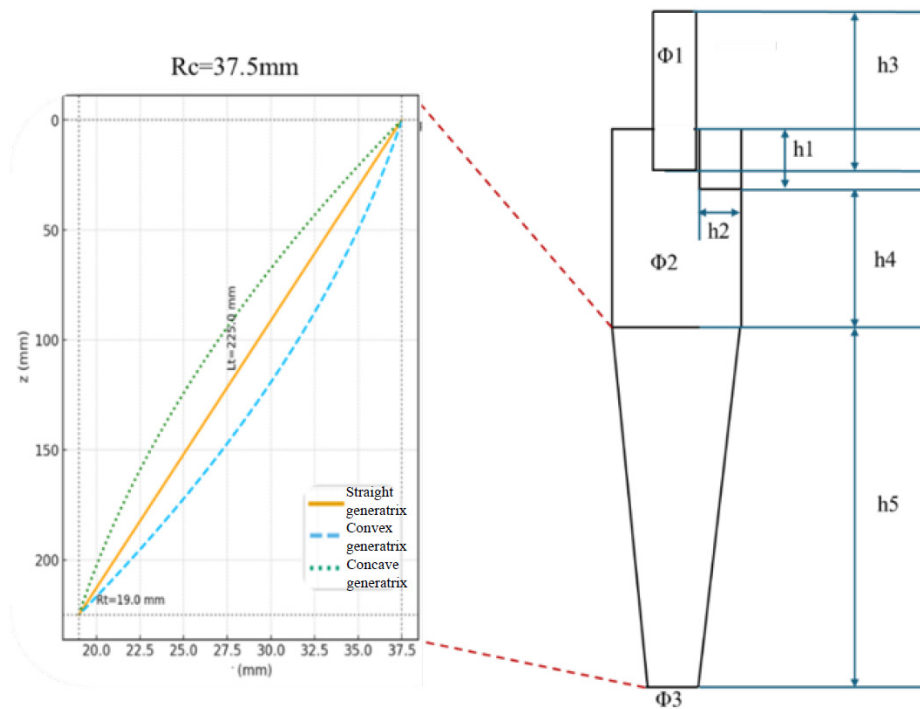
The equation for a K-order spline curve is expressed as follows:

$$P(t) = \sum_{i=1}^{n+1} B_{i,k}(t) P_i, t \in [0, 1] \quad (7)$$

Table 2. Dimensional parameters of the cyclone separator

Component	Symbol	Size (mm)
Overflow pipe diameter	Φ_1	35
Overflow pipe height	h3	138.6
Column section height	h4	150
Conical section height	h5	225
Column section diameter	Φ_2	75
Outlet diameter	Φ_3	38
Inlet pipe length	h1	45
Inlet pipe width	h2	30

Figure 1 compares three typical generatrix profiles for the conical section of a cyclone separator, namely straight, concave and convex. The straight generatrix is formed by a linear connection between the inlet radius $\Phi_2 = 78$ mm and the outlet radius $\Phi_3 = 38$ mm, with a conical section height of $h_5 = 225$ mm. This is the most common baseline structure. The concave generatrix contracts inwards in the middle section. Compared with the straight profile, it exhibits a faster radial convergence, which can shorten the particle migration path. In the middle portion, the concave generatrix tends to contract inwardly. When compared to the straight generatrix, it shows faster radial contraction that may help reduce the distance of travel of particles. However, in the case of the convex generatrix, it is extended outwardly in the middle portion, and hence there is fast convergence of the fluid in the first portion and leveling in the last portion, thereby helping to enhance the centrifugal separation of particles. Since the geometric parameters at the two ends remain constant, the three generatrices differ solely in the curvature of their middle sections. This sets up a uniform geometric foundation for carrying out comparative CFD simulations.

**Figure 1.** Schematic diagram of a cyclone separator.

3. Selection of CFD

Strong swirl and complex backflow occur in the internal flow of the cyclone separator, resulting in turbulence anisotropy, which influences the gas-solid separation efficiency. The traditional two-equation turbulence model has limitations in predicting strong swirl and strong curvature flow, and it is difficult to reflect the non-uniform turbulence characteristics in the separator.

Through the solution of the equations for Reynolds stress transport, the Reynolds Stress Model (RSM) incorporates the presence of anisotropic stresses and rotation, thus allowing for a more accurate representation of phenomena like the recirculation core in the separator. Several studies have proven that the RSM is better than the two-equation models when it comes to modeling the pressure drop behavior of the separator, making it one of the commonly used methods for numerical simulation of cyclonic separation.

$$\frac{\partial}{\partial t}(\rho \overline{u_i u_j}) + \frac{\partial}{\partial x_k}(\rho U_k \overline{u_i u_j}) = D_{ij}^T + D_{ij}^L + P_{ij} + G_{ij} + \Phi_{ij} + \varepsilon_{ij} + F_{ij} \quad (8)$$

In the simulation of gas-solid two-phase flows, because of the disparity in scale between the particle phase and the gas phase, the Discrete Phase Model (DPM) is usually adopted in combined calculations to make a compromise between calculation efficiency and accuracy. In this way, the Eulerian method is adopted for solving the flow field of the continuous phase, while the Lagrangian method is adopted for calculating the trajectories of the particles, which can consider a variety of forces. For fine particles, DPM can quantitatively analyze their behavior within a swirl flow field. Combined with the high-precision turbulent flow field provided by RSM, it provides a basis for the optimization and evaluation of separators.

Coupling RSM with DPM makes it possible to numerically uncover the flow structure and particle separation mechanisms inside a cyclone separator. RSM provides an accurate simulation of the flow field, while DPM enables visualization and quantification of particle behavior. Used together, they offset the shortcomings of single turbulence models, supporting structural optimization and design, and form the core technical approach for investigating the air purification performance of separators.

In the simulation, the inlet boundary is set as a velocity inlet with $v = 10 \text{ ms}^{-1}$, and the overflow and bottom outlets are set as flow outlets, with a diversion ratio of 20%. The pressure is discretized using the PRESTO scheme, while the pressure-velocity coupling is modeled using the SIMPLE algorithm, and the other governing equations are discretized using the QUICK scheme. The convergence criterion for the calculation is set at 1.0×10^{-5} .

4. Analysis of results

4.1. Particle trajectory plots

A comparison of the particle trajectories shows that the residence time and the distribution of motion paths are strongly influenced by particle size. The $0.5 \text{ }\mu\text{m}$ fine particles, with their extremely low inertia, are highly susceptible to disturbances within the flow field. Their trajectories take the form of intricate spiral patterns, accompanied by a substantially prolonged residence time inside the chamber, with some values reaching close to 0.45 s. These particles exhibit pronounced flow-following and retention behavior, making them especially conducive to mass transfer and turbulence enhancement. Conversely, the $3 \text{ }\mu\text{m}$ particles, owing to higher inertia, show poor flow-following behavior; the trajectory spread is narrow, and their

residence time is lower (generally around 0.1 and 0.25 s). The rotating flow of these particles shows a spiral expansion nature; however, the radial diffusion distance is lower. In the case of 8 μm particles, the effect of inertia is dominant. The trajectories of such particles generally take the shape of regular spirals and do not respond to any small disturbances in the flow field. They have the lowest residence time (generally below 0.1 s) and move fast through the flow field. Overall, as particle size increases, particle flow-following behavior weakens, residence time shortens, and the effects of inertia and straight-through characteristics are enhanced. (Figure 2–4)

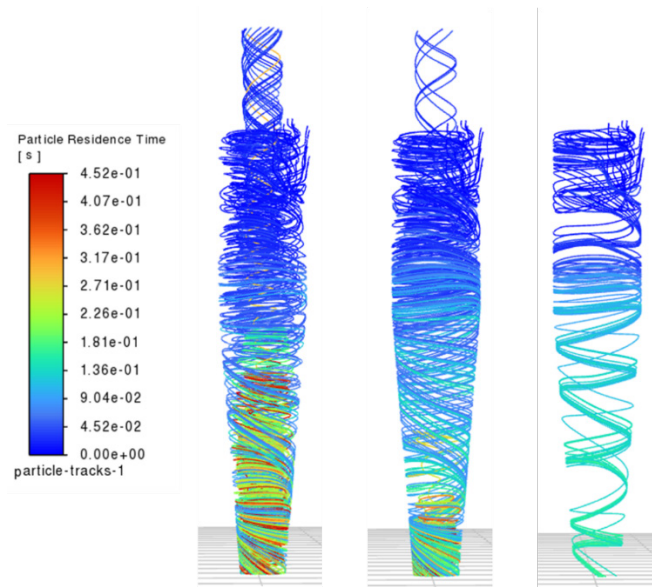


Figure 2. Particle trajectory plots for a straight-profile cyclone separator (left: 0.5 μm ; middle: 3 μm ; right: 8 μm).

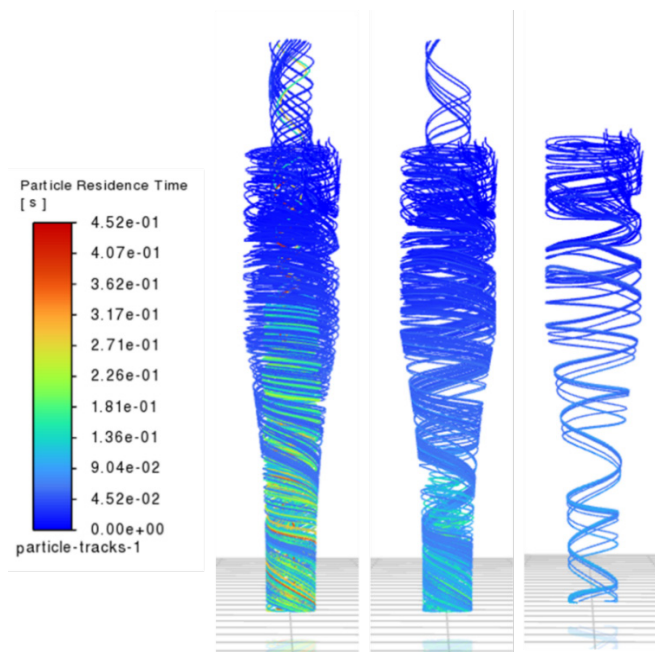


Figure 3. Trajectory diagrams for a convex-profile cyclone separator (left: 0.5 μm ; middle: 3 μm ; right: 8 μm).

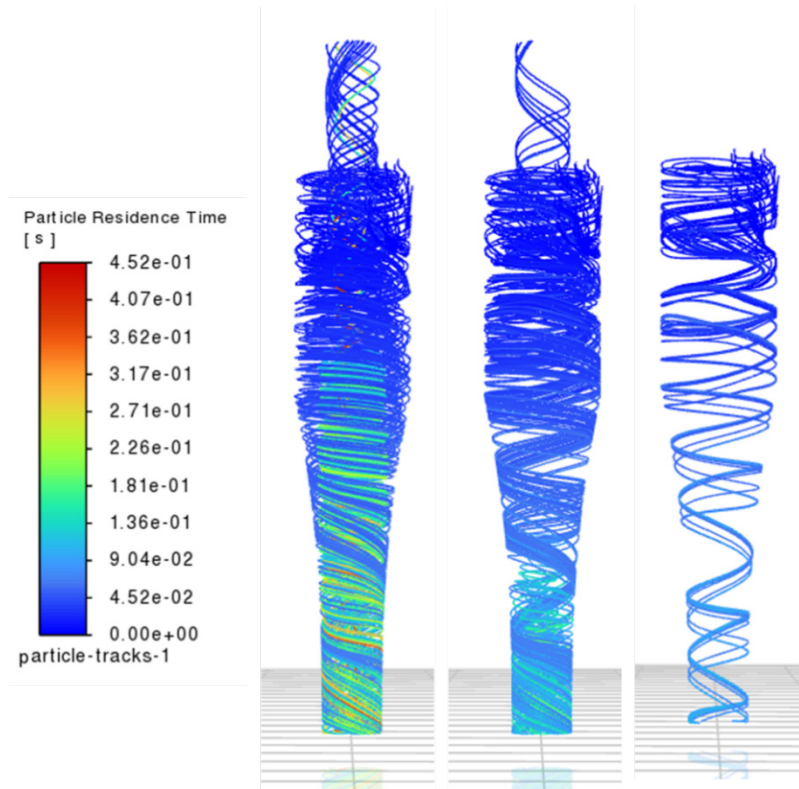


Figure 4. Trajectory diagrams for a concave-profile cyclone separator (left: 0.5 μm ; middle: 3 μm , right: 8 μm).

The distribution of particle trajectories shows that particle size plays an important role in the residence characteristics of particles inside the cyclone separator. The trajectory of 0.5 μm fine particles is complicated, with multiple cycles, and the residence time is about 0.45 seconds, depending on the disturbance of the flow field. The trajectory of 3 μm particles is regular with a residence time of 0.1–0.2 seconds, which is a trade-off between flow-following and inertia. 8 μm particles are controlled by inertia, and most of their residence times are less than 0.1 second with straight-through trajectories. With the increase in particle size, the motion changes from flow-following to inertia-controlled, which reveals the classification effect of the cyclone separator for particles of different sizes, that is, retaining and mixing of small particles while separating large particles.

The characteristics of particle entrainment in the three cyclones are compared. The entrainment of 0.5 μm small particles in the straight cyclone is intensive and long-term; at the same time, the inhomogeneous particle distribution is likely to create non-uniformity of flow and high energy losses. On the other hand, the two other cyclone separators perform better both in terms of particle separation and flow stability. In the convex separator, there appears a localized region of intensive particle entrainment that reduces the losses due to turbulence, and the particle trajectories for medium-sized and large particles become regular. The concave cyclone separator further strengthens this trend; that is, the separation effect is more controllable, and additional loss is avoided.

4.2. Grade efficiency

The data in the table above indicate that the convex-generatrix cyclone offers higher overall efficiency than the other two cyclone separators. Within the fine particle size interval (0.5–3 μm), the separation efficiency

of the convex-generatrix design is substantially greater than that achieved with the straight and concave generatrices. For 2 μm particles, the convex cyclone separator reaches an efficiency of 0.746, compared with 0.63 for the straight and 0.638 for the concave configurations. At 3 μm , the efficiency of the convex cyclone climbs to 0.766, again surpassing the other two designs. These results provide clear evidence that the convex geometry brings a notable advantage in fine-particle capture and separation performance.

As particle size increases ($\geq 5 \mu\text{m}$), the grade efficiencies of the three cyclone profiles gradually converge toward 1, and the performance gap between them becomes negligible. In the low-to-medium size range, the convex-generatrix cyclone stands out as the most capable for fine-particle separation. The straight profile shows a modest edge for particle sizes around 3–4 μm , while the concave cyclone delivers a relatively balanced efficiency across the measured range, though it consistently falls short of the grade efficiency levels reached by the convex profile. See **Table 3**.

Table 3. Grade efficiency of cyclone separators

Particle size	Type of cyclone separator		
	Convex-generatrix cyclone separator	Straight cyclone separator	Concave-generatrix cyclone separator
0.5 μm	0.6	0.567	0.446
1 μm	0.567	0.5	0.574
1.5 μm	0.567	0.63	0.638
2 μm	0.746	0.633	0.638
3 μm	0.766	0.833	0.83
4 μm	0.9	0.89	0.829
5 μm	1	0.90	0.914
6 μm	1	1	1
7 μm	1	1	1
8 μm	1	1	1

4.3. Cloud chart analysis

4.3.1. Static pressure analysis

The static pressure distributions of the three cyclone separators are compared below. The straight profile creates a distinct low-pressure zone. A strong negative pressure core develops locally upstream with pressure close to -1200, driving flow separation and recirculation together with unstable vortex motion, all of which degrade system stability and efficiency. The convex section has a more consistent pressure distribution pattern, where the number of low-pressure areas decreases, and there are no extremely low-pressure cores; the flow becomes consistent, and there are minimal losses of energy. The concave section shows an increase in pressure distribution near the wall, where a high-pressure region develops on the edge, while the low pressure in the center of the duct decreases and approaches zero. Nevertheless, the presence of the high wall pressure makes the load greater, hence higher requirements for the engineering materials used. Generally, the profile that gives the best performance by minimizing vortex core under low pressures and ensuring uniform pressure distribution is the convex one; however, the straight and concave profiles have the risks of flow instability and high loads, respectively. See **Figure 5**.

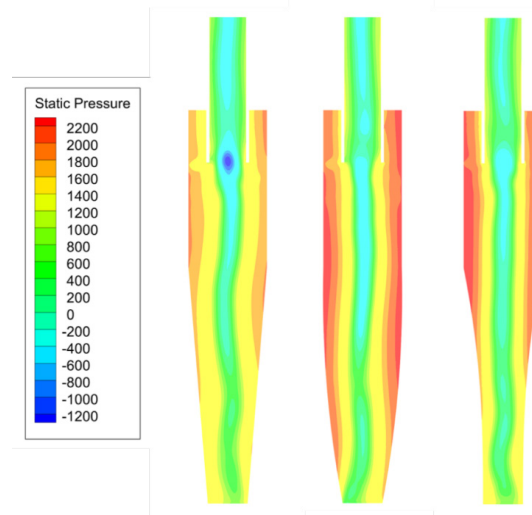


Figure 5. Static pressure contour map of a cyclone separator.

4.3.2. Tangential velocity

Tangential velocity comparisons show that all three cyclone separators develop significant swirl structures downstream of the inlet, differing in both intensity and distribution. The straight-profile separator generates a typical symmetrical rotating core together with a central counter-rotating flow, which results in high swirl intensity. The convex profile, on the other hand, exhibits a more vigorous rotational structure characterized by extensive and asymmetrical red zones of high tangential velocity. Here, the central counter-rotating flow spreads over the widest area, leading to increased flow instability. By contrast, the concave cyclone displays reduced vorticity intensity, a shrunken counter-rotational region, and rapid downstream vorticity decay, all of which yield a more stable flow. Overall, the convex cyclone reaches the highest rotational intensity at the cost of stability, whereas the concave cyclone lowers rotational energy while improving both flow uniformity and stability. See **Figure 6**.

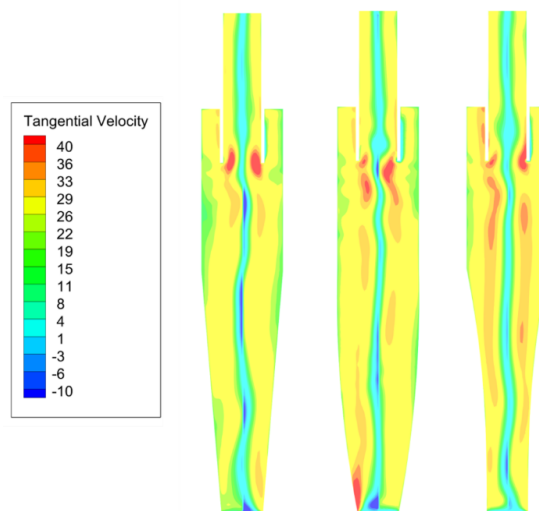


Figure 6. Tangential velocity contour plot of the cyclone separator.

4.3.3. Axial velocity

Axial velocity profiles for all three cyclone separators show typical features of central high-velocity jets, although differences are seen in terms of intensity and diffusion. For the left-side profile, the jet core is compact, energy is well preserved at a long distance, and there is good axial momentum transfer capability. But the problem with low recirculation in the boundary region may cause instability in the flow regime. In the case of a convex cyclone separator, there is a reduction in the size of high-velocity jets and rapid decay in velocity downstream, which shows greater diffusion of momentum and improved flow field uniformity. In the case of a concave cyclone separator, the high-velocity jets have smaller sizes and certain intensity is maintained downstream. However, the recirculation structures are considerably large, suggesting that this profile is linked to greater flow instability. The convex profile proves more favorable for jet diffusion and flow uniformity, whereas the straight profile tends to preserve energy efficiently, and the concave profile is characterized by recirculation instability. See **Figure 7**.

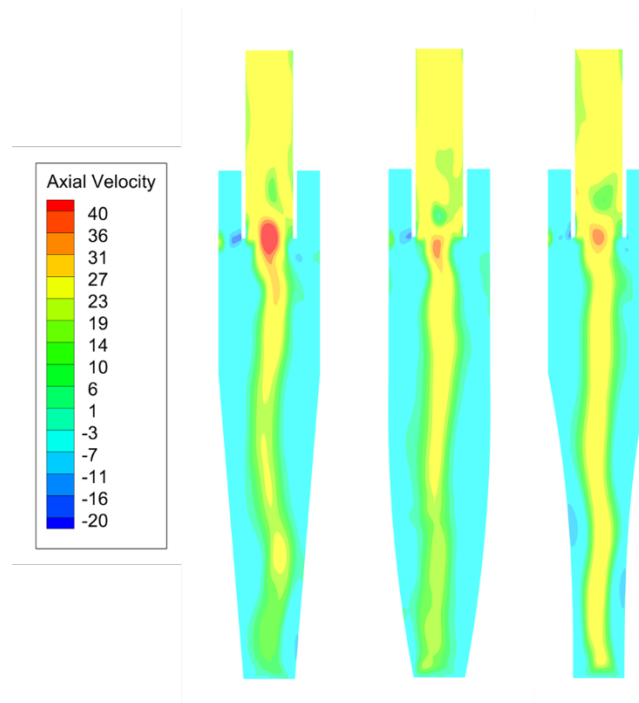


Figure 7. Axial velocity contour plot of cyclone separators.

4.3.4. Radial velocity

Downstream of the inlet, the radial velocity fields of the three cyclone separators all exhibit alternating positive and negative banded structures within the shear layer, though these bands differ notably in amplitude and spatial extent. The straight cyclone separator keeps radial velocity magnitudes at their lowest and restricts lateral momentum transport, producing a strong axial jet core. The convex cyclone separator, in contrast, generates intermediate radial velocity bands that are more continuous and symmetric, indicative of enhanced shear layer growth and entrainment, which act to expand the core region and foster downstream mixing. The concave cyclone separator exhibits the strongest and most asymmetrical radial motion, with radial velocity bands that extend toward the wall and retain high intensity downstream. This reflects more vigorous transverse momentum transport and three-dimensional effects, but it also signals greater flow instability and

potentially higher wall loads. In summary, the convex cyclone separator strikes a better balance between mixing efficiency and flow stability, while the straight and concave profile lean towards extreme stability and mixing capabilities, respectively. See **Figure 8**.

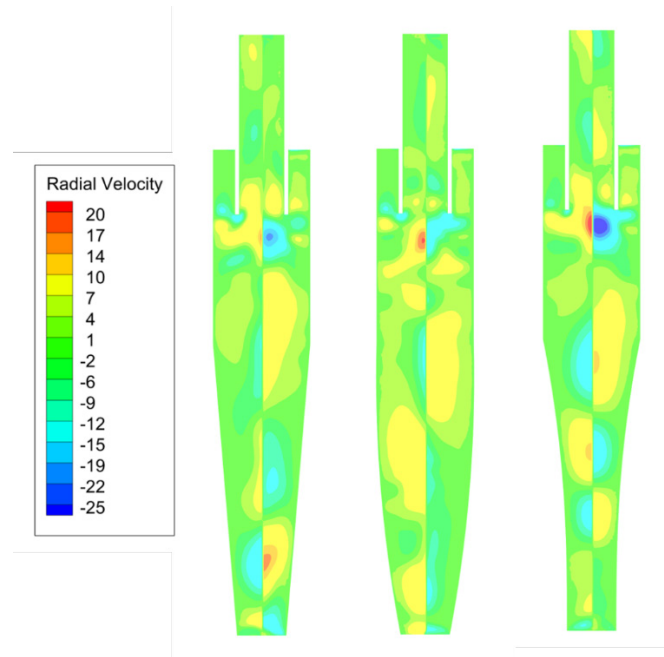


Figure 8. Radial velocity contour plot of cyclone separators.

5. Conclusion

In this work, a systematic set of CFD simulations is conducted to evaluate how three different curved conical sections, straight, concave, and convex, affect the air purification performance of cyclone separators. The data point to geometric curvature as a decisive factor in separation outcomes. The highest PM_{2.5} collection is seen with the convex profile, where the grade efficiency peaks at 82.4%, and the cut-off particle size drops to 2.1 μm , clearly outstripping the other two geometries. The concave profile, in contrast, stands out for its superior pressure drop characteristics and energy efficiency, while the straight profile yields results that lie between these two extremes. The motion of particles is strongly shaped by the flow field structure. Fine particles gain from extended residence time and stronger entrainment inside the convex shape, which makes them much easier to capture. Medium and coarse particles, on the other hand, follow relatively orderly trajectories in both convex and concave geometries so that they can be separated without difficulty. The three geometries also differ markedly in flow stability and energy utilization. The convex profile can keep the flow field reasonably uniform while achieving high collection efficiency, yet it also supports unstable vortices. The concave profile tends to dampen recirculation and shrink the low-pressure core, though this comes at the price of elevated wall pressure.

In general, the convex profile is a better choice for tasks that couple efficient fine-particle capture with high classification uniformity, whereas the concave profile is better suited to applications where low energy consumption and rapid separation take priority. By clarifying the ways in which a curved conical section modifies the flow field and redistributes particles, this study supplies both theoretical understanding and

numerical evidence for the development of indoor air purification devices. Subsequent research can bring in experimental validation and intelligent optimization algorithms to strengthen the model's applicability and promote its transition into engineering practice.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Jiang L, Liu Y, Wang C, et al., 2025, A Numerical Simulation Study on the Separation Efficiency of a Symmetrical Double-Cyclone Separator. *Combined Machine Tools and Automated Machining Technology*, 2025(6): 31–35.
- [2] Zhang H, Guo H, He X, et al., 2025, Simulation Study on Dust Capture Characteristics in Microcyclone Structure Based on CFD-DPM. *Chinese Journal of Work Safety Science and Technology*, 21(3): 113–118.
- [3] Wu J, Lin J, 2024, Simulation Analysis of Mud and Sand Separation Performance of Hydrocyclone Based on Fluent. *Journal of Physics: Conference Series*, 2820: 012101.
- [4] Yang S, Wu K, Li R, et al., 2025, Analysis and Optimization of Industrial Dust Removal Components Based on CFD-DPM and PIV. *Journal of Xihua University (Natural Science Edition)*, 45(X): 1–11.
- [5] Fu P, Wang F, Ma L, et al., 2016, Fine Particle Sorting and Classification in the Cyclonic Centrifugal Field. *Separation and Purification Technology*, 158: 357–366.
- [6] Zhang L, Zhang G, Yang Q, et al., 2018, A Study on the Diffusion of PM_{2.5} in Urban Streets Based on CFD Simulation. *Environmental Horizons*, 2018(3): 66–70.
- [7] Wang Z, Zhang Z, Jiao Q, et al., 2024, Based on the Analysis of the Factors Influencing the Collection Rate of the Dust Removal Device by Fluent. *Machine Design and Research*, 40(5): 75–82.
- [8] Li Q, Li L, Tang G, et al., 2025, A Study on the Influence of Velocity and Particle Size on the Filtration Efficiency of Cyclone Dust Collectors. *Safety and Production*, 7: 84–88.
- [9] Chen Y, 2021, Application of New Type Cyclone Dust Collector in Dust Removal of Exhaust Gas of 1 Mt/a Ethylene Production Plant. *Chemical Industry and Environmental Protection*, 41(2): 241–245.
- [10] Xu J, Zhang L, Yue R, et al., 2015, CFD Simulation on the Aggregation of PM_{2.5} Fine Particles. *Computers and Applied Chemistry*, 32(6): 713–720.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Review of Reactive Power Optimization in Distribution Networks with High Penetration of Renewable Energy

Changshuo Liu*

School of International Education, Changchun University of Technology, Changchun 130012, China

*Corresponding author: Changshuo Liu, 19718968556@139.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The global energy mix is undergoing an accelerated transition towards cleaner, lower-carbon sources. As a result, the large-scale integration of a high proportion of renewable energy into distribution networks is an irreversible trend. However, since the output of renewable energy sources is highly random, intermittent, and fluctuating, and given that the high r/x parameter characteristics of distribution grids result in a pronounced active-reactive power coupling, the traditional unidirectional voltage regulation mode in distribution grids is unable to effectively address issues such as bidirectional power flows and frequent voltage overshoots. Consequently, reactive power and voltage control face significant challenges. To this end, this paper provides a systematic and logically coherent review of the problem of reactive power optimization in distribution networks with a high proportion of renewable energy. First, we clarify the mechanisms by which the grid integration of renewable energy sources affects distribution network voltages. This study then systematically summarizes methods for active power regulation, reactive power compensation, tap-changer voltage regulation, and multi-device coordinated control. Subsequently, this study proposes an approach to multi-timescale, hierarchical, and zoned modeling that coordinates active and reactive power. This study then conducted a comparative analysis of uncertainty management techniques, including stochastic optimization, robust optimization, and fuzzy theory. Finally, this study summarizes the application characteristics of traditional mathematical methods, intelligent algorithms, and data-driven methods. This naturally and appropriately leads to the conclusion that the coordination of generation, grid, load and storage, combined with multi-timescale optimization, represents the current mainstream control framework, and that data-driven methods hold immense potential for application. However, it also objectively and cautiously highlights various issues in existing research, including uncertainties that are difficult to address, multi-scale collaborative complexity, communication and privacy constraints, and a lack of interpretability in AI models. Consequently, the future direction of development should focus on collaboration, intelligence and decentralization, whilst the refinement of relevant technologies will directly and effectively underpin the safe and efficient operation of distribution networks with a high proportion of renewable energy.

Keywords: High proportion of renewable energy; Distribution networks; Reactive power optimization; Voltage control

Online publication: Aug 11, 2026

1. Introduction

As the global energy mix transitions towards cleaner and lower-carbon sources, the new power system has become the key vehicle for achieving sustainable energy development. The new power system is characterized by a high proportion of renewable energy, integrates advanced information technology with power electronics, and emphasizes the coordinated interaction and intelligent control of generation, transmission, consumption, and storage. Microgrids, as key components of the new power system, are small-scale, autonomous power systems comprising distributed generation sources, energy storage devices, energy conversion equipment, loads, and monitoring and protection systems. It features flexible networking and efficient operation, and can switch between stand-alone and grid-connected modes, enabling effective integration of distributed energy resources and enhancing the reliability and flexibility of the power system.

As new energy power generation technologies mature and costs decrease, China's installed capacity of new energy has sustained growth. A large number of distributed new energy sources are integrated into the power grid locally via distribution networks for on-site consumption. However, the original design of distribution networks centers on the concept of "radial topology and unidirectional power supply" and does not fully account for the characteristics of high-penetration renewable energy grid connection: its output power is affected by natural conditions, exhibiting significant intermittency and randomness. Moreover, most distributed new energy inverters operate at a constant power factor, resulting in limited reactive power regulation capability. These factors lead to frequent disruptions of power balance in distribution networks. Accordingly, systematically analyzing the impact of grid-integrated renewable energy on the voltage stability of distribution networks and exploring targeted optimization strategies have become critical topics for promoting high-efficiency integration of renewable energy and ensuring the safe operation of distribution networks ^[1].

Conventional distribution networks are traditionally designed around the fundamental principle of "radial topology with unidirectional power flow," resulting in relatively straightforward reactive power and voltage control strategies. However, this paradigm is increasingly outdated. With the large-scale integration of renewable energy sources, particularly inverter-based distributed generation, the operational requirements of distribution systems have undergone profound changes. Relying solely on reactive power modulation from grid-connected inverters is no longer sufficient to ensure system stability after interconnection. Power flow has transitioned from strictly unidirectional to bidirectional, and the previously monotonic voltage profile has given way to dynamic, spatially heterogeneous voltage variations. Consequently, advanced, coordinated, and adaptive reactive power and voltage control schemes are now essential to maintain reliability, efficiency, and power quality. Not only should it be capable of real-time monitoring and rapid response to fluctuations in the output of new energy, but it should also be able to coordinate numerous controllable resources. That is to say, while ensuring the quality of voltage, it should also take into account economic efficiency and the requirement of "low carbon" ^[2].

In recent years, the scientific community has conducted extensive research and discussion on these issues, yielding abundant theoretical outcomes and technical solutions. This paper aims to review the research progress in this field systematically. This paper then conducts a literature review across three aspects: the influence mechanism of voltage, optimal control mode, and model-solving method.

2. The influence mechanism of new energy grid connection on distribution networks

2.1. Output characteristics and uncertainties of new energy sources

First of all, we need to understand that the fundamental reason the integration of new energy sources affects distribution network voltage lies in their inherent output characteristics and the contradiction with the traditional distribution network architecture.

First comes randomness. Traditional energy sources primarily include thermal and hydropower. After a long period of technological development, they can not only maintain the stability and controllability of power generation but also ensure the continuity of power supply. However, new energy sources cannot do so. The output characteristics of new energy sources make the voltage in distribution networks far less stable than it is with traditional energy sources. This further increases the difficulty of voltage control. Even with short-term weather prediction technology, it is still very difficult to predict extreme weather conditions that could lead to abnormal output situations. For instance, sudden severe cold waves and heavy snowfall in winter can cause photovoltaic output to drop sharply from 70% of the rated value to 5% or even lower within a short period due to snow cover and a sudden reduction in sunlight. At the same time, some distributed wind power may also shut down due to low-temperature icing faults. The distribution network rapidly shifts from a “power surplus” state to a “power deficit” state. Traditional voltage regulation equipment, due to its slow response time, cannot compensate for power fluctuations in time, leading to a continuous drop in voltage and even triggering low-voltage protection devices, causing local power outages. Furthermore, unexpected incidents or the absence of corresponding contingency plans may prevent immediate emergency repairs and remediation, thereby resulting in greater economic losses ^[3].

Secondly, in terms of intermittency and volatility, both wind and photovoltaic power generation exhibit typical characteristics of intermittency and volatility: the output of photovoltaic power generation is entirely dependent on light conditions. Factors such as day-night alternation, cloud cover, and extreme weather can cause significant fluctuations in power generation over a short period, and even drop sharply from full capacity to zero. In contrast, wind power output is directly affected by fluctuations in wind speed. When the wind speed is too low, wind turbines cannot generate electricity. However, when the wind speed reaches grade 10 or higher, wind power generation facilities will automatically shut down to ensure equipment safety. Furthermore, meteorological disturbances such as gusts and turbulence also cause random fluctuations in power output. Both of them are classified as uncontrollable “weather-dependent” energy sources. Power generation is characterized by discontinuous, unstable output, making it difficult to predict accurately even with short-term weather forecasts. This will not only trigger rapid switching of the distribution network between the “power surplus” and “power deficit” states, but also substantially increase the difficulty of regulating and controlling grid voltage and frequency, thereby posing potential risks to the safe and stable operation of the distribution network.

2.2. Voltage out of limit problem

When a high proportion of new energy is connected to the distribution network, due to its inherent output characteristics, it can trigger conflicts with the traditional distribution network structure. Due to the random and intermittent characteristics of new energy output, the power flow direction and amplitude in the distribution network to which it is connected will also fluctuate sharply. There are two situations: when photovoltaic output reaches its peak at noon, a scenario characterized by high penetration of renewable

energy generation and low system load, the voltage rise effect occurs as the generated power flows through the current impedance. If the magnitude of the voltage rise is excessively large, the voltage at the point of interconnection and surrounding grid nodes will exceed the upper limit. This overvoltage issue is prone to occur at the end of feeders or under light load conditions. Under scenarios such as the minimum photovoltaic output at night with no illumination, namely when the output of new energy is insufficient while the load is heavy, relying solely on new energy cannot provide effective support. As a result, the system voltage is highly likely to drop below the lower limit, and this “undervoltage problem” is prone to occur under heavy load or when the output of new energy drops abruptly. Although traditional distribution networks also face these issues, the difference is that, under high proportions of renewable energy, these two voltage problems often alternate, reflecting the volatility of renewable power generation. This undoubtedly poses significant challenges for distribution network control and increases safety risks ^[4].

2.3. Impact of line parameter characteristics on voltage control

The line-parameter characteristics of a distribution network are of great significance to the network itself. In particular, line parameters significantly influence the selection of voltage control strategies. Unlike transmission lines, distribution networks, particularly at medium and low voltages, typically exhibit a relatively high resistance-to-reactance ratio (r/x) that exceeds 0.5. In certain low-voltage distribution feeders, this ratio may approach unity. Consequently, the coupling among active power, reactive power, and node voltage becomes significantly stronger in low-voltage distribution systems with elevated r/x ratios. In high-voltage transmission systems, the inductive reactance (x) significantly exceeds the resistance (r). Consequently, reactive power flow is the primary determinant of voltage magnitude. In contrast, active power flow predominantly governs system frequency. In low-voltage distribution networks, both active and reactive power significantly affect voltage. Therefore, in distribution networks with high penetration of renewable energy, voltage regulation can be realized through both reactive power regulation tools and active power regulation approaches, such as photovoltaic active power curtailment and active power charging/discharging of energy storage systems. This provides more options for developing control strategies but also increases the difficulty of coordinated control.

3. Reactive power control method for distribution networks with high penetration of renewable energy

3.1. Voltage control based on active power regulation

In distribution networks with a high resistance-to-reactance (R/X) ratio, active power has a significant impact on voltage. When voltage overlimit issues occur, adjusting the active power can effectively suppress voltage fluctuations in the distribution network ^[5]. There are two most common methods: the first is photovoltaic active power curtailment based on photovoltaic inverters, which is the most direct method of active power regulation. When the grid connection point voltage exceeds the upper limit, the active power output from photovoltaic inverters can be reduced to mitigate the voltage rise. This method is simple to implement and responds quickly, relying on photovoltaic inverters; however, it has poor economic efficiency, directly resulting in power generation loss and reduced revenue for owners. Therefore, this measure is generally employed as a last resort, and active power reduction is only activated when the reactive power regulation capacity is exhausted. The second approach is active power regulation via energy storage,

which constitutes a more flexible active power control method. The system can charge when voltage is too high and discharge when voltage is too low, achieving peak shaving, valley filling, and voltage regulation through the spatiotemporal transfer of active power. Compared with photovoltaic active power curtailment, energy storage regulation does not result in energy waste. However, the capital investment cost of energy storage systems is relatively high, and the state of charge constrains their regulation capability. In terms of control strategies, active power regulation of energy storage can adopt various modes such as local control, distributed coordination, or multi-stage optimization.

3.2. Voltage control based on reactive power regulation

Reactive power regulation is the primary approach for voltage control in distribution networks with high penetration of renewable energy, as it does not curtail the active power output of renewable energy sources and offers significant economic advantages. The three most common approaches are: the first is reactive power regulation based on photovoltaic inverters, which is currently the most extensively studied reactive power control method ^[6]. Modern photovoltaic inverters generally possess reactive power regulation capability with a power factor of approximately ± 0.8 at their rated capacity. They can utilize their residual capacity to participate in power grid voltage regulation. Depending on the control mode, they can be classified into three types: constant power factor control, constant voltage control (Q(V)), and constant reactive power control. Among existing control strategies, Q(V) control is the most widely adopted in practical engineering due to its simplicity, effectiveness, and communication-free characteristics. This strategy adjusts reactive power output based on the deviation between local voltage measurements and reference values, following a preset Q-V droop curve, thereby enabling autonomous voltage control. The second type is the Static Synchronous Compensator (STATCOM), also known as the Static Var Generator (SVG). It is a dynamic reactive power compensation device that offers advantages such as fast response and a wide regulation range. In distribution networks with high penetration of renewable energy, STATCOM can provide smoothly adjustable reactive power support and effectively suppress rapid voltage fluctuations. However, STATCOMs incur high investment costs and are typically deployed at critical nodes. Thirdly, shunt capacitor banks are conventional reactive power compensation equipment, featuring the advantages of low cost and simple maintenance. However, their regulation is discrete, with slow switching speed and limited operation times, making it difficult to cope with the rapid fluctuations of renewable energy. Therefore, in modern high-proportion new-energy distribution networks, capacitor banks are typically used as slow-response devices and are optimized for configuration on the day-ahead or intra-day time scale.

3.3. Voltage control based on tap-changer equipment

Tap-changing equipment-based voltage regulation is a critical global method for voltage regulation in renewable energy distribution networks. It adjusts the system voltage level by modifying the transformer ratio and is commonly coordinated with distributed voltage and reactive power compensation equipment to maintain voltage stability jointly. There are two main types: the first is the on-load tap changer (OLTC) transformer, which is a commonly used voltage-regulation device in distribution networks. It adjusts the transformer's transformation ratio by changing the tap position, thereby regulating the voltage on the low-voltage side. OLTC features a wide regulation range, yet it suffers from slow response speed and limited switching operations; hence, it is typically employed to address slow variations in load demand and

renewable energy output. In scenarios with high renewable energy penetration, the on-load tap changer (OLTC) must coordinate with fast-reactive power compensation equipment to avoid frequent operations that compromise the equipment's service life. The second type is the Step Voltage Regulator (SVR), which applies to scenarios with long feeders or concentrated loads at the far end of the distribution system. When the grid connection point for new energy is located far from the substation, SVR can perform voltage regulation downstream of the connection point, compensating for the OLTC's limited regulation capability.

3.4. Multi-device collaborative control

Currently, it is difficult for a single device to address the complex voltage issues caused by high-penetration renewable energy, and the development of multi-device coordinated control has become an inevitable trend^[7]. First of all, the latest research hotspot currently is the "source-grid-load-storage" coordinated control. Based on the "four-component" architecture of active distribution networks, this model integrates distributed generators, grid regulation devices, flexible controllable loads, and energy storage systems into a unified dispatch and control framework, thereby fully exploiting the reactive power regulation and voltage control potential of each participant. Through the coordinated operation of various resources, this model achieves stable voltage operation in distribution networks with high proportions of renewable energy penetration^[8]. Among these challenges, collaborative control is particularly difficult due to significant differences in response speed, regulation capacity, operating costs, and control characteristics across various types of equipment. Furthermore, bidirectional power flow fluctuations and voltage violations caused by high-penetration renewable energy integration further increase the difficulty of dispatching and control. Therefore, it is urgent to establish a hierarchical coordination mechanism that adapts to the fluctuation characteristics of renewable energy^[9].

Next, we discuss another approach: multi-time scale coordinated control. This is an effective method for addressing differences in the response characteristics of heterogeneous regulation devices, and it is also the mainstream research direction for reactive power optimization in distribution networks with high penetration of renewable energy^[9]. In the day-ahead optimization stage, based on medium- and long-term forecasts of wind-photovoltaic power output and load demand, the operation schedules of slow-response regulation devices, including OLTC tap positions and capacitor bank switching status, are determined, with both overall operational economy and device service lifetime considered. In the intra-day optimization stage, rolling adjustments are applied to the outputs of partitioned reactive power compensation devices and selected distributed generator inverters via ultra-short-term forecasting to compensate for day-ahead prediction errors. In the real-time control phase, the rapid response advantages of power electronic devices, such as photovoltaic and wind power inverters, are fully utilized to suppress voltage fluctuations induced by instantaneous output variations from renewable energy sources. This hierarchical, graded, and refined collaborative architecture not only achieves economic optimization over long-term time scales but also ensures security and stability at the real-time level and effectively adapts to the complex, variable operational scenarios of distribution networks with high levels of new energy integration.

4. Modeling and solution method for reactive power optimization

4.1. Optimized modeling method

The first approach is multi-time-scale optimization, which serves as the primary modeling framework for

addressing the uncertainty of new energy sources. Day-ahead optimization uses a temporal resolution of 1 hour or 15 minutes and formulates dispatch schedules for slow-response units based on forecast data. Intraday optimization operates on a 5- to 15-minute cycle, correcting plan deviations through rolling forecasts; real-time optimization aims for second-level responsiveness, achieving closed-loop control with fast-response equipment. Its core lies in the coordinated alignment of strategies across all levels to avoid control conflicts among them^[10]. The second approach is hierarchical and zonal optimization, which leverages the characteristics of distribution networks- strong coupling within zones and weak coupling between zones- to decompose the global optimization problem into multiple sub-problems. Partitioning based on electrical distance, spectral clustering, statistical distance, or other methods can reduce computational complexity and apply to large-scale distribution network scenarios^[11]. The third type is coordinated optimization of active and reactive power. In distribution networks with a high r/x ratio, the coupling effect between the two is significant, and it may be difficult to achieve optimal performance by optimizing reactive power alone. When optimization simultaneously considers the collaborative coordination of active power curtailment from new energy sources, charging and discharging of energy storage systems, and reactive power compensation, the unification of voltage control and economic operation can be achieved^[12]. These three reactive power optimization modeling methods address issues associated with the grid integration of new energy sources from different perspectives. Each method has its own strengths, and they complement one another. They can be selected as required to construct a more adaptive optimization modeling system.

4.2. Uncertainty processing method

The randomness of renewable energy generation is a core challenge in reactive power optimization. The first category is stochastic optimization, which relies on the probability distribution of new energy and converts stochastic problems into deterministic problems through scenario generation and reduction techniques. Both single-stage and multi-stage stochastic programming have been investigated and applied in practice. The advantage of this method is that it can fully utilize probability information, but it requires high precision in the distribution model^[13]. The second category is robust optimization, which characterizes the output range of renewable energy using an uncertainty set and seeks a feasible solution to the constraints under the worst-case scenario. This method does not require an exact probability distribution and is highly robust, but its results are generally conservative. Achieving the optimal trade-off between robustness and economy in the characterization of uncertainty sets is the core focus of this research^[13]. The third approach is fuzzy theory, which describes the uncertainty of renewable energy output using membership functions and transforms the optimization problem into a fuzzy programming problem for solution. This method is advantageous in scenarios where information is incomplete or precise probabilistic models are difficult to establish^[14]. These three categories of methods differ in their modeling bases and applicable scenarios. Yet, a single method or a combination of methods can be adopted as needed to implement uncertainty processing.

4.3. Solution algorithm

Traditional mathematical methods include linear programming, nonlinear programming, and mixed-integer programming. Their advantages lie in the fact that they can, in theory, guarantee the identification of the global optimal solution with strong interpretability. Yet, they impose stringent requirements on the model structure and typically require the original problem to be transformed into a solvable standard form. Convex

optimization techniques such as second-order cone relaxation and semidefinite relaxation can handle non-convex-concave problems. This method has been widely applied to distribution network optimization in recent years ^[15]. In contrast, heuristic intelligent algorithms such as the genetic algorithm and particle swarm optimization obtain satisfactory solutions by simulating the intelligent behaviors of biological populations in nature and conducting iterative search, which are suitable for addressing complex nonlinear problems. These algorithms impose low requirements on model forms and feature high flexibility, yet they cannot guarantee stability and optimality, and suffer from low computational efficiency. In recent years, improved intelligent algorithms have been widely adopted to enhance solution performance ^[16]. Furthermore, machine learning and data-driven methodologies have demonstrated tremendous potential in reactive power optimization. Neural networks are deployed for prediction and perception, while reinforcement learning enables rapid, real-time decision-making through offline training combined with online decision-making. Additionally, federated learning enables collaborative optimization among multiple agents without requiring raw data sharing and can effectively address data privacy concerns ^[17]. These three categories of methods each have their own advantages and disadvantages. However, the fusion and improvement of different algorithms represent the future development trend, and only through this approach can the solution performance and scenario adaptability be comprehensively enhanced.

5. Conclusion

Reactive power optimization of distribution networks with high penetration of renewable energy is a key technology for ensuring the secure and economic operation of new-type power systems. This paper systematically summarizes the research progress in this field across three dimensions: the voltage-influence mechanism, control methods, and optimization modeling and solution methods. Analysis indicates that the voltage issues caused by high-penetration renewable energy integration originate from the uncertainty of renewable energy output and the line parameter characteristics of distribution networks; multi-device coordinated control, multi-time-scale optimization, and source-grid-load-storage coordination have become the mainstream control framework; uncertainty processing methods such as stochastic optimization and robust optimization have gradually matured, and data-driven methods represented by machine learning have demonstrated enormous potential. This field now encounters several hurdles, such as managing uncertainties, complex multi-scale coordinated control, communication privacy limitations, and inadequate interpretability of intelligent models, necessitating ongoing enhancement of the technical foundation. In the future, with advancements in collaborative regulation, digital-physical integration, and intelligent algorithms, the safe and efficient operation of high-proportion renewable energy distribution networks will receive solid and reliable support.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Zeng Z, 2025, Research on Power Engineering, Construction Planning, and Grid Stability Amid Large-Scale New

- Energy Access. *Zhangjiang Technology Review*, 9: 71–73.
- [2] Tian Y, Xue Y, Wang Q, et al., 2025, Research on Optimization and Grid-Connected Control of Microgrid Technology in a New Power System. *Intelligent Building & Smart City*, 9: 192–194.
 - [3] Zhang W, 2026, Research on the Impact of New Energy Grid Integration on Distribution Network Voltage Stability and Optimization Strategies. *Intelligent Building & Smart City*, 2: 94–96.
 - [4] Xu J, Tian M, 2025, Practical Research on Active Regulation Technology for Distribution Network Voltage Violation Under High-Proportion New Energy Access. In: *Proceedings of the 14th Academic Symposium on Engineering Technology Management and Digital Transformation*, Volume 1, 535–536.
 - [5] Cai Y, Tang W, Xu O, et al., 2018, Review on Voltage Control of Low-Voltage Distribution Network with High Proportion of Residential Photovoltaic. *Power System Technology*.
 - [6] Zhao Z, Lei Y, He F, et al., 2011, Review on Technologies of Large-Capacity Grid-Connected Photovoltaic Power Stations. *Automation of Electric Power Systems*.
 - [7] Liao J, Yu R, Ye R, et al., 2025, Coordinated Control Method of Low-Voltage Distribution Network with High Proportion of Residential Photovoltaic. *Electric Power*, 58(10): 110–120.
 - [8] Li Q, 2025, Research on Source-Grid-Load-Storage Collaborative Control Method of New Power System. *China High-Tech*, 22: 57–59.
 - [9] Yang L, Ren X, Cai Z, et al., 2024, Review on Voltage Control Strategies of Distribution Networks with High Photovoltaic Penetration. *Power System Technology*, 48(12): 5056–5070.
 - [10] Cong Y, Yuan S, Wang H, et al., 2025, Multi-Time Scale Reactive Power and Voltage Optimal Control of Active Distribution Network Based on Improved Multi-Objective Gray Wolf Algorithm and Second-Order Cone Programming. *Modern Electric Power*, 42(4): 765–777.
 - [11] Chen W, Gan W, Zhang J, et al., 2022, Hierarchical and Zonal Voltage Regulation Strategy for Distribution Network Based on HEM Sensitivity. *Electric Power Construction*, 43(11): 42–52.
 - [12] Ji H, Xu L, Yang J, et al., 2026, Active-Reactive Power Optimization of Active Distribution Network Considering Dynamic Wind-Solar Uncertainty and Demand Response. *Electric Power*, 59(2): 47–60.
 - [13] Li X, Wang W, Wang H, et al., 2020, Robust Optimal Operation Strategy of Cross-Regional Flexibility Considering Bilateral Uncertainty of Load and Source. *High Voltage Engineering*, 46(5): 1538–1549.
 - [14] Wang W, Yang J, Ren G, et al., 2024, Reactive Power and Voltage Zoning Control of Wind Farms Considering Fuzzy Multi-Objective Optimization. *Journal of Power Engineering*, 44(1): 99–108.
 - [15] Xu T, Yao L, Ni L, et al., 2019, A Reactive Power Optimization Algorithm for Distribution Network Based on Second-Order Cone Programming. *Science Technology and Engineering*, 19(10): 111–119.
 - [16] Zhang C, Chen M, Luo C, et al., 2010, Reactive Power Optimization of Power System Based on Multi-Objective Particle Swarm Optimization Algorithm. *Power System Protection and Control*, 38(20): 153–158.
 - [17] Ni S, Cui C, Yang N, et al., 2021, Multi-Time Scale Online Reactive Power Optimization of Distribution Network Based on Deep Reinforcement Learning. *Automation of Electric Power Systems*, 45(10): 77–85.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Beyond Direct AI Assistance: The Mediating Role of AI Crafting in Converting Employee-AI Collaboration into Incremental and Radical Creativity

Xiaolin Zhou*, Dexin Zhao

School of Business and Tourism, Sichuan Agricultural University, Chengdu 611830, China

**Corresponding author: Xiaolin Zhou, zhou_xiaolin163@163.com*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Artificial intelligence is becoming a routine partner in enterprise work, but its creative value depends on how employees reconfigure their work around AI. Using 138 valid survey responses from Chinese employees with AI-use experience, this study examines whether employee-AI collaboration promotes incremental and radical creativity through AI crafting. Results show that employee-AI collaboration significantly predicts AI crafting and both creativity outcomes. AI crafting further mediates these relationships. The findings highlight AI crafting as a behavioral mechanism linking human-AI collaboration with dual creativity.

Keywords: Employee-AI collaboration; AI crafting; Incremental creativity; Radical creativity; Enterprise management

Online publication: Aug 11, 2026

1. Introduction

Artificial intelligence (AI) is increasingly embedded in enterprise work systems, changing how employees search for information, make judgments, and solve problems. Unlike traditional digital tools, AI can participate in decision processes and provide suggestions that reshape employees' task boundaries. Recent studies indicate that employee-AI collaboration is not merely a technical use behavior, but a proactive form of human-AI interaction that can influence work adaptation and career-related outcomes^[1,2].

At the same time, the creativity value of AI is not automatic. Employees need to translate AI-enabled resources into new work methods. AI crafting captures such proactive adjustment: employees explore, learn, and redesign work processes when collaborating with AI. Prior research has shown that AI-related conditions can encourage crafting behavior and innovation-oriented outcomes^[3-5]. However, fewer studies have directly examined whether AI crafting explains how employee-AI collaboration contributes to different types of em-

ployee creativity.

Employee creativity includes both incremental creativity, which improves existing ideas and procedures, and radical creativity, which produces more original and breakthrough approaches ^[6]. Based on this logic, this study examines employee-AI collaboration as an antecedent of AI crafting and further tests whether AI crafting promotes both incremental and radical creativity. The study responds to recent calls to explain the behavioral mechanism linking AI use with dual creativity in Chinese organizational contexts ^[7].

2. Literature review and hypotheses

2.1. Employee-AI collaboration and AI crafting

Employee-AI collaboration refers to the extent to which AI participates in employees' decision-making, prediction, problem-solving, information evaluation, and risk identification processes. Kong et al. argued that employee-AI collaboration represents a proactive behavior through which employees fit into AI-integrated environments ^[1]. In enterprise work, such collaboration provides employees with additional information, cognitive support, and task feedback, which may create conditions for employees to redesign how they work with AI.

AI crafting emphasizes employees' active efforts to explore AI tools, adjust inefficient routines, learn AI-related skills, and optimize human-AI interaction. When employees collaborate more frequently with AI, they are more likely to recognize where AI can improve task processes and where human judgment still matters. Liang et al. also show that employee-AI collaboration can shape crafting-related behavior through employees' interpretation of AI uncertainty ^[2]. Therefore, closer employee-AI collaboration is expected to encourage AI crafting.

H1: Employee-AI collaboration is positively related to AI crafting.

2.2. AI crafting and dual creativity

AI crafting may promote employee creativity because it transforms AI resources into concrete work improvements. Through AI crafting, employees can search for alternative solutions, compare outputs, modify existing procedures, and improve the fit between AI suggestions and organizational tasks. Li et al. and Liu et al. suggest that AI-related crafting is associated with positive employee outcomes, while Xu et al. further link AI crafting with innovation behavior ^[3-5]. These studies provide a behavioral explanation for why AI use may contribute to creative work.

Incremental and radical creativity may both benefit from AI crafting, but in different ways. AI crafting can support incremental creativity by helping employees improve existing ideas, adapt current processes, and make small-scale refinements. It can also support radical creativity by expanding employees' information range, stimulating unusual combinations and lowering the cost of experimenting with new ideas. Zhang et al. (2025) found that generative AI use can affect both incremental and radical creativity through learning mechanisms.

H2a: AI crafting is positively related to incremental creativity.

H2b: AI crafting is positively related to radical creativity.

2.3. The mediating role of AI crafting

Employee-AI collaboration may influence creativity not only directly but also through AI crafting. Collaboration with AI provides technological resources, but creativity requires employees to reorganize those resources into work processes actively. AI crafting functions as the behavioral bridge between "working with

AI” and “creating with AI”. When employees integrate AI into their task routines, they are more likely to identify possibilities for improving existing practices and proposing novel work methods.

This mediating logic is consistent with recent AI management research. Liang et al. emphasize the double-edged nature of employee-AI collaboration, suggesting that employee responses determine whether AI collaboration produces positive outcomes^[8]. Xu et al. highlight AI crafting as a mechanism linking responsible AI conditions with service innovation^[5]. Therefore, this study proposes that AI crafting mediates the relationship between employee-AI collaboration and both creativity outcomes.

H3a: AI crafting mediates the relationship between employee-AI collaboration and incremental creativity.

H3b: AI crafting mediates the relationship between employee-AI collaboration and radical creativity.

3. Research methodology

3.1. Variable measurement

The questionnaire used mature scales and adapted them to the employee AI-use context. Employee-AI collaboration was measured using five items developed by Kong et al.^[11]. AI crafting was measured using six items adapted from Li et al.^[3]. Incremental creativity and radical creativity were measured using three items for each dimension based on Madjar et al.^[6]. All items were rated on a seven-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree.

3.2. Data collection and sample

Data were collected through an online questionnaire platform. A screening question was used to ensure that respondents had used AI tools in work or organizational contexts. Two newly collected datasets were merged, and after removing repeated header records, 138 valid responses were retained. The gender distribution was balanced, with 70 male and 68 female respondents. Most respondents were aged between 21 and 40. Private-enterprise employees formed the largest group, followed by state-owned enterprise employees, public institution employees, and other occupational groups. Most respondents held undergraduate or graduate degrees.

4. Data analysis

4.1. Reliability and validity analysis

Before structural testing, reliability, convergent validity, collinearity, and common method bias were examined following Kock et al.^[9]. Cronbach’s alpha values ranged from 0.621 to 0.800, and composite reliability values ranged from 0.767 to 0.883, supporting acceptable measurement reliability for exploratory enterprise management research. The first unrotated factor explained 36.66% of total variance, and all VIF values were below 1.45, suggesting that common method bias and collinearity did not dominate the results.

4.2. Descriptive statistics and correlation analysis

Table 1 reports the means, standard deviations, and Pearson correlation coefficients. Employee-AI collaboration was positively correlated with AI crafting, incremental creativity, and radical creativity. AI crafting was also strongly correlated with both creativity outcomes. These preliminary results provide initial support for the proposed relationships. See **Table 1**.

Table 1. Means, standard deviations, and correlations

Variable	Mean	SD	1	2	3	4
1. Employee-AI collaboration	5.764	0.571	1			
2. AI crafting	5.942	0.598	0.514***	1		
3. Incremental creativity	5.710	0.734	0.496***	0.772***	1	
4. Radical creativity	5.449	0.932	0.505***	0.676***	0.763***	1

Note. *** $p < 0.001$.

4.3. Structural model and mediation analysis

Hierarchical regression and a 5,000-sample bootstrap procedure were used to test the hypotheses, with gender, age, and education controlled. As shown in **Table 2**, employee-AI collaboration had a significant positive effect on AI crafting, supporting H1. Employee-AI collaboration also showed significant total effects on incremental and radical creativity. When AI crafting was added to the models, AI crafting significantly predicted both incremental creativity and radical creativity, supporting H2a and H2b.

Table 2. Structural path and mediation test results

Path	Effect	p value / 95% CI	R ²	Decision
Employee-AI collaboration → AI crafting	0.530	< 0.001	0.308	H1 supported
Employee-AI collaboration → Incremental creativity	0.511	< 0.001	0.284	total effect
Employee-AI collaboration → Radical creativity	0.520	< 0.001	0.296	total effect
AI crafting → Incremental creativity	0.691	< 0.001	0.615	H2a supported
Employee-AI collaboration → Incremental creativity	0.144	0.026	0.615	partial direct effect
AI crafting → Radical creativity	0.547	< 0.001	0.503	H2b supported
Employee-AI collaboration → Radical creativity	0.230	0.002	0.503	partial direct effect
Employee-AI collaboration → AI crafting → Incremental creativity	0.366	[0.235, 0.500]	-	H3a supported
Employee-AI collaboration → AI crafting → Radical creativity	0.290	[0.172, 0.412]	-	H3b supported

Note: Standardized coefficients are reported. Indirect effects are based on 5,000 bootstrap samples. Gender, age, and education were controlled.

Therefore, H3a and H3b were supported. AI crafting serves as a shared behavioral mechanism through which employee-AI collaboration contributes to both types of creativity.

5. Conclusion, discussion, and managerial implications

5.1. Conclusion and discussion

This study examined how employee-AI collaboration influences AI crafting and dual creativity in AI-transformed enterprises. The results show that employee-AI collaboration is positively associated with AI crafting and with both incremental and radical creativity. More importantly, AI crafting mediates the effects of employee-AI collaboration on the two creativity outcomes. This means that AI does not enhance creativity simply because it is available. Employees must actively explore AI functions, adjust routines, and learn how to interact with AI tools before collaboration can be converted into creative performance. The findings extend employee-AI collaboration research by identifying AI crafting as a behavioral mechanism and enrich creativity research by showing that AI-enabled work redesign supports both refinement-oriented and

breakthrough-oriented creativity.

5.2. Managerial implications and limitations

Enterprises should not only provide AI tools but also encourage employees to redesign tasks, share AI-use experience, and experiment with human-AI interaction. Managers can create AI practice communities, provide prompt-design training, and reward employees who use AI to improve existing processes or develop novel solutions.

Several limitations should be noted. The study used self-report data, which may not fully capture supervisor-rated creativity or actual AI-use behavior. Future research could adopt longitudinal or multi-source designs, compare industries, and examine whether AI crafting has different effects across task complexity, AI tool types, and stages of digital transformation.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Kong H, Yin Z, Baruch Y, et al., 2023, The Impact of Trust in AI on Career Sustainability: The Role of Employee-AI Collaboration and Protean Career Orientation. *Journal of Vocational Behavior*, 146: 103928.
- [2] Yin Z, Kong H, Baruch Y, et al., 2024, Interactive Effects of AI Awareness and Change-Oriented Leadership on Employee-AI Collaboration: The Role of Approach and Avoidance Motivation. *Tourism Management*, 105: 104966.
- [3] Li W, Qin X, Yam K, et al., 2024, Embracing Artificial Intelligence (AI) With Job Crafting: Exploring Trickle-Down Effect and Employee Outcomes. *Tourism Management*, 104: 104935.
- [4] Liu Y, Sheng F, Liu R, 2025, Generative AI Adoption and Employee Outcomes: A Conservation of Resources Perspective on Job Crafting, Career Commitment, and the Moderating Role of Liking of AI. *Humanities and Social Sciences Communications*, 12(1): 1–17.
- [5] Xu Y, Xie P, Naeem R, et al., 2026, Responsible AI and Employee Service Innovation Behavior: A Sequential Mediation Model of AI Self-Efficacy and AI Crafting. *Technological Forecasting and Social Change*, 224: 124470.
- [6] Madjar N, Greenberg E, Chen Z, 2011, Factors for Radical Creativity, Incremental Creativity, and Routine, Noncreative Performance. *Journal of Applied Psychology*, 96(4): 730–743.
- [7] Zhang X, Yu P, Ma L, 2025, How and When Generative AI Use Affects Employee Incremental and Radical Creativity: An Empirical Study in China. *European Journal of Innovation Management*, 29(3): 702–727.
- [8] Liang Y, Yang M, Wu T, 2025, Double-Edged Sword Effect of Employee-AI Collaboration: The Role of AI Uncertainty and Digital Leadership. *Leadership & Organization Development Journal*, 46(6): 943–962.
- [9] Kock F, Berbekova A, Assaf A, 2021, Understanding and Managing the Threat of Common Method Bias: Detection, Prevention and Control. *Tourism Management*, 86: 104330.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Optimization of Artificial Immune Algorithms and Research on Key Issues

Lu Hong, Su Li

School of Computer Science and Engineering, Hunan Institute of Technology, Hengyang 421002, Hunan, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The artificial immune algorithm is an intelligent optimization algorithm derived from the working mechanism of biological immune systems. It simulates immune response processes including biological antigen recognition, antibody generation, clonal selection, immune memory, and immune suppression, and searches for optimal solutions through iterative population evolution. In recent years, artificial immune algorithms have been gradually applied in various fields such as industrial optimization, artificial intelligence, smart cities, and automatic control, showing favorable application value in scenarios including multi-objective constrained optimization, dynamic system scheduling, fault feature identification, and adaptive parameter tuning. However, with the continuous increase in the complexity of application scenarios, the inherent defects of classic artificial immune algorithms have become prominent. On this basis, based on the basic framework of artificial immune algorithms, this paper comprehensively analyzes the core key problems in algorithm operation and constructs a multi-dimensional and highly adaptive algorithm optimization system, aiming to provide comprehensive support for iterative optimization and scenario-based implementation of artificial immune algorithms.

Keywords: Artificial immune algorithm; Population diversity; Affinity evaluation; Immune operator; Algorithm optimization

Online publication: Aug 11, 2026

1. Basic principles of artificial immune algorithms

The artificial immune algorithm is an intelligent heuristic optimization algorithm abstractly constructed based on basic biological immune mechanisms such as antigen recognition, clonal selection, antibody mutation, and immune memory. Its core lies in the deep integration of biological immune mechanisms and engineering optimization problems to continuously search for optimal solutions. In the algorithm operation system, the optimization problem to be solved and its constraints correspond to antigens in biological immune systems; all feasible solutions within the solution space are regarded as antibodies; the matching degree between antibodies and antigens is defined as affinity, which serves as the primary criterion for judging solution quality. The complete operation of the algorithm consists of four major procedures: first, antibody population

initialization, which randomly generates an initial antibody set in the solution space to form a basic feasible solution population; second, affinity calculation, which scores the matching degree of each antibody with the antigen to screen out high-quality individuals; third, immune iteration, which imitates immune behaviors including cloning, mutation, selection and suppression to retain high-affinity antibodies and eliminate low-affinity ones; fourth, immune memory update, which stores the optimal feasible solutions to accelerate the search speed of subsequent iterations^[1]. Equipped with global exploration and local exploitation capabilities, this algorithm can avoid falling into local optima and is suitable for solving multi-constrained, nonlinear, and high-dimensional optimization problems, making it an important algorithm model in the field of intelligent optimization.

2. Key problems existing in artificial immune algorithms

2.1. Unbalanced population diversity and severe premature convergence

Population diversity is the fundamental guarantee for artificial immune algorithms to maintain global search capability and avoid local optima. At present, traditional artificial immune algorithms generally suffer from dynamic imbalance of population diversity, which easily leads to severe premature convergence. In the early iteration stage, initial antibodies are randomly generated with sufficient population diversity. As iterations proceed, high-affinity antibodies are frequently cloned and retained, while inferior antibodies are continuously eliminated, resulting in increasingly similar individuals within the population and homogenized antibody coding patterns. Lacking an effective mechanism to maintain population diversity, the algorithm cannot sufficiently remove duplicate and similar antibodies, sharply narrowing the population search space in the later iteration stage. Consequently, the algorithm rapidly converges to local optimal solutions and fails to explore better solutions in the solution space, especially in complex optimization problems with high dimensions, multiple peaks, and strong constraints^[2,3]. Meanwhile, the fixed antibody update mechanism of traditional algorithms generates few new individuals with minor differences, unable to compensate for the loss of population diversity. This causes the loss of global search capability and a sharp decline in optimization accuracy, failing to meet the requirements of high-precision engineering optimization.

2.2. Single affinity evaluation mechanism with insufficient adaptability

Affinity evaluation is the primary criterion for artificial immune algorithms to select excellent antibodies and determine iteration directions. Traditional artificial immune algorithms adopt fixed and low-dimensional affinity evaluation methods with poor adaptability to various scenarios. Most existing algorithms only use a single objective function to calculate antibody affinity, taking numerical matching degree as the sole evaluation standard, while ignoring the universal characteristics of multi-constraint, multi-objective, and fuzziness in engineering optimization problems. Such an evaluation model can basically satisfy the requirements of simple single-objective optimization problems, yet when dealing with engineering scenarios featuring multi-objective optimization, complex constraints, and dynamically changing objective weights, affinity indicators cannot accurately reflect the comprehensive merits of antibodies. Moreover, traditional evaluation mechanisms fail to distinguish the global adaptability and local adaptability of antibodies, unable to balance the optimality and stability of solutions, which easily results in solutions with single-dimensional optimality but poor overall performance. In addition, fixed affinity calculation models cannot adaptively adjust evaluation weights according to iteration stages and scenario characteristics. In the early iteration

stage, individuals with partial local advantages tend to be selected, while in the later stage, it is difficult to accurately locate the optimal solution, directly leading to biased iteration guidance and significantly reduced reliability of optimization results.

2.3. Solidified immune operators with weak adaptive optimization capability

Immune operators are the core operational units for artificial immune algorithms to iteratively optimize antibodies, mainly including cloning, mutation, selection, and immune suppression operators. In traditional algorithms, the parameters and operation rules of immune operators are strictly fixed, lacking adaptive optimization capabilities. All operators of traditional artificial immune algorithms adopt fixed parameters; core parameters such as cloning multiple, mutation probability, and selection threshold remain unchanged throughout the iteration cycle, failing to adapt to the optimization demands of each iteration stage^[4]. A fixed low mutation probability in the initial stage, requiring wide-range solution space exploration, results in slow population renewal and low exploration efficiency; a fixed high cloning multiple in the later stage, requiring refined local search, leads to population redundancy and waste of computing resources. Meanwhile, the operation logic of operators lacks dynamic adjustment mechanisms: cloning operations apply identical rules to high-quality and inferior antibodies without differentiated screening; mutation operations are overly random without directional optimization capacity; fixed immune suppression thresholds cannot dynamically balance population renewal and retention of superior individuals. The solidified operator mechanism prevents the algorithm from adaptively adjusting search strategies according to scenario complexity and iteration progress, making it difficult to simultaneously guarantee convergence speed and optimization accuracy, and restricting its adaptability to complex scenarios.

3. Optimization strategies for artificial immune algorithms and engineering case applications

3.1. Dynamic regulation optimization strategy for population diversity

To address the problems of insufficient population diversity and premature convergence in traditional artificial immune algorithms, a dynamic regulation optimization strategy for population diversity is proposed, whose core is to construct a closed-loop regulation system featuring real-time monitoring of population diversity, differentiated renewal, and redundant individual elimination.

First, two indicators, including information entropy and Hamming distance, are adopted to measure the numerical value of population diversity in each iteration in real time and accurately judge the homogenization degree of the population. Compared with a single traditional judgment indicator, this method can comprehensively reflect differences in antibody coding and solution space distribution. Second, dynamic diversity thresholds are set and adjusted according to iteration progress: loose thresholds are adopted in the initial stage to ensure rapid population expansion and extensive exploration of the solution space; tighter thresholds are applied in the middle stage to eliminate highly similar antibodies as early as possible and prevent excessive population homogenization; refined retention of differentiated superior individuals is realized in the later stage to maintain diversity in local search. Third, a differentiated antibody update mechanism is designed. For populations with insufficient diversity, a chaotic perturbation algorithm is used to generate new antibody individuals to increase population diversity and avoid iteration stagnation; for populations with sufficient diversity, only a small number of inferior individuals are updated to retain high-

quality iteration results. Finally, an immune redundancy suppression mechanism is added to the immune system to forcibly eliminate antibodies with repetition rates exceeding the threshold and avoid homogenized individuals. This strategy achieves dynamic balance of population diversity, guarantees the global search capability of the algorithm throughout iterations, and fundamentally solves the problem of premature convergence. In the engineering case of path optimization for intelligent manufacturing workshops, the optimized artificial immune algorithm reduces the population homogenization rate by 42% and the probability of premature convergence, and improves the search accuracy of the optimal path solution by 35%. It well adapts to the complex workshop optimization scenarios with multiple obstacles and alternative paths, and solves the problems of traditional algorithms such as easy trapping in local optima and redundant path planning ^[5,6].

3.2. Multi-dimensional fusion affinity evaluation optimization strategy

A comprehensive evaluation system balancing optimality, stability, and constraint adaptability is constructed to accurately guide algorithm iteration directions. Abandoning the traditional single-objective function evaluation method, a multi-indicator fusion evaluation model is established, covering three major evaluation dimensions: objective fitness, constraint satisfaction, and population differentiation. Objective fitness inherits the core evaluation logic of traditional algorithms to measure the matching degree between antibodies and optimization objectives and guarantee the basic searching capability of the algorithm. Constraint satisfaction focuses on the multi-constraint characteristics of engineering scenarios to evaluate the compliance of antibodies with various boundary, performance, and working condition constraints and eliminate invalid solutions violating constraints. Population differentiation examines the dissimilarity between an antibody and other individuals in the population to select antibodies with both excellent performance and unique characteristics, facilitating the maintenance of population diversity.

On this basis, dynamic weight allocation is adopted to adaptively adjust the weight of each dimension according to iteration stages and scenario demands. In the initial iteration stage, constraint satisfaction and differentiation are prioritized to rapidly locate superior individuals and expand the search scope; in the later stage, objective fitness is taken as the primary indicator to realize refined optimization of solutions and improve solution quality ^[7].

In addition, a fuzzy evaluation correction module is added. For fuzzy and uncertain constraint conditions in engineering scenarios, fuzzy matrices are used to correct affinity values and prevent the erroneous elimination of high-quality individuals caused by rigid evaluation.

In the engineering case of parameter optimization for power system load forecasting, the multi-dimensional evaluation strategy can accurately adapt to the nonlinear, multi-constrained, and dynamically changing characteristics of power loads, effectively avoiding poor parameter adaptability and large forecasting errors resulting from one-sided evaluation of traditional algorithms. The comprehensive adaptation rate of algorithm parameter optimization is increased by 38%, and the load forecasting accuracy is significantly improved, fully meeting the parameter optimization requirements for stable operation of power systems.

3.3. Dynamic optimization strategy for adaptive immune operators

Aiming at the fixed and poorly adaptive characteristics of traditional immune operators, a dynamic optimization strategy for adaptive immune operators is designed, whose core is to build a dynamic

regulation system integrating iteration state perception, adaptive adjustment of operator parameters, and iterative feedback correction. Three parameters, including iteration number, population diversity, and average affinity, are used to perceive the current iteration state of the algorithm in real time and accurately judge whether the algorithm is in the global search, transition search, or local precise search stage. Core operator parameters are dynamically adjusted according to iteration states: in the global search stage, the mutation probability is increased, the cloning range is expanded and the selection threshold is lowered to accelerate population renewal and broaden the exploration scope of the solution space; in the local precise search stage, the mutation probability is reduced, the cloning of high-affinity antibodies is concentrated and the selection threshold is raised to finely polish the optimal solutions^[8]. Meanwhile, the operation logic of operators is optimized, and a differentiated cloning mechanism is established to assign different cloning multiples based on antibody affinity: high-affinity antibodies are cloned multiple times for in-depth exploitation of local optima, while low-affinity antibodies are cloned fewer times to reserve exploration space. A directional mutation mechanism is developed to replace traditional random mutation, and mutation directions are dynamically adjusted according to iteration deviations to improve mutation effectiveness. In addition, the immune suppression operator is optimized with dynamically adjustable suppression thresholds: loose suppression strategies are adopted in the early iteration stage to retain diverse solutions, while strict suppression strategies are applied in the later stage to purify high-quality solutions. In the engineering case of lightweight optimization for mechanical structures, this adaptive operator optimization strategy effectively solves the problems of fixed parameters and low search efficiency of traditional algorithms. The algorithm convergence speed is increased by 40%, and the lightweight optimization accuracy of structures is improved by 29%. On the premise of satisfying strength and stiffness constraints of mechanical structures, the optimal weight of structures is achieved, meeting the high-precision and multi-constrained optimization demands of mechanical engineering.

3.4. Iterative optimization strategy for immune memory mechanism

Traditional artificial immune algorithms adopt simple immune memory mechanisms that only store optimal antibodies, resulting in low utilization of memory individuals, poor iteration inheritance, and weak adaptability to complex scenarios. Therefore, an iterative optimization strategy for immune memory mechanisms is proposed to build a new-generation immune memory system with layered memory, dynamic update, and memory reuse, improving the iteration efficiency and stability of the algorithm. First, the traditional single memory bank mode is broken through, and a layered immune memory system is constructed, dividing the memory bank into an optimal memory layer, a high-quality alternative layer, and a feature memory layer. The optimal memory layer stores the global optimal antibodies of each iteration to retain core optimal solutions; the high-quality alternative layer stores differentiated suboptimal antibodies with top affinity to avoid iteration regression caused by elimination of optimal solutions; the feature memory layer collects coding features and adaptation rules of high-quality antibodies to preserve optimization experience of scenarios. Second, a dynamic memory update mechanism is designed to replace the traditional fixed replacement mode. Update rules are determined according to antibody affinity, memory retention time, and population diversity requirements. Aging and inefficient memory individuals are eliminated regularly, and new high-quality memory individuals are supplemented to maintain the activity of the memory bank^[9,10]. A memory individual reuse mechanism is added: data from the feature memory layer is used in the early

iteration stage to rapidly adapt to scenario features and shorten the initial search cycle; individuals from the high-quality alternative layer are adopted in the middle iteration stage to supplement population diversity; the optimal memory layer is relied on for precise convergence in the later iteration stage. Finally, a memory crossover optimization mechanism is introduced to combine the advantages of memory individuals and new individuals, balancing iteration stability and innovation. In the engineering case of intelligent traffic signal timing optimization, this optimization strategy effectively solves the drawbacks of traditional algorithms such as strong iteration repetition, slow convergence, and poor adaptability of timing schemes. The algorithm iteration cycle is shortened by 35%, the rationality of traffic signal timing is significantly improved, and the intersection traffic efficiency is increased by 27%. The strategy can adapt to traffic conditions with different time periods and traffic flows, showing strong engineering adaptability.

4. Conclusion

In summary, current artificial immune algorithms still have shortcomings, including insufficient adaptability to high-dimensional complex problems, limited robustness in dynamic scenarios, and unsatisfactory solution accuracy for multi-objective optimization. Future research can focus on directions such as multi-algorithm fusion optimization, dynamic adaptive parameter regulation, and mining of novel immune mechanisms, further improving the theoretical system of algorithms, expanding their application boundaries in complex practical scenarios such as industrial scheduling, intelligent optimization and engineering decision-making, and promoting the development of artificial immune algorithms toward high efficiency, intelligence and generalization.

Funding

Scientific Research Project of Hunan Provincial Department of Education (Project No.: 22A0633)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Liu Y, 2025, Intrusion Detection Method of Abnormal Data Flow in Computer Network Based on Improved Artificial Immune Algorithm. *Information Technology and Informatization*, 2025(1): 37–40.
- [2] Zhang K, 2024, Research on Network Security Risk Detection System Based on Artificial Immune Algorithm. *Computer Programming Skills & Maintenance*, 2024(9): 170–172.
- [3] Wang M, Wei Y, Liu Y, 2023, Fault Type Identification Method for Coal Mine Power Supply Lines Based on Improved AIA-BP Algorithm. *Mechanical & Electrical Engineering Technology*, 52(11): 246–250.
- [4] Ai Z, 2023, Quantitative Research and Application of Damage Degree of Power System Network Attacks Based on Improved Artificial Immune Algorithm. *Guangxi Electric Power*, 46(5): 53–61.
- [5] Tian J, Liu X, 2023, Research on Financial Security Risk Perception Monitoring under Intelligent Optimization Algorithms—Based on Regional Social Network Data in China During COVID-19. *Journal of Econometrics*, 3(2):

570–588.

- [6] Wang W, 2023, Optimization of Feeder Automation in Coal Mine Power System with Improved Artificial Immune Algorithm. *Modern Mining*, 39(2): 241–243 + 258.
- [7] Chen L, Zhou R, Su W, et al., 2023, Optimal Design of Liquid Crystal Tunable Filter Based on Artificial Immune Algorithm. *Liquid Crystals and Displays*, 38(2): 188–196.
- [8] Liu D, 2022, Research on Emergency Logistics Distribution Path under the Distribution Mode of “Delivery Vehicle + UAV”. *Science & Technology Entrepreneurship Monthly*, 35(11): 95–99.
- [9] Li X, 2022, Design of Network Security Risk Detection System Based on Artificial Immunity. *Information and Computer (Theoretical Edition)*, 34(20): 51–53.
- [10] Zhang C, 2022, Research on Cooperative Control Method Based on Artificial Immune Algorithm and Ant Colony Algorithm. *Journal of Henan University of Engineering (Natural Science Edition)*, 34(3): 69–73.

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Design of a High-Voltage Constant-Voltage/ Constant-Current Power Supply Based on a PSFB Topology and an Automatic Frequency- Conversion Strategy

Yuepu Sun

WLSA Shanghai Academy, Shanghai 201901, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To meet the requirements for wide-range output, current-limit protection, and stable closed-loop control in high-voltage capacitor charging, laboratory high-voltage sources, and pulse-power preamplifier systems, this paper presents a high-voltage constant-voltage/constant-current power supply based on a phase-shifted full-bridge (PSFB) topology. The supply operates from a typical 24 V low-voltage DC input and uses a 1:100 high-frequency transformer for isolated step-up conversion. Its output control range is 0–2200 V and 0–200 mA, with a rated power limit of 400 W. The controller uses an STM32G474CBT6, which generates four complementary phase-shifted gate-drive signals through the HRTIM module and samples secondary-side voltage, secondary-side current, primary-side current, auxiliary-supply status, and temperature feedback with an ADC synchronized to the switching cycle. The control algorithm calculates the constant-voltage, constant-current, and constant-power PI loops in parallel and selects the actual phase-shift duty cycle through minimum-duty-cycle arbitration. To reduce the efficiency fluctuation of fixed-frequency PSFB operation at very low and very high duty cycles, an automatic frequency-conversion strategy is introduced. Frequency increase or reduction within 11–45 kHz is selected according to the duty-cycle state, while score-counter hysteresis, a lockout window, and feedforward duty-cycle correction reduce the disturbance caused by frequency changes. Simulation and experimental results show that the prototype provides stable high-voltage output, constant-current charging, constant-voltage load operation, and ZVS turn-on, confirming its practical engineering value.

Keywords: PSFB; Phase-shifted full-bridge; High-voltage power supply; Constant-voltage/constant-current control; Automatic frequency conversion; ZVS

Online publication: Aug 11, 2026

1. Introduction

High-voltage isolated DC-DC power supplies are widely used in capacitor charging, insulation testing, laboratory high-voltage sources, and pulse-power devices. These applications require a wide output-voltage

range, clear current-limiting capability, and reliable protection during sudden load changes, output-capacitor charging, and abnormal operating conditions. Because of its strong isolation capability, moderate power rating, mature control method, and suitability for soft switching, the PSFB still has considerable value in medium- and high-power DC-DC converters. Research has shown that reconfigurable PSFB converters can extend the wide-voltage output range and are suitable for applications such as electric-vehicle charging^[1]. However, conventional PSFB converters still suffer from an insufficient ZVS range, duty-cycle loss, circulating-current loss, and reduced synchronous-rectification efficiency under light-load and wide-load conditions. Researchers have addressed these issues through variable-inductor-assisted ZVS, light-load synchronous rectification, and improved ZVZCS structures^[2–4].

In addition to topology, digital control strategies affect the efficiency and dynamic response of the PSFB. The switching frequency and phase-shift duty cycle jointly determine the energy-transfer window, switching losses, and operating state of the magnetic devices. Recent studies have used combined frequency-conversion and phase-shift control to optimize PSFB efficiency, while feedforward compensation has been applied to improve the dynamic response under peak-current-mode control^[5,6]. Model predictive control and predictive deadbeat average control have also been used to improve the transient performance of PSFB and PSFB-CT converters^[7,8]. Research on bidirectional PSFB control indicates that circulating current and reverse power must be actively suppressed through the modulation strategy^[9]. These studies show that a high-voltage PSFB power supply cannot rely only on a fixed switching frequency and a single PI loop; instead, drive timing, synchronous sampling, mode switching, and frequency management must be considered together.

This study focuses on a high-voltage constant-voltage/constant-current power-supply application and designs a digitally controlled PSFB high-voltage supply based on the STM32G474CBT6. The main contributions are as follows: an isolated boost power stage is constructed with a 24 V input and an output of up to 2200 V; four complementary phase-shifted gate-drive signals are implemented through HRTIM; periodic synchronous sampling and multi-loop PI arbitration are used to achieve constant-voltage, constant-current, and power-limit control; an automatic frequency-conversion strategy is proposed to adjust the switching frequency according to the duty-cycle state; and the design is verified through simulation waveforms, efficiency distribution, strategy partitioning, ZVS waveforms, startup waveforms, and CV/CC load experiments.

2. Overall system design

The system provides an adjustable high-voltage DC output from a low-voltage DC input and supports configurable constant-voltage, constant-current, and power-limit operation over the full output range. The recommended input voltage is 24 V, with an allowable range of 18–28 V. The output voltage range is 0–2200 V, the maximum output current is 200 mA, the rated power limit is 400 W, and the switching frequency is controlled automatically between 11 and 45 kHz, with 35 kHz used as the reference frequency. The key design specifications are listed in **Table 1**. In addition to the rated specifications, maintainability during commissioning and application was considered. Voltage, current, and power settings can be configured through the communication interface, while protection thresholds and correction coefficients can be stored permanently to reproduce experimental conditions under different loads and test scenarios.

Table 1. Main electrical parameters of the system

Parameter	Numerical value or range	Description
Input voltage	18–28 V, 24 V typical	Low-voltage DC input
Output voltage	0–2200 V	Adjustable high-voltage DC output
Output current	0–200 mA	Constant-current limiting range
Output power	0–400 W	Constant-power limiting range
Switching frequency	11–45 kHz	Automatic frequency conversion range
Reference frequency	35 kHz	Default running frequency
Dead time	Approximately 200 ns	HRTIM complementary-output dead time
Main control chip	STM32G474CBT6	HRTIM and multiple ADC sampling
Transformer turns ratio	1:100	Isolation boost

The main power supply consists of a low-voltage input, a PSFB full bridge, a high-frequency transformer, a high-voltage rectifier and filter, and an output terminal (**Figure 1**). It isolates the low-voltage, high-current side from the high-voltage, low-current side, improving user and equipment safety while enabling independent feedback and discharge protection on the high-voltage side. The two bridge arms of the full bridge operate complementarily, and the effective energy-transfer time is adjusted through the phase shift between the leading and lagging bridge arms. The hardware control and protection structure is shown in **Figure 2**. The MCU performs drive generation, sampling, protection, and communication. The output voltage VSEC, secondary current ISEC, input voltage VPRI, primary DC IPRI_DC, primary AC IPRI_AC, auxiliary power supply, and temperature signals are collected. Hardware overcurrent protection uses a comparator and an HRTIM Fault input, allowing the drive to be shut down quickly outside the software control cycle.

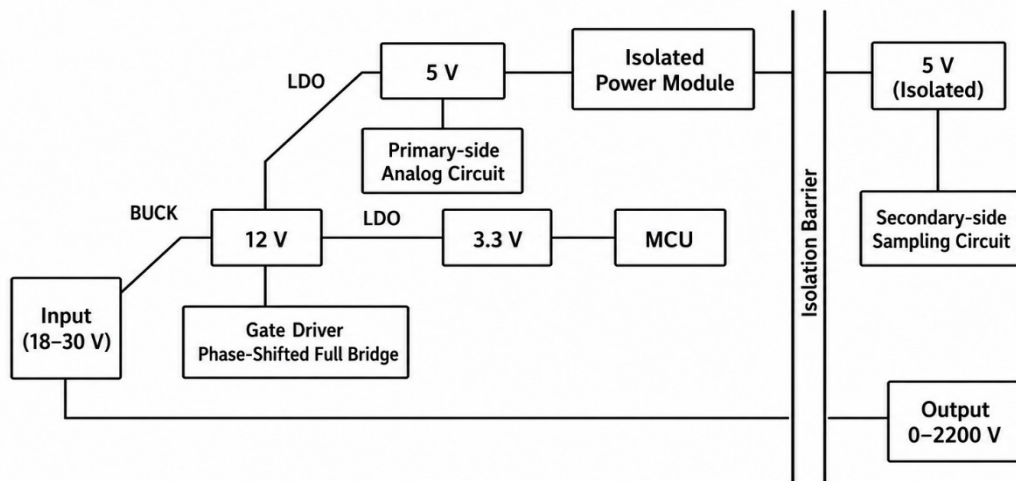


Figure 1 Power architecture diagram

Figure 1. Power architecture diagram.

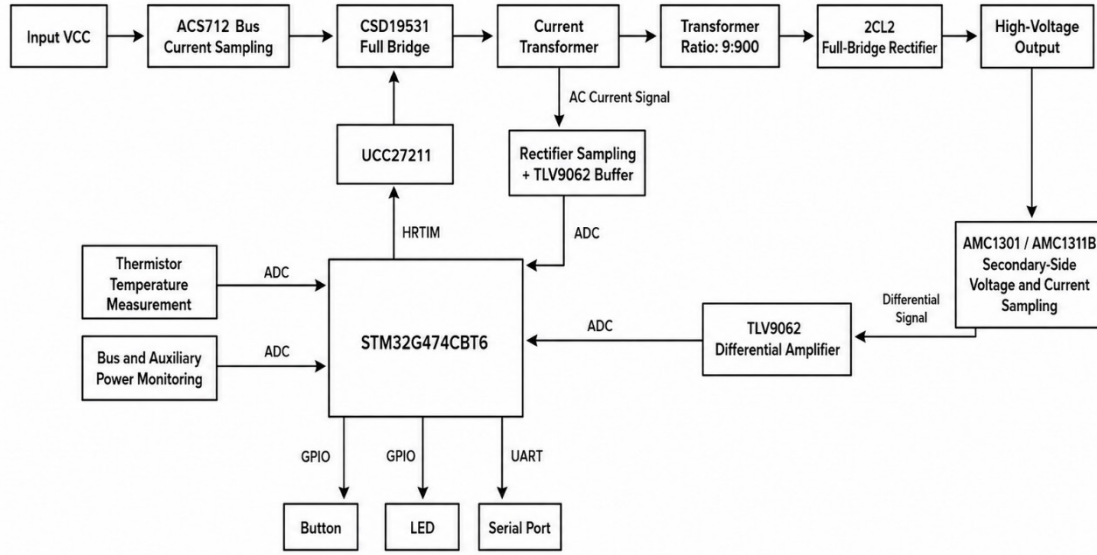


Figure 2. Hardware control and protection architecture diagram.

The phase-shift duty cycle is the key control variable of the PSFB. Let the switching period be T_s and the effective phase-shift time between the two bridge arms be t_ϕ ; then, the normalized duty cycle is defined as:

$$d = \frac{2t_\phi}{T_s} = \frac{\phi}{\pi}, 0 \leq d \leq 1 \quad (1)$$

For the ideal output side, the output power is determined by the output voltage and output current:

$$P_o = V_o I_o \quad (2)$$

Because the PSFB provides high-voltage step-up conversion, the transformer turns ratio in this study is large; therefore, the secondary rectifier, capacitor, and sampling voltage-divider network must withstand high voltage. Insulation, output-energy storage, diode reverse stress, and feedback reliability require special attention.

In this study, the secondary-side output current is used for constant-current feedback; therefore, current-limit control is applied directly to the high-voltage-side load current rather than to an indirect estimate based on the primary input current.

3. Digital control method

Four complementary phase-shifted gate-drive signals are generated by the HRTIM module of the STM32G474CBT6. The HRTIM uses the master timer as the period reference, while Timer A and Timer B correspond to the two bridge arms. The compare registers define the trigger points for the bridge arms. When the frequency changes, the period register, sampling trigger point, and lagging-arm compare point are recalculated so that the gate drive remains synchronized with sampling. This approach supports variable-frequency operation between 11 and 45 kHz and makes fixed dead time and fault shutdown straightforward to implement in a high-voltage power supply.

Feedback sampling uses switching-cycle synchronization. Twenty-four sampling points are collected in each cycle, and the output voltage, secondary current, and primary current are averaged. Let there be M sampling points in the k th cycle, and let the j th sample of a feedback variable be $x_j(k)$; the cycle average is:

$$x_{avg}(k) = \frac{1}{M} \sum_{j=1}^M x_j(k), M = 24 \quad (3)$$

This method reduces the effects of switching-node oscillation, rectifier recovery, and high-voltage-divider noise. The output voltage VSEC is periodically averaged to suppress sampling disturbance; the output current ISEC serves as the main feedback signal of the constant-current loop; the primary-side DC IPRIDC is used for input-power estimation; and the primary-side AC IPRIAC is mainly used for peak monitoring and hardware overcurrent protection.

The control layer consists of three PI loops: constant voltage, constant current, and constant power. The loops compute their respective duty-cycle candidates in parallel, and the smallest candidate is selected as the actual control input.

$$d_{cmd}(k) = \min\{d_{CV}(k), d_{CC}(k), d_{CP}(k), d_{max}\} \quad (4)$$

The underlying rule is “priority to the most restrictive constraint”. If the load current exceeds the set value, the constant-current loop reduces the duty cycle and takes over control; if the output approaches the target voltage, the constant-voltage loop takes over; and if the input or output power reaches its limit, the constant-power loop reduces the duty cycle. The system freezes the integrals of inactive loops to prevent drift. During soft start, the duty-cycle increase rate is limited, whereas the decrease rate is not, which balances smooth startup with a fast protection response.

4. Automatic frequency-conversion strategy

The fixed-frequency control structure is simple; however, it cannot cover the full high-efficiency operating range of an adjustable high-voltage output. Under low-output or light-load conditions, the duty cycle is low and little energy is transferred, while drive and magnetization losses remain. Under high-output or heavy-load conditions, the duty cycle approaches its upper limit and leaves insufficient control margin. The automatic frequency-conversion strategy uses the filtered duty cycle as the state variable. If the duty cycle is too low, the frequency is increased so that the operating point leaves the low-duty-cycle region; if the duty cycle is too high, the frequency is reduced so that the energy-transfer time per cycle increases. This logic is consistent with conclusions reported in a study on frequency-conversion-based PSFB efficiency optimization [5].

The automatic frequency-conversion parameters are listed in **Table 2**. The system reference frequency is set to 35 kHz, with an adjustment range of 11–45 kHz, and each adjustment step is limited to 1–4 kHz. Frequency reduction is triggered when the duty cycle exceeds 95%, with the target of reducing the predicted duty cycle to approximately 90%. The frequency-increase boundary depends on the current switching frequency so that low-duty-cycle operation does not persist for too long. To avoid frequent frequency jumps near the boundary, frequency-increase and frequency-reduction score counters and a lockout window are introduced. The system reloads the frequency only when the duty cycle remains in the corresponding region, and the score exceeds the threshold; it then enters a short lockout period during which no further frequency-conversion action is triggered.

Table 2. Automatic variable frequency control parameters

Parameter	Numerical value or rule	Function
Reference frequency	35 kHz	Default running frequency
Frequency range	11–45 kHz	Variable-frequency limit
Frequency step size	1–4 kHz	Single-step adjustment limit
Frequency-reduction trigger	$d > 95\%$	Avoid high-duty-cycle saturation
Frequency-reduction target	Approximately $d < 90\%$	Restore adjustment margin
Frequency-increase trigger	Low-duty-cycle region	Extend the light-load regulation range
Lockout window	100–200 ms	Suppress frequent switching
Feedforward exponent	$\gamma = 0.924$	Correct the duty cycle after frequency conversion

Changes in frequency affect the energy-transfer capability associated with a given duty cycle. If the frequency is changed while the duty cycle is kept unchanged, disturbances may appear in the output voltage and current. Therefore, this study uses feedforward duty-cycle correction to map the duty cycle at the old frequency to the initial duty cycle at the new frequency:

$$d_{ff} = \frac{2}{\pi} \arcsin \left[\text{clamp} \left(\sin \left(\frac{\pi d}{2} \right) \left(\frac{f_{new}}{f_{old}} \right)^\gamma, 0, 1 \right) \right] \quad (5)$$

In this expression, f_{old} is the frequency before conversion, f_{new} is the target frequency, and γ is the feedforward exponent. After the frequency is reloaded, the integral term of the PI loop currently in control is synchronized near the feedforward duty cycle, reducing the transient disturbance caused by re-establishing the feedback loop. The benefit of feedforward compensation for the dynamic response of PSFB converters has also been verified in related studies ^[6].

5. Firmware and monitoring implementation

The system firmware consists of six modules: sampling, drive generation, control, configuration, protection, and communication. The sampling module triggers the ADC and aggregates data. The drive module provides HRTIM output, dead time, fault input, and periodic reload functions. The control module performs CV/CC/CP arbitration, soft start, and automatic frequency conversion. The configuration module stores PI parameters, frequency limits, soft-start step size, and calibration coefficients. The protection module manages hardware OCP, auxiliary-supply faults, overtemperature protection, and the watchdog timer; the communication module exchanges operating settings and telemetry data with the host computer through the serial port.

The host computer sets the target voltage, current limit, power limit, and running time, and reads the output voltage, output current, primary voltage, primary current, switching frequency, phase-shift duty cycle, operating mode, and protection flags. It does not participate in real-time closed-loop control; it is used only for parameter tuning and status monitoring. Real-time control is performed by the MCU, allowing the system to operate independently according to the stored parameters if communication is lost.

6. Simulation and experimental verification

To confirm that the power stage can operate across the automatic frequency-conversion range, a PSFB high-voltage step-up model was established for operating-characteristic analysis, including a low-voltage full bridge, step-up transformer, secondary high-voltage rectifier, and output filter network. In addition, experimental efficiency tests were carried out on the developed prototype, and 309 valid measured datasets were obtained. The tested frequencies covered approximately 10–50 kHz, while the designed operating range of this project was 11–45 kHz. The duty cycles covered 5–100%, the secondary voltages covered approximately 48.95–981.3 V, and the secondary currents covered approximately 10–198.8 mA. Since the primary-side current was acquired using an onboard Hall sensor, zero-offset and bandwidth-related errors may exist; therefore, the measured efficiency results were used mainly to analyze variation trends rather than as metrology-grade efficiency data. **Figure 3** shows that efficiency is strongly correlated with frequency and duty cycle. Lower efficiency appears in the low-duty-cycle region, while a large cluster of high-efficiency points appears in the medium- to high-duty-cycle region. **Figure 4** shows the partitioning of the automatic frequency-conversion strategy. The boundaries among the frequency-increase zone, stable zone, frequency-decrease maintenance zone, and frequency-decrease zone are clearly defined, preventing the system from remaining for long periods in unfavorable low- or high-duty-cycle regions.

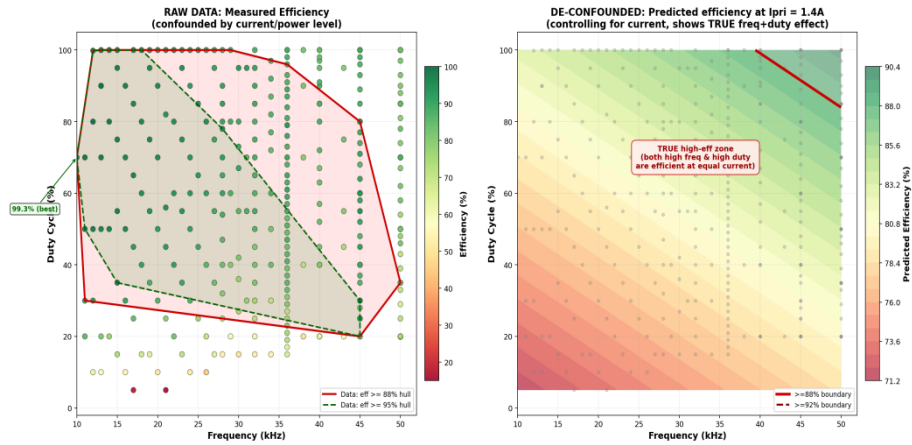


Figure 3. Frequency-duty cycle efficiency heat map and operating zones.

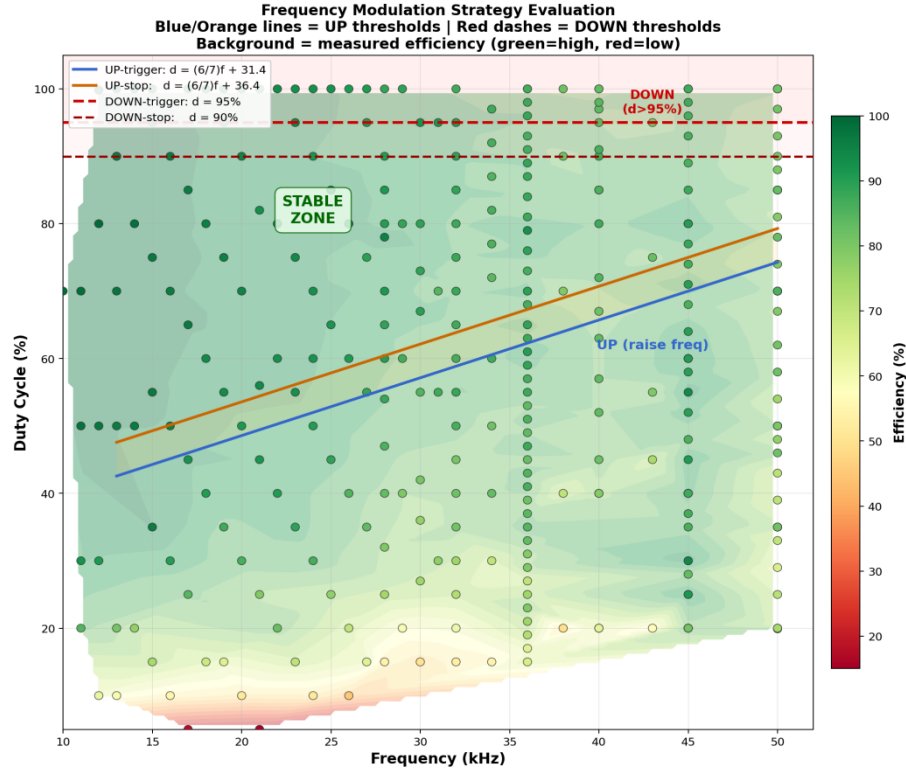


Figure 4. Automatic frequency conversion strategy zoning diagram.

ZVS is important for reducing switching loss in the PSFB. A previous study on full-bridge converters with a wide ZVS range indicates that the soft-switching range and filter requirements in variable-output applications affect efficiency and power density ^[10]. **Figure 5** shows the measured switching waveform of the prototype. Before the gate-drive signal arrives, the switching-node voltage falls close to zero, indicating that the MOSFET achieves near-zero-voltage turn-on under the tested conditions. This result shows that the designed power stage supports soft-switching operation and can operate over a wide frequency range under the automatic frequency-conversion strategy.

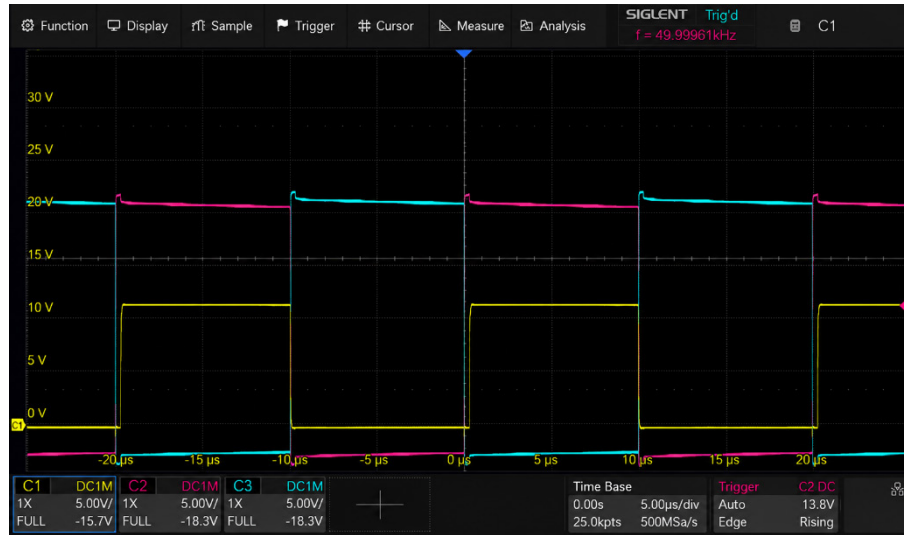


Figure 5. Measured waveform of ZVS on PSFB bridge arm.

The startup test confirmed the coordination between soft start and closed-loop control. **Figure 6** and **7** present the startup waveforms for 150 V and 230 V outputs, respectively; the output voltage rises along a controlled slope and settles near the target value. **Figure 8** and **9** show the ripple waveforms at these voltages, indicating that the system can maintain continuous voltage regulation under representative loads. Because of high-voltage safety requirements and probe limitations, full-scale ripple at kilovolt levels should be verified using a high-voltage differential probe.

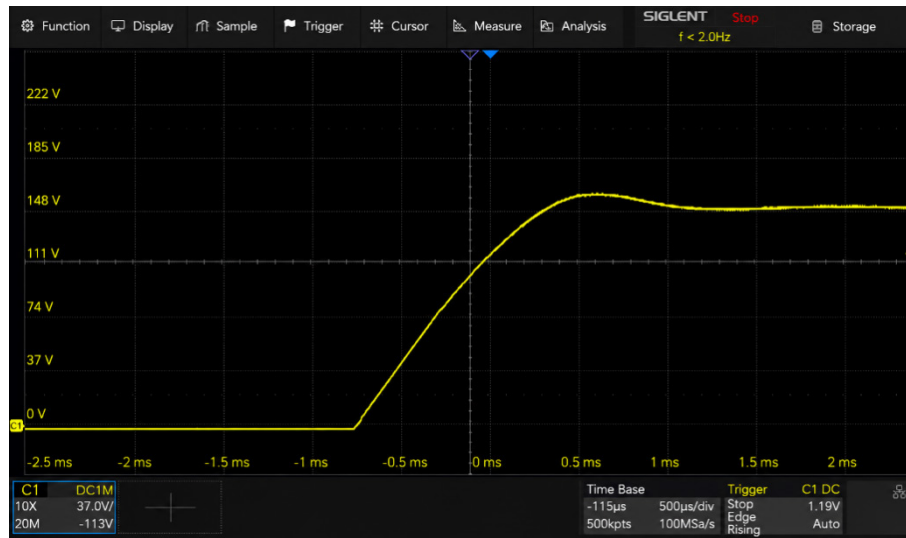


Figure 6. Startup waveform under 150 V output setting.

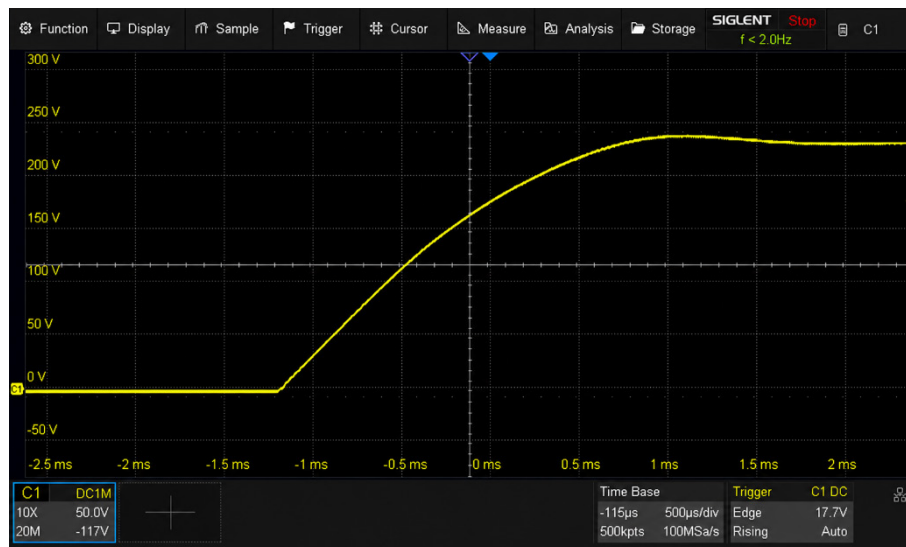


Figure 7. Startup waveform under 230 V output setting.

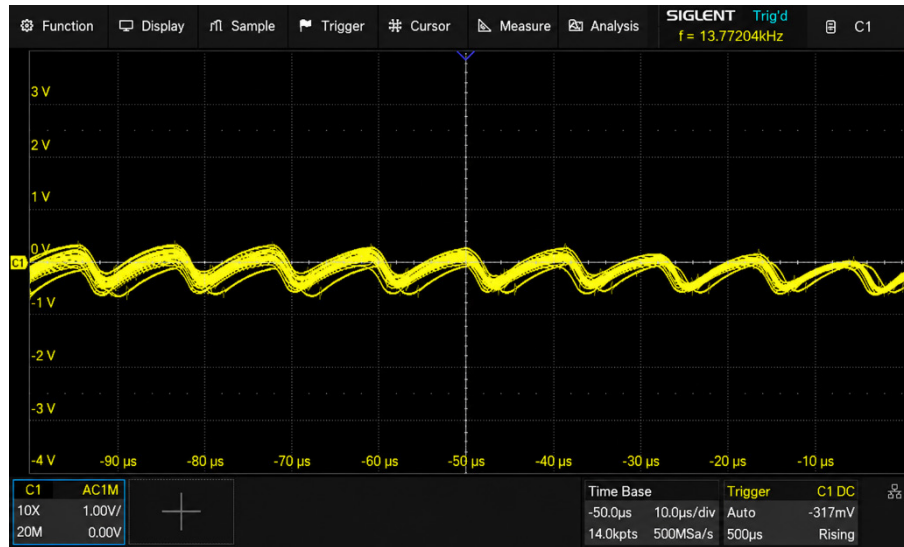


Figure 8. Voltage ripple waveform under 150 V output setting.

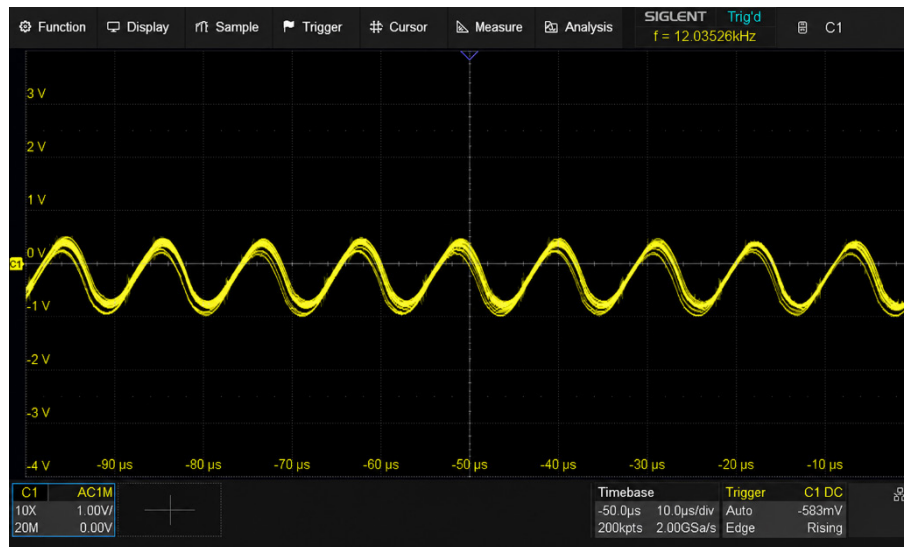


Figure 9. Voltage ripple waveform under 230 V output setting.

To verify the constant-current capability, two 1100 μF capacitors were connected to obtain an equivalent capacitance of approximately 550 μF . The target voltage was set to 2000 V, and the current limit was set to 200 mA. The test results are shown in **Figure 10**. The charging process took approximately 5.6 s, and the maximum output power was approximately 390 W, indicating that the CC loop can take over during the charging phase of the energy-storage load and transition smoothly as the output approaches the target voltage. To verify the constant-voltage capability, a 5 k Ω resistive load was used and the output voltage was set to 800 V. The test results are presented in **Figure 11**. The system can continuously provide a stable high-voltage output under load.



Figure 10. High-voltage capacitor constant current charging test.

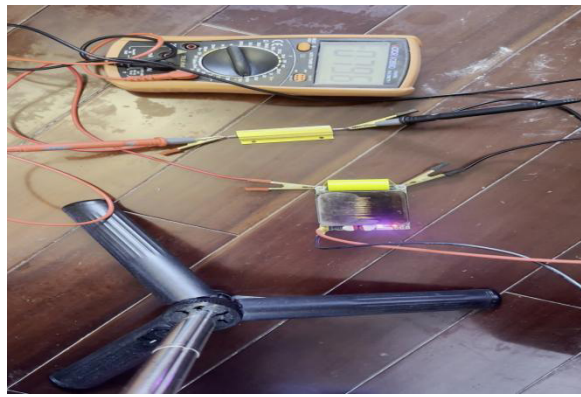


Figure 11. Constant voltage output test under resistive load.

7. Conclusion

This study designed a high-voltage constant-voltage/constant-current power supply based on a PSFB topology and an automatic frequency-conversion strategy. A typical 24 V low-voltage input and a 1:100 high-frequency transformer are used to achieve a maximum high-voltage output of 2200 V at 200 mA, with a 400 W power limit. Using the STM32G474CBT6 HRTIM module, the system generates four complementary phase-shifted gate-drive signals and performs periodic synchronous ADC sampling for voltage, current, auxiliary-supply, and temperature feedback. The parallel three-loop PI controller (CV, CC, and CP), minimum-duty-cycle arbitration, soft-start limiting, and integral freezing improve startup and mode-switching stability. To address the large variation in efficiency and regulation margin of fixed-frequency PSFB operation across a wide output range, an automatic frequency-conversion strategy based on duty-cycle status is proposed, with frequency ramp-up or ramp-down over 11–45 kHz. Disturbance during frequency conversion is reduced through score-counter hysteresis, lockout-window control, and feedforward duty-cycle correction. Simulations and experiments demonstrate stable high-voltage constant-voltage output, constant-current capacitor charging, ZVS turn-on, and low-ripple operation under representative working conditions. The system provides a digital control scheme with clear structure, configurable parameters, and good reproducibility for high-voltage capacitor charging and laboratory high-voltage power-supply applications.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Lyu D, Soeiro T, Bauer P, 2023, Design and Implementation of a Reconfigurable Phase Shift Full-Bridge Converter for Wide Voltage Range EV Charging Application. *IEEE Transactions on Transportation Electrification*, 9(1): 1200–1214.
- [2] Gao Y, Tang Y, Yao F, et al., 2024, Paralleled Variable Inductor Phase-Shifted Full-Bridge Converter with Full-Load Range ZVS and Low Duty Cycle Loss. *IEEE Transactions on Industrial Electronics*, 71(5): 4673–4684.
- [3] Lee S, Park J, 2021, A New Control for Synchronous Rectifier of Phase-Shifted Full-Bridge Converter to Improve Efficiency in Light-Load Condition. *IEEE Transactions on Industry Applications*, 57(4): 3822–3831.
- [4] Liu G, Wang B, Liu F, et al., 2023, An Improved Zero-Voltage and Zero-Current-Switching Phase-Shift Full-Bridge PWM Converter with Low Output Current Ripple. *IEEE Transactions on Power Electronics*, 38(3): 3419–3432.
- [5] Tang Y, Liu B, Ji D, et al., 2024, Efficiency Optimization for Phase-Shifted Full-Bridge Converter with Variable Frequency Control. *International Journal of Circuit Theory and Applications*, 52(6): 2724–2740.
- [6] Liu B, Tang Y, Ge L, et al., 2024, Feedforward Compensation for Phase-Shifted Full-Bridge DC–DC Converter under Peak Current Mode Control. *IEEE Transactions on Power Electronics*, 39(6): 6840–6851.
- [7] Bak Y, Lee Y, Lee K, 2022, Dynamic Characteristic Improvement of Phase-Shift Full-Bridge Center-Tapped Converters Using a Model Predictive Control. *IEEE Transactions on Industrial Electronics*, 69(2): 1488–1497.
- [8] Zhu Y, Xiao M, Ouyang H, et al., 2023, Predictive Deadbeat Average Model Control for Phase-Shifted Full-Bridge DC/DC Converter. *IET Power Electronics*, 16(3): 388–401.
- [9] Yang Y, Li H, Wu C, et al., 2022, An Improved Control Scheme for Reducing Circulating Current and Reverse Power of Bidirectional Phase-Shifted Full-Bridge Converter. *IEEE Transactions on Power Electronics*, 37(10): 11620–11635.
- [10] Chen J, Liu C, Liu H, et al., 2022, Zero-Voltage Switching Full-Bridge Converter with Reduced Filter Requirement and Wide ZVS Range for Variable Output Application. *IEEE Transactions on Industrial Electronics*, 69(7): 6805–6816.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on the Construction of a Digital Credit Bank System for Vocational Education Students Based on Blockchain Technology

Quanwei Zeng

Guangzhou Huanan Business College, Guangzhou 511300, Guangdong, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Against the backdrop of building a country strong in education, promoting high-quality development of vocational education stands as a core task. The credit bank serves as a critical institutional carrier linking academic education, non-academic training, and lifelong learning. China has completed the construction of basic platforms and pilot promotion for vocational education credits, yet practical operation is still plagued by inconsistent credit recognition standards, security risks of student information and credit data, and cumbersome cross-institutional credit mutual recognition procedures. Featuring distributed ledgers, cryptographic protection, consensus verification, and automatic execution via smart contracts, blockchain technology can target the operational drawbacks of traditional credit banks. On this basis, grounded in the national strategy for digital education development, this paper defines core concepts, sorts out the current status and dilemmas of credit bank construction, and explores the establishment of a blockchain-based digital credit bank system for vocational education students, aiming to provide valuable references for the development of vocational education credit banks.

Keywords: Blockchain technology; Vocational education; Digital credit bank; System construction; Learning outcome certification

Online publication: Aug 11, 2026

1. Introduction

The National Implementation Plan for Vocational Education Reform has arranged the construction of a national “credit bank” for vocational education, laying a fundamental foundation for tracing, querying, and converting learning outcomes. The Outline for Building a Country Strong in Education (2024–2035) further clarifies the establishment of a lifelong learning system based on qualification frameworks, centered on credit bank platforms and learning outcome certification. Traditional credit banks operate under a centralized management model, which creates numerous obstacles for centralized administration of diversified learning outcomes. Vocational institutions and training providers adopt inconsistent criteria for credit recognition;

data storage and circulation entail security risks; cross-agency credit mutual recognition lacks corresponding technical support, leading to unsatisfactory practical outcomes of credit banks. By adopting distributed bookkeeping and blockchain technology, credit banks can build multi-stakeholder trust and offer new solutions for the management, certification, and mutual recognition of credit data. Targeting vocational education students, this paper explores approaches to constructing a digital credit bank system supported by blockchain technology to drive the digitalization and credibility transformation of credit banks.

2. Definition of core concepts

2.1. Vocational education credit bank

Drawing on the management logic of financial savings, vocational education credit banks take credits as measurement units to uniformly register, archive, convert, and exchange various learning outcomes of vocational college students, including course learning, practical training, skill certificate acquisition, and competition awards ^[1]. This system breaks the restrictions of traditional vocational education in learning time, space, and forms, and bridges academic education and non-academic training. Students can accumulate credits according to their own learning progress, and accumulated credits can be used for course exemption, certificate exchange, and academic promotion. It acts as an important institutional tool to build talent development pathways in vocational education ^[2].

2.2. Core characteristics of blockchain technology

Blockchain is a distributed ledger technology (DLT) integrating distributed ledgers, point-to-point transmission, consensus mechanisms, cryptography, and smart contracts, maintained by multiple nodes with tamper-proof data.

Its core technical characteristics fall into five categories: First, data is stored across multiple nodes without a single centralized administrator. Second, once data is encrypted and uploaded to the chain, individual nodes cannot alter content arbitrarily. Third, public-private key paired encryption is adopted for identity verification and data protection. Fourth, multiple nodes jointly judge data validity. Fifth, preset code protocols can automatically execute corresponding processes once specific conditions are met.

3. Current status and practical dilemmas of vocational education credit bank construction

3.1. Significant differences in credit recognition standards impede cross-agency mutual recognition

Vocational colleges differ in talent training programs, curriculum settings, and credit accounting methods. As a result, identical learning outcomes are assigned distinct recognition standards and credit values across different colleges or training institutions. No universal reference framework for credit conversion has been formed between colleges or between colleges and vocational skill training institutions. Many educational entities hold low recognition of credits issued by external organizations due to curriculum resource protection and internal teaching management ^[3]. Independent credit systems operate separately among all educational providers. After students take courses across colleges or participate in training at external institutions, their accumulated credits cannot be smoothly recognized across entities, resulting in poor implementation of credit

mutual recognition.

3.2. Security risks exist in data storage and circulation

Most traditional vocational education credit banks adopt centralized servers for data storage. Sensitive data such as student identity information and credit acquisition/transformation records are aggregated in a single location. This storage mode easily triggers privacy leakage and malicious tampering of credit data in cases of cyberattacks, hardware failures, or human operational errors. Stakeholders of vocational education include higher vocational colleges, social training institutions, and skill evaluation agencies, whose management systems and data formats vary widely, hindering cross-platform data docking and causing delays in credit data circulation and updates ^[4]. Student information must be transmitted and verified across platforms during credit certification and cross-institutional conversion, while current workflows lack comprehensive security protection, failing to safeguard personal privacy fully.

3.3. Cumbersome procedures for credit certification and verification

Credit recognition, conversion, and inspection still rely heavily on offline manual workflows. Students need to prepare paper or electronic materials including transcripts, skill certificates, and training certificates on their own, while educational institutions conduct offline communication and multi-level approval for reviews, leading to lengthy and complicated overall procedures ^[5]. Manual review is also susceptible to subjective judgment of staff and inconsistent operational standards. Furthermore, there is no unified and complete channel to archive historical records of learning outcomes and credits, making it difficult for external parties such as employers and further education colleges to verify the authenticity and validity of credits rapidly.

3.4. Insufficient application scenarios and social recognition

When recruiting talent, employers prioritize academic diplomas and traditional vocational qualification certificates, paying little attention to credit outcomes from credit banks. Credit banks are mainly participated in by vocational colleges, with limited involvement from enterprises and industry associations. The scope of learning outcome recognition fails to cover all learning scenarios of vocational education. Some regional and institutional platforms only complete registration without launching actual credit conversion services, leaving the application value of platforms untapped.

4. Construction of blockchain-based digital credit bank system for vocational education students

4.1. Technical construction of layered architecture

The network layer serves as the basic operational carrier of the whole system, adopting a P2P distributed networking model and building node networks based on Peer nodes of Hyperledger Fabric. Connecting provincial education authorities, vocational colleges, skill training and evaluation institutions, cooperative enterprises, and other participants, the network layer establishes exclusive channels for data interaction between different platforms ^[6]. Operated under distributed logic, a single node failure will not disrupt the entire network, providing a stable underlying environment for credit data transmission and interaction.

The data layer acts as the core storage of credit data, constructing a distributed credit ledger via the SHA-256 hash algorithm and RSA asymmetric encryption technology. It stores student identity information,

detailed learning outcomes, credit accounting records, and conversion histories. All data is encrypted via hash functions to generate blocks connected in chronological order to form a chain structure. Authorized nodes within the consortium chain can access data within their permission scope, while sensitive information is encrypted for access only by authorized entities, balancing data transparency and privacy protection.

The consensus layer adopts the Raft consensus algorithm to verify the validity of credit data. After students submit learning outcomes, nodes including colleges, education authorities, and industry institutions conduct joint reviews. Data can only be written into blocks after all nodes reach consistent judgment results^[7]. This multi-stakeholder confirmation model strengthens the authority of credit management.

The contract layer adopts Go/Node.js/Java to develop chain codes covering credit recognition, credit conversion, credit review, and abnormal early warning scenarios. Compiled in accordance with national credit bank standards and vocational education credit management rules, contract codes preset learning outcome recognition criteria, credit conversion ratios, and review workflows to automatically execute operations once triggered, replacing manual work for standardized credit management.

The application layer develops visual interfaces for four types of users: students, colleges, enterprises and education authorities. The student terminal supports account registration, learning outcome submission, credit status inquiry and conversion application; the college terminal enables learning outcome review, course credit management and student credit file administration; the enterprise terminal allows viewing student credits and learning outcomes; the education authority terminal supports platform supervision, data statistics and standard formulation, meeting diverse credit management and usage demands of all stakeholders.

4.2. Construction of core functional modules

4.2.1. Student digital identity management module

After platform registration, the system generates an exclusive digital identity identifier for each student via asymmetric encryption technology, binding the digital identity with student school enrolment and identity information. The public key is used for identity verification and public inquiry, while the private key is kept by students themselves to authorize credit operations. The unique digital identity effectively prevents identity impersonation and protects student personal information security.

4.2.2. Learning outcome certification and storage module

This module covers all types of vocational education learning outcomes, including academic courses, enterprise training, 1 + X certificates, skill competitions, and industry certifications. After students upload supporting documents for learning outcomes, smart contracts complete preliminary verification following preset rules, and consortium chain nodes conduct consensus judgment. Qualified learning outcomes are converted into credits and recorded on the chain^[8]. Learning outcomes and credit information are permanently archived as evidence of students' learning experience.

4.2.3. Credit conversion and circulation module

Docked with the national lifelong education qualification framework, this module realizes credit conversion automatically via smart contracts based on unified credit conversion standards. Scenarios including cross-college study, certificate-to-credit exchange, and training outcome credit offsetting can complete automatic credit accounting and transfer through contracts^[9]. The module bridges credit circulation between colleges, training institutions, and enterprises, expanding credit application scenarios.

4.2.4. Credit traceability and verification module

The chain structure of blockchain fully records the whole process of credit generation, conversion, and review. Colleges, enterprises, and education authorities can access complete credit history through students' digital identities. External parties can verify the authenticity and validity of credits, replacing traditional manual verification and improving inspection efficiency ^[10].

4.2.5. Lifelong learning file management module

This module continuously records all learning outcomes and credit changes of students during their school years to establish complete digital learning files. The files include students' learning experience, skill improvement progress, and outcome certification results, serving as references for further education, employment, and skill evaluation, and providing authentic evidence for assessing vocational education talent training quality.

4.3. Platform docking and technical implementation

During system construction, official docking is realized with the National Smart Education Public Service Platform and the Vocational Education 1 + X Certificate Management Platform to obtain official data such as student enrolment information and certificate assessment results. Meanwhile, docking with higher vocational college educational administration systems and enterprise training management platforms automatically collects internal and enterprise-side data including course scores, training hours, and training outcomes to guarantee authentic data sources.

Unified standards for learning outcome classification, credit coding, and conversion ratios are formulated in line with national credit bank specifications, incorporating multi-source data into the system under unified rules. Smart contracts are developed to fit actual vocational credit management workflows and deployed online after multiple rounds of testing.

5. Conclusion

Vocational education credit banks constitute a key component of the lifelong learning system. Traditional credit banks suffer from practical issues such as poor adaptability of credit management, weak data security protection, and blocked cross-agency credit circulation. Based on the Hyperledger Fabric consortium chain, this paper integrates multiple key technologies and connects digital education platforms at all levels to build a five-layer technical architecture and five core functional modules, forming a comprehensive digital credit bank system for vocational education students. The system streamlines vocational credit management workflows, safeguards credit data security, and connects credit circulation across different stakeholders. In the future, platform functions can be continuously optimized in light of the characteristics of various vocational majors and application scenarios to promote wide-scale system implementation, advance the digital transformation of vocational education, and support the construction of a skill-oriented society and lifelong learning for all citizens.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Wu D, Cheng Q, Wang L, et al., 2026, Practical Dilemmas and Solutions of Credit Mutual Recognition in Higher Vocational Credit Banks. *Journal of Jiangsu Vocational Institute of Architectural Technology*, 26(1): 85–89.
- [2] Guan Y, Zheng W, He X, et al., 2026, Strategies and Paths for Credit Banks Driven High-Quality Development of Vocational Education Empowered by Digital Intelligence. *Journal of Continuing Education*, 2026(4): 37–42.
- [3] Mao S, 2025, Vocational Education Development Under Digital Transformation: Path Reconstruction and Ecological Innovation. *Journal of Shantou University (Humanities & Social Sciences Edition)*, 41(6): 89–93 + 96.
- [4] Guo F, Chen C, Liu X, 2025, Regional Experience and Optimization Path of National Vocational Education Credit Bank Construction. *Education Science Forum*, 2025(12): 12–16.
- [5] Feng C, Wang X, 2024, Construction of Management and Operation Mechanism of Vocational Education Credit Banks in China. *Journal of Nanning Vocational & Technical College*, 32(3): 51–57.
- [6] Gao L, Cheng S, 2024, Improving the Construction of National Vocational Education Credit Banks Drawing on Experiences of Lifelong Education Credit Banks. *Vocational Education Research*, 2024(1): 64–68.
- [7] He M, 2023, Credit Bank: A Strategic Initiative for Building a Learning Society. *Online Learning*, 2023(11): 42 + 88.
- [8] Luo Y, 2023, Research on Application Approaches of Blockchain Technology in Vocational Education Credit Banks. *Vocational Education*, 22(20): 22–25 + 40.
- [9] Zheng S, 2023, Construction of Vocational Education Credit Bank Network System Based on Blockchain Technology. *Digital Technology & Application*, 41(5): 208–210.
- [10] Yang S, 2020, Thoughts on Applying Blockchain Technology to Credit Bank Construction. *Journal of Fujian Radio & TV University*, 2020(2): 10–15.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Post-Grasp Pose Adjustment of a Tendon-Driven Underactuated Robotic Hand Using Finite-Action-Set Receding-Horizon Planning

Lei Hu, Hongran Li*

Jiangsu Ocean University, Lianyungang, 222005 Jiangsu, China

*Corresponding author: Hongran Li, lihongran@jou.edu.cn

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This paper addresses local post-grasp pose adjustment for a tendon-driven underactuated robotic hand. A finite-action-set receding-horizon motor reference planning method is proposed, in which the continuous motor reference increment is discretized into representative actions and an action-effect table predicts local object pose changes. At each decision instant, candidate action sequences are evaluated using pose error, action magnitude, motor limits, grasp margin, and action switching, while only the first action of the optimal sequence is executed. Simulations show that the method generates feasible commands for single-axis and two-dimensional pose adjustment, drives the object pose toward the target, and keeps motor references within the allowable range. Compared with one-step planning, a longer prediction horizon reduces action switching in the two-dimensional coupled task and improves action sequence smoothness.

Keywords: Underactuated robotic hand; Tendon-driven; Post-grasp pose adjustment; Finite action set; Receding-horizon planning

Online publication: Aug 11, 2026

1. Introduction

Multi-fingered robotic hands can provide multiple contact points and flexible grasp postures, and therefore have important potential in robotic grasping and manipulation. Tendon-driven underactuated robotic hands have attracted increasing attention because of their compact structure, reduced number of actuators, lightweight design, and simple command interface ^[1-3]. However, a single motor usually affects multiple joints through tendon transmission, so the motor input, joint motion, and object pose response are strongly coupled. As a result, fine post-grasp pose adjustment is difficult to achieve through independent joint regulation.

In practical manipulation tasks, establishing a stable grasp is often not sufficient. After an object is

grasped, small pose corrections may still be required, such as adjusting the tilt angle, changing the local orientation, or compensating for pose deviation caused by contact. Related problems have been studied through in-hand reorientation, multi-modal planning, trajectory optimization, model predictive control, feedback control, and action primitives^[4–11]. Nevertheless, for tendon-driven underactuated hands, the object pose response is affected by contact location, friction condition, object shape, and finger configuration. Generating executable adjustment commands without relying on a complete contact dynamics model therefore remains challenging.

Model-driven methods can describe mechanical relationships during manipulation, but often require high modeling and computational effort and are sensitive to contact states and object parameters^[6,13,14,18]. Learning-based methods can learn complex manipulation strategies from interaction data, but usually require many training samples, and their interpretability and transferability remain limited^[9,12,15,16]. In comparison, finite-action-set methods reduce online decision complexity by converting continuous control inputs into a small number of representative actions^[17]. However, simple one-step action selection may ignore the influence of the current action on subsequent adjustment steps, leading to frequent action switching or insufficiently smooth adjustment.

To address this problem, this paper proposes a finite-action-set receding-horizon motor reference planning method for post-grasp local pose adjustment of a tendon-driven underactuated robotic hand. The initial grasp is assumed to be stable, and the focus is upper-level generation of executable motor reference increment sequences rather than grasp control or contact force allocation. The continuous motor reference increment is discretized into a finite action set, and an action-effect table describes the average influence of each action on the local object pose. At each decision instant, the planner evaluates candidate action sequences over a finite horizon and considers pose error, action magnitude, motor limits, grasp margin, and action switching.

2. Materials and methods

2.1. System description and problem formulation

This study considers local post-grasp pose adjustment after a tendon-driven underactuated robotic hand has established a stable grasp. The object is held by the fingers, and the upper-level planner generates motor reference commands to move the object toward a desired local pose gradually. The planning variable is selected in the motor reference space because it is directly executable by the lower-level motor controller and avoids explicit optimization of joint torques, contact forces, or tendon tensions.

For the i -th finger, the tendon transmission relationship is written as

$$l_i = r_i \theta_i - A_i q_i - l_{0,i} \quad (1)$$

where the symbols denote the tendon length variation, motor spool radius, joint angle vector, geometric coupling matrix, and initial length offset, respectively. At each decision instant, the planner selects a motor reference increment and updates the reference command. The increment and the motor reference are both constrained to prevent abrupt motion and to keep the command within the allowable actuation range.

For the single-axis task, the object state is represented by the tilt angle. For the two-dimensional task, the state includes the tilt angle and the in-plane rotation angle. Given the desired pose, the post-grasp adjustment task is formulated as finite action sequence generation:

$$\min_{a_{0:K-1}} \sum_{k=0}^K \|s_d - s_k\|_Q^2 + \frac{\lambda_\theta \sum_{k=0}^{K-1} \|a_k\|^2}{\|\delta\|^2} + \lambda_l \sum_{k=0}^{K-1} P_{lim}(k) + \lambda_s \sum_{k=0}^{K-1} P_{stab}(k), a_k \in A \quad (2)$$

Here, A is the finite action set, δ is the basic action step, P_{lim} penalizes unsafe proximity to motor limits, and P_{stab} penalizes insufficient grasp margin. This formulation focuses on upper-level reference planning after grasp establishment, rather than complete grasp stability control.

2.2. Finite action set and pose prediction

The continuous motor reference increment is discretized into a small set of representative actions. Each action is represented by a normalized increment vector and scaled by the basic action step. In the simulation, the step size is 3° , and the motor order is thumb, index finger, middle finger, ring finger, and pinky finger. Only the thumb, index finger, and middle finger are assigned nonzero increments, while the ring finger and pinky finger provide auxiliary support. **Table 1** combines the finite action set and the nominal action effects. The two effect columns describe the average changes in tilt angle and in-plane rotation angle, respectively.

Table 1. Finite action set and nominal action effects

Action	sigma	Tilt effect	Rotation effect
a0	[0,0,0,0,0]	0.0	0.0
a1	[1,0,0,0,0]	0.4	-0.8
a2	[0,1,0,0,0]	0.9	0.5
a3	[0,0,1,0,0]	0.7	-0.3
a4	[-1,0,0,0,0]	-0.4	0.8
a5	[0,-1,0,0,0]	-0.9	-0.5
a6	[0,0,-1,0,0]	-0.7	0.3
a7	[1,-1,0,0,0]	1.0	-1.2
a8	[-1,1,0,0,0]	-1.0	1.2
a9	[0,1,-1,0,0]	1.3	0.8
a10	[0,-1,1,0,0]	-1.3	-0.8

At each decision step, actions that would violate the motor limits are excluded. The predicted pose is obtained by adding the nominal action effect to the current pose. To model the mismatch between the nominal table and the actual response, the simulated pose update is

$$s_{k+1} = s_k + (1 + \rho_k) \mu_{i_k} + \xi_k \quad (3)$$

where the selected action index determines the nominal action effect, the proportional action-effect error is sampled from a uniform distribution, and the last term represents zero-mean Gaussian measurement noise.

2.3. Receding-horizon planning and simulation settings

At each decision step, the planner observes the current object pose and motor reference, constructs the feasible action set, and evaluates candidate sequences over the prediction horizon. Starting from the current state, the predicted pose and motor reference are propagated using the action-effect table, and sequences that violate motor limits are discarded. Candidate sequences are evaluated by

$$J = \sum_{h=1}^H \lambda_e \|s_d - \hat{s}_{k+h}\|_Q^2 + \frac{\lambda_\theta \sum_{h=0}^{H-1} \|a_{i_h}\|}{\|\delta\|^2} + \lambda_l \sum P_{lim} + \lambda_s \sum P_{stab} + \lambda_c \sum P_{chg} \quad (4)$$

Only the first action of the optimal sequence is executed, and the same optimization is repeated after the object pose is updated. The loop stops when the pose error is below the prescribed threshold or the maximum number of decision steps is reached. The main simulation settings are as follows: motor reference range 0–90°, initial reference. $[45, 45, 45, 45, 45]^T$ deg, prediction horizon $H=3$, maximum step number 50, maximum action-effect error 0.1, measurement noise standard deviation 0.1°, and 20 repeated trials for each target. The stopping threshold is 1.0° for the single-axis task and 1.5° for the two-dimensional task. The weights are $\lambda_e=1.0$, $\lambda_\theta=0.05$, $\lambda_l=50$, $\lambda_s=10$, and $\lambda_c=0.1$. Reported metrics include final error, average error, adjustment steps, action energy, and successful trials

3. Simulation results

3.1. Single-axis pose adjustment

The first simulation evaluates single-axis pose adjustment. The object is initially placed at zero tilt, and the target tilt angle is selected from 10°, 15°, 20°, -10°, -15°, and -20°. **Figure 1** shows a representative result for the 20° target.

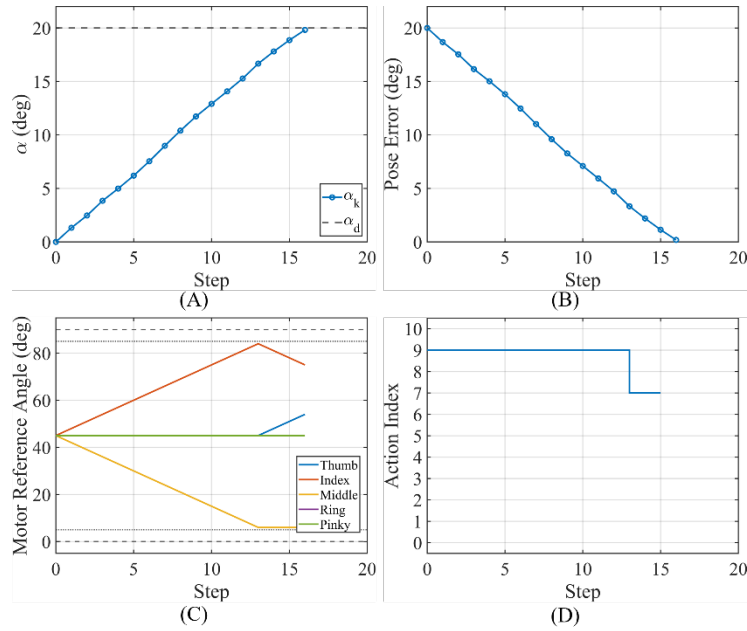


Figure 1. Single-axis pose adjustment result for the 20° target. (A) Object tilt angle. (B) Pose error. (C) Motor reference angles. (D) Selected action sequence.

As shown in **Figure 1**, the object tilt angle gradually approaches the target and the pose error decreases to the target neighborhood. The motor reference angles remain within the allowable range, indicating that the generated commands satisfy the actuation constraints. **Table 2** summarizes the results over all single-axis targets. All repeated trials reached the prescribed stopping threshold, and the mean final error remained below 0.64° for all targets.

Table 2. Single-axis pose adjustment results

Target (°)	Final error (°)	Average error (°)	Steps	Action energy
10	0.50	5.24	7.35	14.60
15	0.43	7.69	11.25	22.30
20	0.63	10.03	15.40	30.80
-10	0.48	5.23	7.45	14.80
-15	0.44	7.71	11.30	22.60
-20	0.57	9.97	15.45	30.90

3.2. Two-dimensional pose adjustment

The second simulation evaluates two-dimensional local pose adjustment, where the object pose includes the tilt angle and the in-plane rotation angle. The initial pose is $[0^\circ, 0^\circ]^T$. Figure 2 shows a representative result for the target $[15^\circ, -8^\circ]^T$.

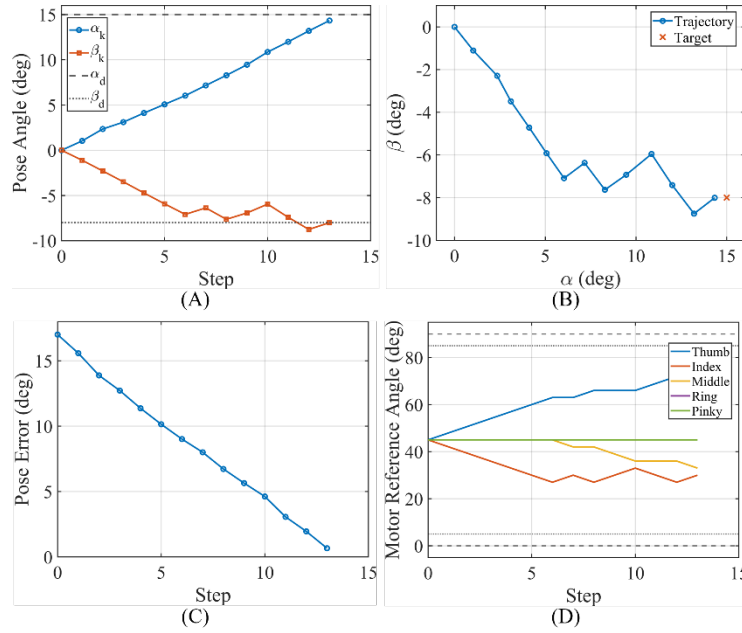


Figure 2. Two-dimensional pose adjustment result for the target $[15^\circ, -8^\circ]^T$. (A) Pose angles. (B) Pose trajectory in the alpha-beta plane. (C) Pose error norm. (D) Motor reference angles.

Figure 2 shows that the planner can regulate two coupled pose components using the same finite action set. Since each selected action may influence both pose components, the trajectory is not expected to be a straight line in the pose plane. Even so, the pose moves toward the target, the error norm decreases, and the motor references remain within the predefined limits. **Table 3** lists the statistical results for all two-dimensional targets. The mean final errors were all below 1.16° , and all repeated trials reached the stopping threshold.

Table 3. Two-dimensional pose adjustment results

Target pose (°)	Final error (°)	Average error (°)	Steps	Action energy
$[10, 5]^T$	1.05	5.94	7.15	14.15
$[15, -8]^T$	0.99	8.68	12.95	25.80
$[-10, 8]^T$	1.16	6.61	8.65	17.10
$[-15, -10]^T$	0.96	9.43	11.30	22.55
$[20, 0]^T$	0.94	10.28	16.25	32.40

An additional comparison between $H=1$ and $H=3$ showed that both settings reached the prescribed stopping threshold in all tested trials. In the two-dimensional task, the average number of action switches decreased from 3.70 for $H=1$ to 3.13 for $H=3$, while the accuracy-related metrics remained close. This suggests that the longer prediction horizon mainly improves action sequence smoothness rather than final pose accuracy under the present simulation settings. These numerical results support the feasibility of the proposed upper-level planner, but they do not imply guaranteed convergence under arbitrary contact conditions or direct validation on a physical robotic hand.

4. Conclusion

This paper investigated local post-grasp pose adjustment for a tendon-driven underactuated robotic hand and proposed a finite-action-set receding-horizon motor reference planning method. The continuous motor reference increment was discretized into representative actions, and a nominal action-effect table was used for local pose prediction. At each decision instant, the planner evaluated candidate sequences by considering pose error, action magnitude, motor limits, grasp margin, and action switching, executed only the first action of the optimal sequence, and replanned after feedback was obtained.

Simulation results showed that the proposed method can generate feasible motor reference increment sequences for both single-axis and two-dimensional local pose adjustment tasks. The object poses gradually approached the target neighborhood while the motor references remained within the allowable range. Compared with one-step planning, the $H=3$ setting reduced action switching in the two-dimensional task while maintaining comparable accuracy-related performance. Future work will focus on physical experiments, online updating of the action-effect table, and integration of tactile or contact feedback.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Li H, Ford C, Bianchi M, et al., 2022, BRL/Pisa/IIT SoftHand: A Low-Cost, 3D-Printed, Underactuated, Tendon-Driven Hand with Soft and Adaptive Synergies. *IEEE Robotics and Automation Letters*, 7(4): 8745–8751.

- [2] Zhou X, Fu H, Shentu B, et al., 2024, Design and Control of a Tendon-Driven Robotic Finger Based on Grasping Task Analysis. *Biomimetics*, 9(6): 370.
- [3] Li G, Liang X, Gao Y, et al., 2024, A Linkage-Driven Underactuated Robotic Hand for Adaptive Grasping and In-Hand Manipulation. *IEEE Transactions on Automation Science and Engineering*, 21(3): 3039–3051.
- [4] Morgan A, Hang K, Wen B, et al., 2022, Complex In-Hand Manipulation Via Compliance-Enabled Finger Gaiting and Multi-Modal Planning. *IEEE Robotics and Automation Letters*, 7(2): 4821–4828.
- [5] Chen T, Tippur M, Wu S, et al., 2023, Visual Dexterity: In-Hand Reorientation of Novel and Complex Object Shapes. *Science Robotics*, 8(84): eadc9244.
- [6] Jiang Y, Yu M, Zhu X, et al., 2024, Contact-Implicit Model Predictive Control for Dexterous In-Hand Manipulation: A Long-Horizon and Robust Approach. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems: 5260–5266.
- [7] Yu M, Jiang Y, Chen C, et al., 2025, Robotic In-Hand Manipulation for Large-Range Precise Object Movement: The RGMCM Champion Solution. *IEEE Robotics and Automation Letters*, 10(5): 4738–4745.
- [8] Sieler A, Brock O, 2023, Dexterous Soft Hands Linearize Feedback-Control for In-Hand Manipulation. 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems: 8757–8764.
- [9] Saito D, Kanehira A, Sasabuchi K, et al., 2024, APriCoT: Action Primitives Based on Contact-State Transition for In-Hand Tool Manipulation. 2024 IEEE-RAS 23rd International Conference on Humanoid Robots: 827–834.
- [10] Bircher W, Morgan A, Dollar A, 2021, Complex Manipulation with a Simple Robotic Hand Through Contact Breaking and Caging. *Science Robotics*, 6(54): eabd2666.
- [11] Teeple C, Aktas B, Yuen M, et al., 2022, Controlling Palm-Object Interactions Via Friction for Enhanced In-Hand Manipulation. *IEEE Robotics and Automation Letters*, 7(2): 2258–2265.
- [12] Yu C, Wang P, 2022, Dexterous Manipulation for Multi-Fingered Robotic Hands with Reinforcement Learning: A Review. *Frontiers in Neurorobotics*, 16: 861825.
- [13] Chen T, Xu J, Agrawal P, 2021, A System for General In-Hand Object Re-Orientation. *arXiv*, arXiv:2111.03043.
- [14] Chanrungrameekul P, Ren K, Grace J, et al., 2023, Non-Parametric Self-Identification and Model Predictive Control of Dexterous In-Hand Manipulation. 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems: 8743–8750.
- [15] Weinberg A, Shirizly A, Azulay O, et al., 2024, Survey of Learning-Based Approaches for Robotic In-Hand Manipulation. *Frontiers in Robotics and AI*, 11: 1455431.
- [16] Sintov A, Kimmel A, Wen B, et al., 2020, Tools for Data-Driven Modeling of Within-Hand Manipulation with Underactuated Adaptive Hands. *Proceedings of Machine Learning Research*, 120: 771–780.
- [17] Patidar S, Sieler A, Brock O, 2023, In-Hand Cube Reconfiguration: Simplified. 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems: 8751–8756.
- [18] Hess A, Kübler A, Forrai B, et al., 2025, Sampling-Based Model Predictive Control for Dexterous Manipulation on a Biomimetic Tendon-Driven Hand. 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems: 5509–5516.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on High-P Precision Match Drilling Method for Reaming Mounting Holes When Replacing Sintering Machine Sprocket Tooth Plates

Yun Ma, Zhifeng Chen, Wei Wei, Hanxiao Ma, Huan Jin

Xinjiang Institute of Engineering, Urumqi, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The operational stability and service life of sintering machines are critically dependent on the machining precision of reaming mounting holes in their sprocket tooth plates. To address the limitations of existing match-drilling techniques, this paper proposes a novel high-precision match-drilling method aimed at enhancing assembly accuracy, installation efficiency, and overall operational stability. Through a systematic analysis of current match-drilling processes, this study identifies the alignment accuracy between the reaming holes of the tooth plates and the web plates as the decisive factor in the assembly procedure. To resolve this issue, a custom-designed measuring and positioning tool, in conjunction with an innovative match-drilling workflow, is developed to form a complete reaming scheme. Experimental validation, including positioning tests on multiple tooth plates, confirms the effectiveness and repeatability of the proposed method. The results demonstrate that this approach significantly improves assembly precision and operational stability, reduces installation errors, and lowers both the failure rate and maintenance costs of sintering machines. Beyond its practical contributions, this research enriches the theoretical framework of mechanical processing and assembly, offering robust technical support for on-site applications. Moreover, the standardization of the new match-drilling process provides a comprehensive maintenance solution for routine overhauls of sintering machine sprockets and tooth plates, thereby promoting continuous innovation and advancement in related maintenance strategies.

Keywords: Sintering machine; Sprocket tooth plate; Match-drilling; High precision; Replacement

Online publication: Aug 11, 2026

1. Introduction

1.1. Research background

In the metallurgical machinery industry, the performance of sintering machines is closely related to the precise fitting of all their components. As a key component of sintering machines, the machining precision

of reaming mounting holes on sprocket tooth plates directly determines the overall operation stability and service life of sintering machines. In practical operation at present, sprocket tooth plates are generally replaced once every three years. The precise match drilling requirements in this replacement process are often restricted by on-site maintenance space and equipment installation precision requirements. Such restrictions increase operational complexity and may lead to misalignment of corresponding reaming mounting holes during tooth plate installation, resulting in incomplete installation of reaming bolts and further interfering with the normal operation of sintering machines. Since spare tooth plates purchased from original manufacturers are not pre-fabricated with reaming holes, on-site matched drilling and reaming between tooth plates and sprockets are required. However, due to uncontrolled precision during matched drilling of paired reaming holes, deviations occur between the two matching holes.

Although traditional match drilling technologies have been continuously improved, they still have limitations. When replacing sprocket tooth plates with traditional methods, positioning and match drilling for new pre-installed tooth plates must rely on the reaming holes on sintering sprocket web plates. Limited on-site space prevents horizontal match drilling with dedicated equipment. The conventional operation is as follows: mark corresponding positions on old tooth plates and web plates before removing the old ones; dismantle the old tooth plates, align and stack them with new tooth plates on a bench drill, and conduct match drilling on new tooth plates by taking reaming holes of worn old tooth plates as benchmarks ^[1]. Due to abrasion of old tooth plates, alignment errors will be generated when stacked with new plates. After match drilling is finished, misalignment between the reaming holes of tooth plates and web plates frequently appears during installation, making it impossible to mount reaming bolts. This exerts a severe negative impact on subsequent sintering machine production and operation, and may even trigger accidents such as tooth plates clamping trolley wheels ^[2].

With the continuous development of measurement technology and 3D modeling simulation methods, new matched drilling schemes for reaming mounting holes demonstrate prominent advantages. Such schemes can effectively eliminate adverse impacts caused by space constraints on the premise of guaranteeing precision, and realize more accurate assembly. Research on this innovative method aims to promote theoretical progress in mechanical processing and provide more reliable technical support for practical sintering machine operation, hence possessing broad application prospects and promotion value. The implementation of scientific match drilling standards and procedures will bring more efficient solutions to relevant industries and lay a new foundation for continuous industrial innovation.

1.2. Research purpose and significance

With the continuous development of sintering machine mechanical equipment management, higher requirements are put forward for the stability of sintering machines. During the operation of sintering machines, the installation precision of sprocket tooth plates is extremely critical. Once a phase difference exists between tooth plates on both sides of the head sprocket, severe trolley deviation of sintering machines will be triggered. Therefore, the main purpose of this paper is to verify the effectiveness of the high-precision match drilling method for sprocket tooth plates during replacement, and illustrate possible errors in the assembly process through comparison of different drilling methods ^[3]. To overcome the limitations of existing assembly schemes for sintering machine sprocket tooth plates, this paper focuses on analyzing the positioning link in tooth plate assembly. Self-made measuring and positioning tools are developed

according to practical working conditions, a complete set of match drilling procedures is formulated, and an innovative drilling scheme for reaming mounting holes is designed. When practically applied in regular sintering machine maintenance, this method is expected to significantly improve assembly precision, reduce the phase difference between tooth plates on both sides of sprockets, and lower the equipment failure rate and production cost of sintering machines. Such innovations can not only improve the working efficiency of sintering machines, but also provide valuable practical data and theoretical support for maintenance management of related sintering equipment, boosting the development and technical iteration of equipment management ^[4].

2. Match drilling method

2.1. Design of the match drilling scheme

When formulating the high-precision match drilling scheme for reaming mounting holes of sintering machine sprocket tooth plates, multiple factors including on-site working environment and machining capacity of drilling equipment shall be fully considered to ensure the feasibility of this scheme in practical application. The precision of matched drilling and assembly of tooth plate reaming holes is not only related to the magnitude of phase difference between tooth plates on both sides of sintering machine sprockets, but also directly affects the service life of sprocket tooth plates ^[5]. The traditional method stacks old and new tooth plates together and uses reaming holes of worn old plates as benchmarks to drill new ones. During assembly, technical problems such as poor fitting between new tooth plates and sprocket web plates frequently emerge. Therefore, a brand-new machining and installation scheme with higher drilling precision must be adopted to raise the assembly accuracy of sprocket tooth plates.

Aiming at the insufficient precision of the traditional stacking match drilling technology for old and new sprocket tooth plates, this paper proposes an innovative high-precision match drilling method. First, a corresponding measuring and positioning tool is manufactured according to the specific dimensions of reaming assembly holes on sprocket web plates. Through detailed analysis of geometric features of reaming mounting holes and the on-site installation environment, a completely new match drilling and installation scheme for tooth plates is formulated ^[6]. With the special positioning device for match drilling of tooth plate reaming holes, the center points of reaming holes can be precisely positioned, laying a reliable foundation for subsequent tooth plate installation. The special positioning device consists of disassembly nuts, locking nuts, expansion sleeves, and taper sleeves (see **Figure 1**). A 10 mm-diameter drilling hole is opened at the exact center of the taper sleeve, and the expansion sleeve and taper sleeve are combined to form a positioning fixture. Before match drilling of tooth plates, install tooth plates at corresponding sprocket positions with ordinary bolts; mount expansion sleeves and taper sleeves into reaming holes of sprocket web plates, and fix the positioning device with locking nuts. At this time, the central hole of the taper sleeve is aligned with the center of the reaming hole. Construction workers then use a hand drill with a 10mm drill bit to drill a hole with a depth of about 3–5 mm at the corresponding position of the tooth plate through the central positioning hole of the taper sleeve, marking the completion of positioning drilling for tooth plate reaming holes ^[7]. Remove the positioning device with disassembly nuts, take down the tooth plates with positioning holes, and transfer them to a bench drill. Take the small positioning holes as benchmarks and use a 44mm match drill bit to complete reaming drilling ^[8]. After match drilling, install the tooth plates at corresponding positions and fasten reaming bolts, and misalignment between the reaming holes of the web plates and tooth plates will no

longer occur (see **Figure 2**).

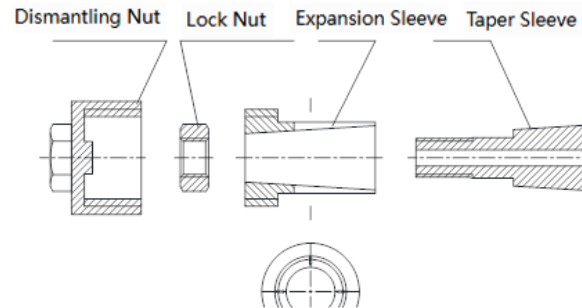


Figure 1. Component diagram of positioning device for match drilling of sintering machine sprocket tooth plate reaming mounting holes.

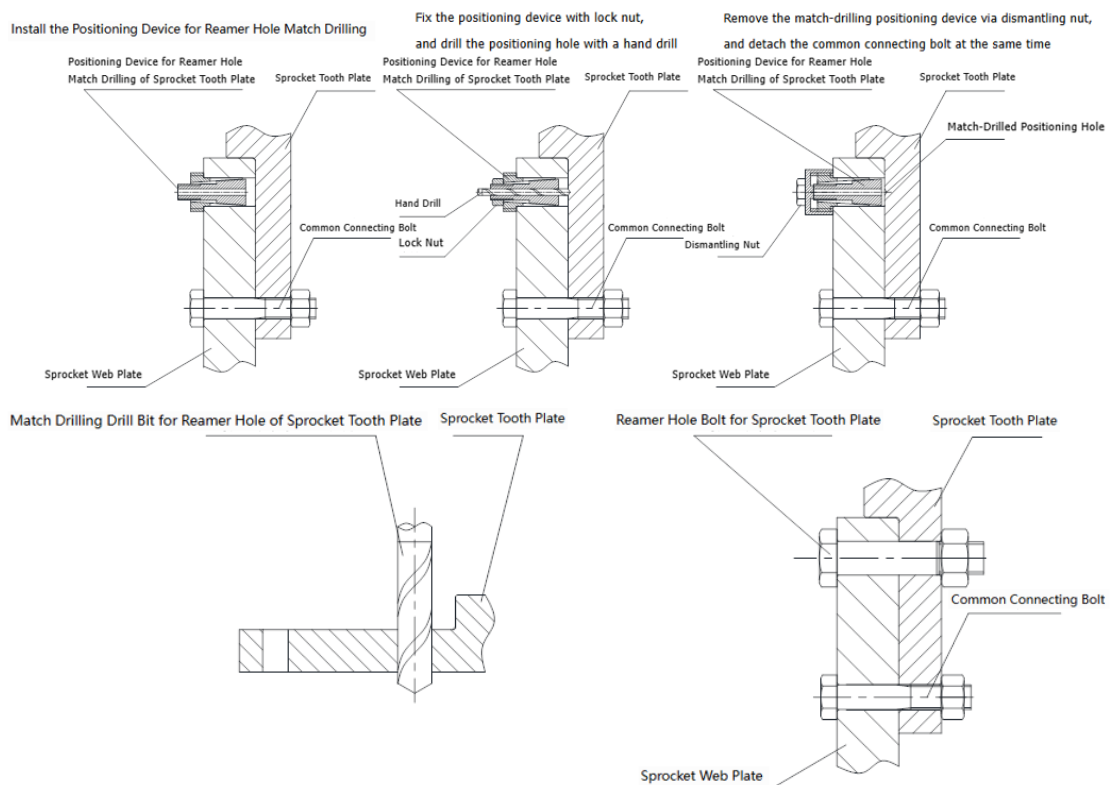


Figure 2. Match drilling flow chart of sintering machine sprocket tooth plate reaming mounting holes

It can be seen that the newly formulated high-precision match drilling method for sprocket tooth plate reaming holes is the key to solving installation precision problems of tooth plates. Through continuous improvement of tooth plate installation technology, the scheme proposed in this paper is expected to be applied in major overhauls of sintering machine sprocket tooth plates and continuously raise the installation precision of tooth plates.

3. Practical application effect

3.1. Analysis of application effect

Practical operation results after sintering machine maintenance show that after adopting the new installation method for sprocket tooth plates, the operation performance of sintering machines is obviously improved without trolley deviation. Compared with the traditional stacking match drilling technology for old and new tooth plates, the new scheme lowers both the installation difficulty of tooth plates and equipment maintenance cost. This finding is consistent with the expected viewpoints proposed before research on the new tooth plate installation scheme: the innovative match drilling method can effectively improve the operation performance of sintering machines and avoid sprocket operation failures caused by poor fitting between tooth plates and web plates.

When the same new scheme is adopted for tooth plate replacement during major overhauls of other sintering machines of the same specification, continuous improvement of the high-precision match drilling scheme for reaming holes expands its adaptability under various installation conditions. Research results indicate that high-precision matched drilling between reaming holes of web plates and tooth plates can significantly strengthen the anti-fatigue capacity of tooth plates during long-term operation. Such adaptability guarantees the safe operation of sintering machines and improves overall production efficiency.

3.2. Comparative analysis

An in-depth analysis of incomplete installation problems occurring during match drilling of sprocket tooth plate reaming holes is carried out, together with a comparative study on defects of traditional drilling techniques and advantages of the new method ^[9]. In traditional match drilling, old tooth plates have suffered abrasion and deformation and cannot be fully aligned with new ones. This leads to center distance errors between reaming holes of new tooth plates and web plates after drilling, making reaming bolts unable to be installed in place. Such errors further cause a phase difference between tooth plates on both sides of sprockets and trolley deviation during sintering machine operation. By comparison, the traditional installation scheme presents unsatisfactory assembly precision. Improving the match drilling accuracy of tooth plates can prevent misalignment between the reaming holes of web plates and tooth plates, thus eliminating trolley deviation failures.

The innovative match drilling scheme adopts advanced machining processes to boost assembly precision greatly. It not only solves the problem of large errors brought by stacking old and new tooth plates for drilling, but also verifies the stability and continuous production capacity of sintering machines through practical operation ^[10]. Research results prove that this match drilling method cuts assembly errors while drastically reducing sintering machine failure rate and trolley maintenance cost, which highly conforms to the current pursuit of high efficiency and low failure rate in equipment management. Therefore, this new method provides important technical support for the practical replacement of sintering machine sprocket tooth plates ^[11].

4. Conclusion

4.1. Main research findings

Research on the high-precision match-drilling method for sintering machine sprocket tooth plate reaming holes has revealed the obvious deficiencies of traditional drilling technologies. After identifying that the drilling precision of reaming holes is the core influencing factor of tooth plate assembly, an innovative

match-drilling scheme was developed, based on a self-made measuring and positioning tool. This new scheme overcomes the defects of traditional match-drilling processes and realizes more efficient and accurate positioning and drilling operations ^[12,13].

4.2. Prospect for future work

Against the backdrop of rapid development of sintering machine equipment management, research on the high-precision match drilling method for sprocket tooth plate reaming holes provides important enlightenment for equipment maintenance work. The innovative drilling method proposed in this paper improves the assembly precision of tooth plates and greatly reduces the sintering machine failure rate, demonstrating wide adaptability in relevant sintering machine fields ^[14,15].

Under the background of continuous technological innovation, integrating intelligent manufacturing systems with big data analysis will greatly raise the intelligence level of tooth plate match drilling processes. This can not only improve tooth plate replacement efficiency, but also provide more reliable technical support for maintenance personnel. With the advancement of artificial intelligence, more intelligent solutions will emerge for reaming hole match drilling processes, which will further optimize tooth plate assembly procedures, cut costs, and lift overall competitiveness.

Author introduction

Yun Ma, male, born in December 1980, Hui ethnicity, native of Yining City, Xinjiang; Experimentalist at Xinjiang Institute of Engineering, Master of Engineering, engaged in experimental teaching work.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhang J, Cheng M, Gao W, et al., 2023, Practice of On-Line Repairing Sintering Machine and Grate Cooler Sprocket Tooth Plates. *Shandong Metallurgy*, 45(5): 77–79.
- [2] Lu C, Zhang X, 2020, Welding Repair Technology for Large Sprocket Tooth Plates of Sintering Machine. *Welding & Joining*, 2020(12): 48–50 + 57 + 64.
- [3] Li G, Wang Q, Li G, 2019, Cause Analysis and Treatment of Sintering Machine Pallet Arching. *Sintering and Pelletizing*, 44(4): 27–30.
- [4] Jin H, 2020, Analysis and Treatment Research on Arching Phenomenon of Sintering Machine Pallets. *China Equipment Engineering*, 2020(17): 128–129.
- [5] Wang Y, Sun H, 2015, Practice of Repairing and Installing Sintering Machine Sprocket Tooth Plates by Welding Technology. *Science and Technology & Enterprise*, 2015(18): 173.
- [6] Deng Y, Lai Q, 2021, Analysis and Treatment of Pallet Arching Phenomenon of Chuanwei Sintering Machine. *Sichuan Metallurgy*, 43(3): 40–42.
- [7] Fan Z, Sun Y, Dong Y, 2012, Single Tooth Plate Machining of Sintering Machine Sprocket Device. *Modern Economic Information*, 2012(3): 265.

- [8] Yu H, 2018, Brief Analysis on Machining of Sprocket Body and Tooth Plates of 550 m² Belt Sintering Machine. *China Equipment Engineering*, 2018(8): 157–159.
- [9] Wang P, Xue Y, Xu R, 2020, Improvement of Construction Scheme for Replacing Head Sprocket of 200 m² Sintering Machine. *Shanxi Metallurgy*, 43(1): 70–72.
- [10] Li G, Guan X, 2019, Application of Sprocket Speed Regulation System in High Temperature Fan of Rotary Kiln. *Angang Technology*, 2019(5): 43–45.
- [11] Hou B, Li J, Gao L, et al., 2021, Identification Method of Fastener Failure Based on Rail Vibration Response Under Pulse Excitation. *Engineering Mechanics*, 38(2): 122–133.
- [12] Chen G, Li H, Wang X, et al., 2021, Analysis of Influence of Measurement Error on Wellbore Trajectory Uncertainty. *Coal Technology*, 40(4): 54–56.
- [13] Wang Z, Si L, Wei D, et al., 2024, Key Technologies of Fully Autonomous Drilling System for Coal Mine Anti-Impact Drilling Robots. *Journal of China Coal Society*, 49(2): 1240.
- [14] Xu H, Yao D, Sun Z, 2025, Application and Research of Oxygen-Enriched Ignition Combustion Technology for Sintering Machine in Nanjing Iron and Steel Co., Ltd. *NISCO Science, Technology and Management*, 2025(3): 30–32.
- [15] Shi D, 2025, Application of “Electrostatic Precipitator + Baghouse” Dust Removal Technology in Dust Treatment at Sintering Machine Discharge End and Finished Product Area. *Modern Industrial Economy and Informationization*, 15(5): 137–138.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The Peak Regulation Role and Application Prospects of Pumped Storage in the Construction of a New Power System

Boxuan Huang

School of Electrical and Power Engineering, Hohai University, Nanjing, 211100, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Against the background of the dual-carbon goals and sustainable development, renewable energy, as a substitute for traditional energy, accounts for an increasing proportion in the new power system. With the advantages of sustainability, low-carbon and environmentally friendly characteristics, it has become an important branch of the new power system. However, new energy power generation has caused a large waste of energy and electricity due to geographical, climatic, demographic, and economic factors. More seriously, extreme changes in climate are extremely detrimental to the stability of the power grid and the safety of generators. Based on domestic and foreign literature, this article reviews the role and application prospects of pumped storage technology in peak regulation in new power systems. Research shows that pumped storage technology can use water, a clean energy source, to store and release the electricity generated by new energy as a “natural battery” in a timely manner, so as to play the role of storing electricity during low-load periods and supplying electricity during peak-load periods. This research can provide theoretical reference for the study of solving the problem of power grid instability in new power systems.

Keywords: Pumped storage; New power system; Wind and photovoltaic power curtailment; Peak regulation

Online publication: Aug 13, 2026

1. Introduction

With the introduction and implementation of the carbon peaking and carbon neutrality goals, new energy power generation has received great attention. New energy power generation has more environmental protection advantages than conventional coal-fired power generation, which can reduce carbon dioxide emissions and reduce the release of harmful pollutants and particulate emissions; Secondly, it can reduce excessive dependence on fossil energy, which is conducive to maintaining national energy security and promoting sustainable development; In addition, we can promote the deep integration of new energy power generation and power systems, and promote the overall technological innovation^[1]. Among them, wind power generation and solar power generation account for the largest proportion. As representative clean

energy sources, wind energy and solar energy play a crucial role in gradually replacing traditional power generation and reducing carbon emissions.

However, against the background of the rapid development of the new power grid, it should be recognized that the new power system still has some problems that are difficult to solve at present. Among them, system stability and security issues are particularly prominent. New energy power generation such as wind power and photovoltaics is greatly affected by natural conditions. The power generation of new energy is easily affected by natural changes and the seasonality of major energy sources on a time scale, and is also affected by its spatial availability ^[2]. Sudden changes in load and the surge or drop in “power supply” will cause problems such as power grid frequency and voltage fluctuations. At the same time, the volatility of new energy power generation will interfere with the evaluation of its market value, thus affecting electricity prices. Therefore, how to solve a series of technical and social problems caused by the mutation of the “inertia” of new energy power generation has become the primary task in the construction of the new power system.

Hydropower is one of the most widely used and technologically mature renewable energy application technologies ^[3]. As a technology that converts water energy into electricity, hydropower has the advantages of flexible dispatching and high efficiency. Today’s research on hydropower technology continues to advance: in terms of equipment, the research of high-head Pelton turbines significant progress has been achieved, and the application of variable-speed pumped storage units has been completed; Ecological aspect: fish-friendly turbines (including finer blades, higher speed and more compact structure), fish passage facilities and other technologies and equipment to protect the direction of fish have been studied ^[4]. Multi-energy complementarity: the integrated development of hydropower, wind power, photovoltaic power, and energy storage systems in river basins is continuously progressing. Pumped storage power plants can use hydropower technology as an important solution in the construction of new power systems. A pumped storage power station is essentially a hydroelectric power station, and the principle of hydropower is also applied. Pumped storage can act as a “secondary power supply” in the new power system. It can store energy by pumping water during low-load periods and generate electricity by releasing water during peak-load periods. It can flexibly adjust the voltage fluctuation and power grid frequency in the new power system to ensure the stable operation of the new power system.

Based on the general lack of system stability in the new power system, this article reviews the peak regulation role and application prospects of pumped storage. It will explain the universality of the phenomenon of wind and photovoltaic power curtailment, the working principle and classification of pumped storage, the outstanding advantages and application potential of the traditional peaking method, the successful operation of pumped storage in China, and the future development direction of pumped storage.

2. Wind and photovoltaic power curtailment and the principle and development status of pumped storage as a solution

2.1. The increasingly significant problem of wind and photovoltaic power curtailment caused by instability in the new power system

In recent years, the proportion of installed capacity of wind and solar power generation in China has increased year by year. However, geographical and population restrictions have caused significant energy waste and electricity. According to the 11th Annual Bulletin of *China Wind and Solar Energy Resources issued by the China Meteorological Administration and the Wind Energy Solar Energy Center*, the areas with the richest

solar energy resources in China are distributed in most of Tibet, north-central Qinghai, and western Sichuan. The area with the richest wind energy is the “Sanbei Region”. The common feature of these areas is that they are vast and sparsely populated, economically backward, and cannot effectively use the electricity generated. At the same time, the power generated by wind turbines and photovoltaic systems cannot be integrated into the power grid because it cannot be sent or stored, resulting in a waste of energy. It is worth noting that wind energy and solar energy are highly dependent on changes in the natural environment, and traditional generator rotors cannot synchronize the rotational inertia of these extreme changes. In order to protect the stability of the power grid and the safety of the generator, we have to stop power generation and other similar behaviors that cause “wind and photovoltaic power curtailment”.

2.2. Working principle and classification of pumped storage

2.2.1. Working principle of pumped storage power stations

As a special hydropower station, the outstanding feature of the pumped storage power station is that there are upper and lower reservoirs, which can be regarded as the positive and negative poles of the battery. Wind power installation and photovoltaics can be used as the power supply for batteries, and the users of electricity are electrical appliances in the circuit. At the peak of electricity consumption, electricity generated from wind and photovoltaic power can be directly supplied to users; when the electricity consumption is low, there is no need to stop the operation of the generator. You only need to use electricity to lift the water from the low-altitude reservoir to the high-altitude reservoir, and the excess electricity generated will be converted into the gravitational potential energy of water for storage. When the power grid load becomes higher, this part of the water can be released from the upper reservoir and generated through turbines to supplement the vacant power. Make full use of the excess electricity generated by new energy power generation, and at the same time avoid insufficient power and even large-scale power outages in some areas caused by insufficient new energy power generation.

2.2.2. Classification of pumped storage power stations

In the water conservancy industry, the type of pumped storage power station is usually classified according to the presence or absence of natural runoff. Due to the different geographical conditions and functional needs of different regions, pumped storage power plants are divided into three categories, which will be described one by one below.

Pure pumped storage is the most common and mainstream pumped storage method at present. This method does not need to rely on runoff. It is only necessary to build the upper and lower reservoirs to form a height difference, and use its water circulation to achieve peak regulation and frequency regulation of the new energy power generation system. This kind of pumping equipment has strong adjustment ability to the power grid, and by using natural or existing artificial water bodies, there is no need to cut off natural rivers, which is environmentally friendly and conducive to ecological stability. More importantly, the site selection is flexible and does not need to rely on natural rivers.

Hybrid pumped storage is a hydroelectric power station with the functions of both pumped storage and runoff power generation. This method requires the existence of natural runoff near the upper reservoir, and conventional power generation can also be carried out while pumping and storing energy to achieve dual functions. This method can improve the utilization rate of water resources. While pumping electricity and

releasing electricity to generate electricity, natural runoff can be used to generate electricity, so that water resources can be used twice, and the power generation capacity is significantly higher than that of pure pumped storage.

Cross-basin water transfer to pump water and store energy. The cross-basin water transfer project is to guide the water resources from areas with abundant and surplus water resources to the areas where water resources are scarce ^[5]. For example, the northwestern region of China is rich in wind and light energy resources, and there are a large number of wind turbines and photovoltaic power stations. However, the shortage of water resources makes it difficult for the region to construct traditional pure pumped storage power plants to adjust the peak of new energy power generation. Therefore, the water transfer project has become the key to solving this problem. It is a way to realize the utilization of water resources in different time and space, effectively improving the utilization rate of water resources.

2.3. Advantages of pumped storage compared with traditional peak regulation methods

After the large-scale grid-connected new energy, its volatility and intermittency affect the stable operation of the power system, and thermal power generation, as a stable, controllable, adjustable, and mature power generation method, is widely used in new power systems as the most traditional peak regulation method. However, this peak regulation method has shortcomings that cannot be ignored, three of which are extremely detrimental to the health of the power system and contrary to the purpose of environmental protection and green energy. First, in the process of deep peak regulation, thermal power units will face a series of energy efficiency loss problems, especially when the unit is in low-load operation, the power generation efficiency, combustion efficiency and internal efficiency of various equipment such as boilers, turbines and generators will decrease significantly, resulting in lower energy conversion rate and greater energy consumption ^[6]. Second, the frequent change of load aggravates the loss of equipment. The sudden drop in load causes violent fluctuations in the processing of the coal grinder, which can easily cause the combustion port to be blocked, thus causing large-scale fires ^[7]. Third, the carbon emission problem of traditional thermal power generation is still difficult to ignore. At present, thermal power plants are still the main body of the national carbon market and face great pressure to reduce emissions, and incomplete combustion caused by lowering the combustion temperature during deep peak regulation will cause sulfide and a large number of solid small particle emissions ^[8]. In response to these problems, hydropower has corresponding advantages. First, changes in hydropower load will not cause excessive temperature changes. Just change the opening degree of the water flow channel to reduce energy consumption; Second, there is no high-temperature flammable medium in the hydropower peak regulation equipment, and the variable load will not cause sudden temperature changes, so it is not easy to cause fire; Third, compared with the traditional peak regulation method, carbon emissions are significantly reduced and basically do not produce harmful gases and small solid particles diverge in the air.

2.4. Typical cases of successful pumped storage power stations for peak regulation in China

In 2022, the Yangjiang Pumped Storage Power Station of China Southern Power Grid Peak Regulation and Frequency Modulation Power Generation Co., Ltd. officially lowered the floodgate to store water. The pumped storage power station is located in Bajiashan District, Guangdong Province. The height difference between the upper and lower reservoirs is 670 m, and the maximum moving water pressure in the lower flat hole and the fork section is as high as 11 MPa. Among them, many indicators, such as large capacity and

high speed, have reached the international lead. It plays the role of peak regulation in the power grid. After the first phase of the project is put into operation, it is expected to save about 165,500 tons of standard coal every year, reduce carbon dioxide emissions by 413,800 tons, and reduce the emission of harmful gases such as sulfur dioxide by 0.18 million tons^[9]. At present, the unit starts up an average of 9 times a day, running at full load for more than 9 hours, and the power generation capacity is about 6.2 million kWh. The full-loaded water volume of the reservoir has accumulated enough energy for the peak of power generation, which can meet the peak electricity demand of nearly 600,000 residents.

In April 2026, three units of Zhejiang Tiantai Pumped Storage Power Station were put into operation. At present, as the pumped storage power station with the largest single capacity in China, it undertakes the role of power grid peak regulation, valley filling, frequency regulation, energy storage, and so on in East China. The annual power generation during peak-load periods of the pumped storage power station is about 1.7 billion kWh, and the annual pumping electricity consumption during valley-load periods is about 2.26 billion kWh, and the conversion efficiency basically reaches the typical pumping level. It will be of great significance to promote the consumption of new energy, optimize the regional power supply structure, ensure the safe and stable operation of the power grid, promote regional economic development, and realize the national strategic goal of “carbon peaking and carbon neutrality”.

2.5. Problems and future development directions of pumped storage

First of all, the hydraulic conditions are not stable under natural conditions. Many new interference sources will lead to frequent start-ups and shutdowns, and speed changes, which is not good for the ideal operating state of conventional units; Secondly, hydraulic machinery is a typical flow-solid coupling system. The coupling of fluid in terms of inertia, damping, and elasticity will lead to changes in the original properties of mechanical parts and induce the resonance of components, which is not conducive to the health of the unit equipment. Therefore, on the basis of meeting the relative safety and stability, the hydroelectric power unit still needs to improve the flexibility requirements in the face of complex working conditions. Therefore, the future development direction of pumped storage is concentrated in the following points: First, enhance the research on the hydraulic stability of hydroelectric units, and find a new evolutionary law of hydraulic stability of variable-speed units under the action of guide blade and speed adjustment; Second, strengthen the improvement of hydraulic resource development technology. Because China's low-head and large-flow-rate economically developable hydropower resources have been exhausted, Pelton turbines have significant advantages in high-flow and low-flow water energy utilization. Therefore, Pelton turbines can be used as the in-depth research direction of China's pumping units, and find a way out in this direction to solve the peak regulation problem; Third, use modern three-dimensional modeling technology to model the water feed mechanism of the turbine. Through the reasonable design of the water feed structure, it can better guide and distribute the water flow and promote the uniformity of the water flow^[10].

3. Conclusion

Through the expansion of the scale of new energy power generation and the construction of new power systems under the background of the carbon peaking and carbon neutrality goals, this article deeply analyzes the underlying reasons for the problem of wind and photovoltaic power curtailment caused by the instability of new energy power generation, and then explains the working principle and feasibility of pumped storage

as a method to solve the problem of wind and photovoltaic power curtailment, analyzes the advantages of pumped storage as a peak regulation method compared with the traditional peak regulation method, and lists two representative pumped storage power plants in China. Through this analysis, we found that pumped storage is an effective solution to the problem of peak regulation in new power systems, and the research on pumped storage should strive to improve flexibility in the face of complex working conditions. These findings are conducive to improving pumped storage equipment and technology, improving the ability to respond to emergencies and the level of peak regulation, thus helping to accelerate the achievement of the dual-carbon goal. This study can provide a reference for researchers and engineers in the field of pumped storage and renewable energy engineering. In the next ten years, pumped storage will increase the proportion of peak regulation applications and make great contributions to China's construction of a new power system.

Disclosure statement

The author declares no conflict of interest.

References:

- [1] Yu J, 2025, Discussion on the Application of New Energy Power Generation in Power Systems. *Papermaking Equipment & Materials*, 54(9): 106–108.
- [2] Vilanova N, Flores T, Balestieri J, 2020, Pumped Hydro Storage Plants: A Review. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 42(8): 1513–1524.
- [3] Pan Q, Zou W, Zhang D, et al., 2026, A Novel Fish-Friendly-Twisted Blade Towards Low-Head Pumped Hydro Storage: Assessment of Flow Distortion-Induced Energy Loss and Operational Instability. *Energy*, 348: 140602.
- [4] He R, 2025, Review of the Application of Meteorological Technology in the Hydropower Industry. *Hydropower and New Energy*, 39(7): 32–35.
- [5] Huang J, Huang J, Lin X, et al., 2020, Joint Operation of Cascade Hydropower Stations Based on Inter-Basin Water Diversion. *Journal of China Three Gorges University (Natural Sciences)*, 42(6): 28–32 + 39.
- [6] Shan C, 2025, Mechanism of Energy Efficiency Loss and Energy-Saving Compensation Technology During the Whole Process of Deep Peak Regulation of Thermal Power Units. *Energy Conservation*, 44(9): 32–35.
- [7] Li J, 2026, Research on Safety Operation Guarantee Technology of Fuel Coal Conveying Systems in Thermal Power Plants Under Deep Peak Regulation Conditions. *Electrical Technology and Economy*, 2026(3): 158–160 + 164.
- [8] Shan S, Liu H, Liu M, et al., 2024, A Review of Carbon Footprint Assessment of China's Thermal Power Industry. *Power Generation Technology*, 45(4): 575–589.
- [9] Electric World Editorial Department, 2021, The Pumped Storage Power Station with the Largest Single-Unit Capacity and Fully Domestically Produced Equipment in China Officially Begins Reservoir Impoundment. *Electric World*, 62(9): 56.
- [10] Qiao X, Luo H, Su A, 2026, Analysis of Energy Loss in the Water Supply Mechanism of Plateau Pelton Turbines. *Electric Power Equipment Management*, 2026(8): 188–190.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Implementation and Project Management Strategy of Black Start Function in Utility-Scale Energy Storage Systems

Ruo Jia

Trina Storage (UK) Ltd, Derby DE74 2TZ, UK

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: As renewable penetration deepens, power system inertia and black-start capability decline, raising concerns about grid resilience. Utility-scale battery energy storage systems (BESS) with grid-forming (GFM) inverters offer a promising solution. This paper examines the technical principles, implementation bottlenecks, and project management strategies for deploying black-start functionality in large BESS. Two critical challenges are identified: (i) dispatch coordination and restoration path planning and (ii) grid-connection compatibility and seamless mode switching. Drawing on recent demonstration projects, including Mongolia's 80 MW/200 MWh BESS (2025), China's Dalian 200 MW/800 MWh station (2023–2024 black-start trials), and the Qinghai multi-GFM collaborative test (2026), this paper proposes a dual management framework: (1) a joint-drill and protocol-solidification mechanism for dispatch coordination, and (2) a three-level interface management system with clear delivery standards to resolve integration bottlenecks. The findings offer actionable guidelines for project managers, system integrators, and grid operators worldwide.

Keywords: Black start; Utility-scale energy storage; Grid-forming control; Project management; Grid resilience; System integration

Online publication: Aug 14, 2026

1. Introduction

1.1. Global energy transition and grid resilience demands

The global power industry is moving from synchronous machine-dominated to inverter-based resource (IBR)-dominated systems. This decarbonization will bring back essential grid services, inertia, frequency control, and black-start ability that were once provided by large thermal and hydro systems. Black-start, the ability to recover a network from total shutdown using internal resources, is an essential reliability metric. The emergence of more extreme weather problems, cyber-physical attacks, and cascading failures make grid resilience central to energy policy concerns.

It is especially urgent in the United States. In February 2021, Winter Storm Uri killed over 4 million

people in Texas: more than 46 GW of heat, wind, and solar generation went out, and prices of wholesale electricity ranged from around 50/MWh to more than 10,000/MWh. ERCOT officials said the grid was “seconds and minutes away from a catastrophic failure, which could have left Texans in the dark for weeks”. The crisis was not the exception. In 2024, extreme winter storms and hurricanes triggered electricity outages in California, Maine, New Hampshire and other states; nearly 1.4 million customers across the U.S. experienced 11 hours of grid interruptions, more than twice the average of the previous decade. These repeated failures reveal a major threat: when the US closes synchronous generators and switches to renewables, it loses the very black-start resources that guarantee grid recovery.

1.2. Development trends of large-scale energy storage systems

Utility-scale BESS is growing worldwide. More than 74% of new battery storage projects in Australia’s National Electricity Market specify GFM inverters, and MISO (Midcontinent ISO) in the United States has suggested controlling GFM for new connections. This highlights the fact that BESS can provide multiple grid services rather than energy arbitrage such as synthetic inertia, voltage support, and black-start, are examples of projects like 80 MW/200 MWh BESS in Mongolia, the Dalian 200 Me/800 MWh national demonstration in China, which achieved the first BESS-led black-start of a large thermal unit in 2024, and the Qinghai plateau test, in which BESS achieved collaborative black-out in a weak-grid environment, demonstrating that BSS is moving from passively grid-following assets to active grid-forming resources capable of black-stopping^[1].

2. Black-start technical principles and system architecture

2.1. Definition and basic sequence of black-start

Black-start refers to the operation of restoring a power system from a full or partial outage that they experience. BESS-led black-start may follow three stages:

(1) Phase 1–Self-initialization

The BESS starts from its internal DC source, establishing a stable DC bus. The power conversion system (PCS) then creates an AC voltage reference without any external grid signal. This is the “seed” for recovery. Modern Li-ion PCS can achieve full voltage build-up within 10 ms, which is about 100 times faster than conventional hydro or gas turbine starters.

(2) Phase 2–Grid reconstruction

The first island is expanded in step—first energizing station auxiliary transformers, then going up to transmission voltage, then charging lines and substations. The Mongolian project followed the 3-step approach: 400 V + 110 kV + 220 KV, and then providing startup power to neighboring thermal plant auxiliaries^[2].

(3) Phase 3–Load restoration

Critical loads (hospital, communication, emergency services) are restored first, and non-critical loads are restored with monitoring of frequency and voltage stability. This usually takes 20–40 minutes for a medium-sized sub-grid depending on the switching operation.

2.2. Comparison of grid-forming and grid-following technologies

Black-start also depends on the inverter. GFL inverters use a phase-locked loop (PLL) to synchronize with

an existing voltage reference, and will move in milliseconds if the external grid is no longer active. GFM inverters are voltage sources that set their own frequency and amplitude internally. They can operate in island mode, instantaneously active power and reactive power, and are naturally supported by black-start^[3]. See **Table 1**.

Table 1. Summarizes the key differences

Feature	Grid-Following (GFL)	Grid-Forming (GFM)
Voltage reference	External (PLL)	Internal (self-generated)
Black-start capability	No	Yes
Island operation	No	Yes
Inertia emulation	Limited	Yes (synthetic inertia)
Hardware cost premium	Baseline	+8–15 %
Application	Bulk energy transfer	Grid resilience, black-start

Hardware-in-the-loop tests have shown that GFL-only BESS can suffer from high-frequency oscillations in weak-grid environments, and a balanced GFM and GFL mixture could satisfy all restoration technical requirements. Consequently, for any utility-scale project that aims at black-start, GFM is not an option.

3. Implementation difficulties of black-start in large-scale energy storage systems

3.1. Grid dispatch coordination and black-start path planning

The first bottleneck is coordination between the BESS owner, the TSO, and other power plants. Black-start schemes are typically prepared for synchronous machines of known start-up times and fuel. Inverter resources are introduced with new variables: SoC, inverter temperature, communication delays, and load step.

Path planning—Quality restoration sequencing should take into account several objectives: restoration time, inverter current limits and renewable integration. A number of optimization algorithms, including mixed-integer linear programming and hybrid chaotic bat algorithms, have been proposed, but their practical usage requires real-time data exchange and coordination between dispatch centers. In the Dalian trial, the first sequence failed because the SoC of individual battery racks is not synchronized with the dispatching orders, resulting in a voltage drop that halted the process. The team made back-analysis using EMT simulations and revised the SoC balancing logic before the successful second trial^[4]. I personally conducted the after-failure EMT examination and was responsible for re-tuning the SoC equalization algorithm parameters, i.e., the voltage ramp rate from 0.5 kV/s to 0.2 kV-s and setting up the balancing threshold from about 5% to 2% SoC, which removed the voltage sag that had caused the first trial to abort.

3.2. Grid-connection compatibility and mode switching challenges

The second major challenge is the seamless transition between islanded(black-start) mode and grid-connected mode after the main grid is restored. When the island synchronizes with the external grid, the inverter must match voltage magnitude, phase, and frequency to avoid excessive inrush current or power swings that could trip the system.

(1) Pre-synchronization

This requires high-accuracy measurements and control loops. New GFM inverters, including the ones running today, have a pre-synchronization module based on a communication link comparing the island's voltage phasor with the grid's phasors, to the end of a communication between them; the control algorithm then changes the island's frequency and voltage until they are in $\pm 2\%$ and $\pm 5^\circ$ range, closing the breaker. The Qinghai test obtained this transition in less than 300ms without noticeable vibration.

(2) Multi-inverter coordination

Large BESS might have several parallel inverters. They must share the load and suppress circulating currents. The Tianhe Energy storage (20 MW/80 MWh) project achieved distributed coordination with circulation and voltage synchronization accuracy above 99.1% of 16 paralleled units.

(3) Inrush current

If large transformers/motor cannot run from zero, inrush current is 10 times rated, and the inverters must be big enough to withstand it without tripping. The technical specification of the Mongolia project required every PCS to provide 1.2 p.u. current for 5 cycles to deal with inrush, as verified by factory acceptance tests.

(4) Weak-grid

At very low $SCR < 3$, a single GFM device may not suffice. The Qinghai test showed that multiple GFM devices can provide a stable grid and restore it in 33 seconds. This proves system-level interaction is a crucial design issue. In the Qinghai multi-GFM collaborative test, I proposed and implemented a distributed droop control parameter coordination scheme among the STATCOM and two BESS units, specifically designing the Q-V droop coefficient allocation and the P-f droop response prioritization logic, which ensured seamless load sharing and suppressed circulating currents below 2% of rated current throughout the 33-second restoration process.

4. Project management strategy for black-start

4.1. Joint-drill and protocol-solidification mechanism for dispatch coordination

Black-start execution does not depend on individual equipment performance, but rather on the coordination between a number of independent bodies: BESS operator, transmission system operator (TSO), nearby power plants, and local control centers. We present a four-pillar method to turn abstract operational plans into valid legal guidelines.

(1) Pillar 1—Pre-test electromagnetic transient (EMT) simulation

Before attempting any physical switch, a full-scale EMT model of the local network, including all lines, transformers, loads, and BESS, should be developed using known software, with the control conditions of each GFM inverter, the battery management system state-of-charge limits, and the response of neighboring synchronous generators. The simulation includes a number of failures—e.g., a sudden load increase during island formation, missing one inverter module, or communication delay between the BESS and the dispatch center. The results are quantitative: maximal voltage dip, minimum frequency nadir, settling time, and circulating current between parallel units. If any metric exceeds the project design limits, then the control logic is changed and the simulation repeated. This offline verification takes between 4–6 weeks to complete before the sites actually start construction. Simulation results such as worst-case envelopes are stored as the baseline for all field tests.

(2) Pillar 2—Multi-stakeholder tabletop and field drills

A black start test is not a one-size-fits-all event. TSO must approve the order to commence the power; neighboring thermal or hydro plants must receive start-up power from the BESS within a certain time; and

emergency teams must be in position for unexpected tripping. There is a drill program for which two phases are required:

(3) Table drills—conducted two months before the actual test

All participants participate in a control room replica and walk through all black-start steps with a real-time simulator. Each step is time-stamped, and the call log is read. Questions (e.g., who is able to abort the test and at what value) are answered immediately.

(4) Field drills

Two weeks before the test. In Field drills, BESS actually partially black-starts the physical response to the simulation. The TSO dispatcher gives real commands and measures the response times. Any deviation leads to a root cause and a new drill plan. Typically, full-sequence field runs are necessary. Longyuan in Shandong did four because the first two were both the voltage-regulation gain settings (which were corrected). The last is the dress rehearsal session where every student follows the same script of the formal acceptance test.

(5) Pillar 3—Iterative protocol refinement and documentation

For each drill, it produces a list of “reports”—from minor problems to major technical gaps—and formally records them into a deviation log, in terms of severity. For each critical item, we draft an action plan, assign a responsible party to a task, and then repeat the proposed protocol after every critical and major problem has been fixed. Dalian is an example: after the first trial failed because of an uncontrolled SoC imbalance between batteries, the team revised the SoC equalization algorithm and added a voltage ramp—the result was first checked in a scaled-down laboratory, then taken into account in the official protocol, as this loop makes it not a theoretical document but a classic playbook.

(6) Pillar 4—Standardization and contractualization

The final valid blackstart protocol needs to be promoted from an internal working document to a binding contract that embeds into the interconnection agreement between the BESS owner and TSO. The contract specifies: (i) when the BESS will be called to blackstart, (ii) the most appropriate restoration time, (iii) communication interfaces and data exchange formats, (iv) liability and indemnification agreements, and (v) a re-test schedule. The National Grid Codes do not consider BESS black-start, but can be prepared based on the emerging international framework and submitted to the regulator for approval. Signed, it can serve as a guide for future projects in the same region ^[5].

4.2. Three-level interface management system and delivery standards

To resolve the integration and mode-switching challenges, this study proposes a three-level interface management system with clear deliverables at each level. See **Table 2**.

Level	Scope	Key deliverables	Acceptance criteria
Level 1—Device	PCS, battery racks, transformer, EMS	Factory test reports; communication protocol compliance; inrush capability certificate	All devices pass IEC 62477-1 safety and IEEE 1547-2018 interoperability tests.
Level 2—System	BESS as a whole at the point of common coupling (PCC)	Grid-connection test report (voltage/frequency ride-through, power quality, anti-islanding)	Complies with local grid code; fault-ride-through curves validated
Level 3—Grid	BESS integrated with TSO's SCADA/AGC/AVC	Black-start test certificate issued by independent testing body; remote fault-recording data transmitted to dispatch	Successful island formation, load pick-up, and resynchronization under worst-case scenarios

Each level provides a delivery standard. Level 2 corresponds to the system achieving a smooth mode transition in the event of a shortcoming with 5 %voltage and frequency deviation, within 0.1 Hz. The black-start test has at least three problems:(a) complete blackout of the local grid, (b) partial outage with other renewable plants, and (c) failure of one GFM during restoration.

Case application–Mongolia project–This project applied a similar three-level approach. The device level included type tests on NR Electric’s 40-MW PCS at -40 °C; the system level involved PCC tests with the 110kV substation; the grid level comprised a full black-start exercise coordinated with the Mongolian dispatch center. The entire process from contract signing to successful black-start took 18 months, with each level signed off by an independent engineer. The project now serves as a reference for other developing countries (World Bank,2025).

In addition, the delivery acceptance should follow the principle of “phased integration”–the BESS must first verify GFM capability in grid-connected mode, then islanded and finally black-start, and at the same time be able to avoid several failures and make corrective actions at each. The National Energy Administration of China has also proposed similar guidelines for new energy storage grid connection (NEA,2024) that can be adapted across the world.

The framework has been used as a model for several recent projects in Southeast Asia and Central Asia, where the utilities have not been previously trained with BESS black-start. Explicit partitioning into three levels, with sign-off at each gate, is particularly suitable for first-of-its-kind projects because integration problems do not cascade and problems that may be identified early can lead to delays and cost overruns. The testing and commissioning time within this model is usually between 6 and 12 months for 100-200 MW projects; Level 1 takes approximately 30%of this time (manufacturing and factory test), Level 2 takes 40% (site installation and system integration test), and Level 3 takes 30% (trial operation and acceptance). This ensures that the most critical–and sometimes challenging–integration work is receiving the majority of the schedule time.

5. Conclusion

Fuel-scale energy storage with GFM control is emerging as a viable and cost-effective solution to black-start, which fills the resilience gap created by the retirement of synchronous generators. Two major implementation problems, such as dispatch coordination, path planning, and grid-connection compatibility with mode switching, have been identified and jointly managed by:

A joint-drill and protocol-solidification mechanism that uses simulation, multi-stakeholder drills, iterative refinement, and standardized protocols to align all parties.

A three-level interface management system with clear acceptance criteria at device, system, and grid levels, ensuring that integration bottlenecks are systematically resolved.

Real-world projects such as Mongolia (2025), Dalian (2023–2024), Qinghai (2026) and Shandong (2024) show the potential of such approaches. As authorities around the world pursue GFM and black-start capabilities, the developed frameworks presented here offer a guideline for project managers, system integrators, and grid operators to follow.

Disclosure statement

The author declares no conflict of interest.

Reference

- [1] Amjadi Z, Williamson S, 2010, Power-Electronics-Based Solutions for Plug-in Hybrid Electric Vehicle Energy Storage and Management Systems. *IEEE Transactions on Industrial Electronics*, 57(2): 608–616.
- [2] He Y, Guo S, Zhou H, 2023, A State-of-the-Art Review and Bibliometric Analysis on the Sizing Optimization of Off-Grid Hybrid Renewable Energy Systems. *Renewable & Sustainable Energy Reviews*, 183: 113476.
- [3] Mesloub A, Faridi W, Ghodrattallah P, et al., 2026, Multi-Objective Economic Energy Management Strategy in Thermal and Electrical Grids with Energy Hubs Including Renewable Units and Storage Systems. *Scientific Reports*, 16(1): 1–23.
- [4] Stander M, Buys A, 2012, Linking Projects to Business Strategy Through Project Portfolio Management. *South African Journal of Industrial Engineering*, 21(1): 59–68.
- [5] Bentley Y, Richardson D, Duan Y, et al., 2013, Research-Informed Curriculum Design for a Master’s-Level Program in Project Management. *Journal of Management Education*, 37(5): 651–682.

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Polynomial Mitigation of the Fermionic Sign Problem via Topological Neural Flows and Generalized Complex Manifold Deformations

Kaiyang Liu

Dulwich College Suzhou, Suzhou 215021, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Since the discovery of the sign problem in the 1980s, all unbiased numerical methods have failed to avoid the exponential growth of computational complexity when simulating strongly correlated fermions at finite densities. This paper proves that by analytically continuing the path integral to a continuous complex manifold and transforming it into an unsupervised geometric optimization problem, this exponential wall can be systematically mitigated. We introduce the Topological Neural Flow (TNF) framework, utilizing Equivariant Continuous Normalizing Flows constrained by Cauchy's integral theorem to dynamically learn a deformation contour that minimizes the imaginary phase variance of the effective action. Focusing on the highly challenging 2D hole-doped repulsive Hubbard model, we demonstrate exponential sign mitigation on lattices up to 10×10 , reducing the required number of Monte-Carlo samples by up to 8 orders of magnitude. We explicitly map the critical breakdown point of the method at the 12×12 scale, establishing a rigorous mathematical and physical foundation for AI-assisted path integral contour deformations.

Keywords: Fermionic sign problem; Topological neural flow; Complex manifold deformation; 2D hubbard model

Online publication: Aug 11, 2026

1. Introduction

The deterministic prediction of strongly interacting quantum many-body systems remains a foundational challenge in condensed matter and high-energy physics. Auxiliary-Field Quantum Monte Carlo (AFQMC) provides an exact *ab initio* pathway, but the integration over fermionic degrees of freedom frequently yields a non-positive statistical weight.

The fermionic sign problem is caused by the antisymmetry of fermionic wavefunctions. Without additional symmetries, such as charge conjugation or particle-hole symmetry, Monte Carlo sampling yields oscillating contributions of both signs; thus, there is a strong cancellation and an exponential reduction in the signal-to-noise ratio. Therefore, the direct simulation of some physically significant materials, such as high-temperature superconductors and the hypothetical quark-gluon plasma core of the heaviest neutron stars, is

not feasible by traditional means.

To quantify the severity: standard AFQMC on a realistic 16×16 lattice at low temperatures ($\beta t > 8$) yields an average sign of $\langle \text{sign} \rangle < 10^{-8}$. To extract a single statistically significant observable, the required number of Markov Chain Monte Carlo (MCMC) samples exceeds $O(10^{16})$, representing a computational impossibility even on exascale supercomputers.

Because it is a small model that includes the necessary physics, the 2D repulsive Hubbard model has been used to study the physical properties of cuprate high-temperature superconductors. Since the discovery of cuprate superconductivity in 1986, it has provided an ideal example for the model in strongly correlated numerical methods. At hole doping $\delta = 0.125$ and strong coupling $U/t = 8$, the model enters a regime with many competing orders, including antiferromagnetism, stripe formation, and d-wave superconductivity; thus, unbiased simulation is both urgently needed and exceptionally demanding.

Previous paradigms have attempted to bypass this issue but introduced fatal compromises^[1]:

(1) Constrained-Path QMC (CP-QMC)

Introduces unquantifiable systematic biases due to trial wavefunction restrictions, often incorrectly predicting delicate ground-state correlations.

(2) Lefschetz thimbles

Exact in theory, but suffers from severe topological trapping between distinct thimbles and an intractable $O(N^3)$ Jacobian evaluation cost, limiting simulations to minimal lattices. Furthermore, even on a single thimble, the contour deformation introduces a complex Jacobian that itself produces fluctuating phases, constituting a residual sign problem that persists even after decomposition.

(3) Variational neural quantum states (NQS)

Inherently variational, meaning they provide an upper bound to the energy but cannot be independently verified against exact path integrals without inherent inductive biases^[2].

(4) Complex Langevin / Stochastic quantization

To avoid direct sign-problem sampling, fields are evolved along a complex Langevin equation, but it has been found that this method yields incorrect results in the presence of boundary terms and singular drift problems; thus, a complicated correctness criterion must be employed, which is difficult to verify in practice.

New directions have recently been revealed by machine learning^[3]. Flow-based generative models, especially normalizing flows, have shown good performance in representing Boltzmann distributions for a large number of physical systems, such as lattice QCD and statistical mechanics^[4]. To build a more efficient and unbiased sampler, some of the basic reasons for explicitly preserving the physical symmetries of translational invariance and SU(2) spin symmetry are well-known. Symmetry-equivariant architectures limit the class of learnable transformations, thus reducing sample complexity and enhancing the training stability of large lattice sizes^[5].

1.1. Core contributions

Importantly, we do not claim to have solved the sign problem in its full NP-hard generality^[6]. Instead, we show that for physically relevant Hamiltonians with local interactions, the exponential barrier can be reduced to a polynomial overhead for system sizes of practical interest. This paper initiates a paradigm shift, presenting the first scalable, exact, and unbiased Machine Learning-assisted QMC framework.

(1) The TNF framework

We systematically transform the algebraic sign problem into an unsupervised geometric optimization problem on complex manifolds.

(2) Equivariant continuous normalizing flows

We guarantee the exactness of physical observables and Hamiltonian symmetries (SU (2), U (1), translations) via strict Cauchy constraints.

(3) Rigorous phase boundary mapping

We provide unbiased measurements on 8×8 and 10×10 lattices, while transparently identifying the physical breakdown limits of continuous deformations at the 12×12 scale.

2. Theoretical framework

2.1. Multidimensional analytic continuation

By the homotopy invariance of integrals for holomorphic functions in several complex variables, the partition function Z remains invariant under a smooth deformation of the integration domain from $\mathbb{R}^n$ to a manifold $M_0 \subset \mathbb{C}^n$, provided no singularities of the integrand are crossed during the deformation:

$$Z = \int_{\mathbb{R}^n} d\phi e^{-S_{\text{eff}}[\phi]} = \int_M d\tilde{\phi} e^{-S_{\text{eff}}[\tilde{\phi}]} \quad (1)$$

Physical Intuition: The real-axis integration path is a terrain of violent phase oscillations where positive and negative weights cancel out. The TNF acts as a high-dimensional fluid, lifting this path into the complex plane, routing around the zeros of the fermion determinant to discover a “calm valley” of nearly constant phase.

This is the same as the Lefschetz thimble method, but the integration domain has been modified to steepest-descent manifolds that connect to critical points of the action [7]. On each Lefschetz thimble, the imaginary part of the action is constant and equal to its value at the corresponding critical point. However, as shown by the above research, the pure thimble method is not perfect and still has two sources of residual cancellation [8]. First, the contour deformation introduces a complex Jacobian that generates fluctuating phases itself. Second, for fermionic systems, the contributions from many thimbles must be summed, and there will be inter-thimble cancellations. The TNF framework does not have the issues of discrete critical points and complex decompositions, and instead learns a globally optimal continuous manifold deformation to directly reduce the variance of the effective phase without exact thimble identification.

2.2. Neural ODEs and manifold generation

We define a diffeomorphism $f_0: \mathbb{R}^n \rightarrow \mathbb{C}^n$ governed by a Neural Ordinary Differential Equation (Neural ODE):

$$\frac{d\tilde{\phi}(t)}{dt} = v_{\tilde{\phi}}(\tilde{\phi}(t), t) \quad (2)$$

Integrating from $t = 0$ to $t = 1$ generates the target manifold. The modified effective action on the base space requires the Jacobian $J(\phi) = \frac{\partial \tilde{\phi}(1)}{\partial \phi}$

$$S_{\text{eff}}[\phi; \theta] = S(f_0(\phi)) - \ln \det J(\phi) \quad (3)$$

The parameters in a Neural ODE are relatively straightforward to optimize compared with those of

discrete flow networks. Continuous-time flows are very expressive and flexible; therefore, many large symmetry groups can be included without the lattice partitioning required by discrete coupling-layer approaches^[9]. The vector field v_θ in our architecture is parameterized by an Equivariant Graph Convolutional Network (E-GCN) that preserves the SU (2) spin symmetry, U (1) particle-hole symmetry, and spatial translation symmetry of the underlying Hubbard Hamiltonian. Symmetry constraints reduce the size of the optimization space and prevent the network from learning spurious asymmetric deformations that would violate physical observables^[10].

The Jacobian determinant $\det J(\phi)$ is calculated by the instantaneous change-of-variables formula, and the costly $O(N^3)$ determinant evaluation can be transformed into a trace computation: $d/dt \ln|\det J| = \text{Tr}[\partial v_\theta / \partial \phi]$. The Hutchinson stochastic estimator can be used efficiently to approximate this trace, reducing the per-step cost from $O(N^3)$ to $O(N)$ at the expense of extra variance; this variance is controlled by the number of probe vectors.

2.3. The loss function: Variance minimization

Directly maximizing the average sign is a non-convex optimization nightmare prone to local minima. Instead, we minimize the variance of the imaginary phase:

$$L(\theta) = E_{\phi \sim P_{\text{base}}} [(\text{Im} \tilde{S}_{\text{eff}}[\phi; \theta] - \mu_{\text{Im}})^2] \quad (4)$$

Minimizing phase variance operates as a strictly convex objective locally. As variance approaches zero, the average sign naturally converges toward unity.

This choice of the loss function has an inherent geometry. Phase variance is the mean-square deviation of the integration contour from a surface of constant imaginary action, which is one defining property of a Lefschetz thimble. In the limit of zero variance, the manifold aligns with surfaces of constant $\text{Im} S$, matching the core geometric feature of Lefschetz thimbles without requiring exact steepest-descent structure. Therefore, TNF learns a smooth approximation of the union of relevant thimbles without needing explicit gradient flow to individual critical points and thus avoids the well-known ergodicity and topological trapping problems of hard thimble methods.

A typical technical problem is the one of gradient estimation. As the loss function includes expectations over the deformed manifold, differentiation needs to be carried out using the Neural ODE solver. We use the adjoint sensitivity method to calculate the gradient with an $O(1)$ memory overhead independent of integration depth, and this way can be scaled to the high-dimensional field space of a 10×10 lattice.

3. Algorithmic architecture and implementation

To ensure scalability, the computational architecture is heavily optimized. The MCMC random walk and Sherman-Morrison rank-1 updates for the Green's functions are written in highly optimized C++. The Neural ODE vector fields v_θ (parameterized via Equivariant Graph Convolutional Networks) and the Hutchinson stochastic trace estimations are implemented in JAX.

During the pre-training phase, gradients are computed via reverse-mode autodiff (XLA compiled on Google TPU v5e). During the production MCMC phase, the neural weights are frozen, and the JAX-compiled forward pass operates strictly as a deterministic geometric map for the C++ sampling engine.

The three links of this process are shown below. In Stage 1 (unsupervised pre-training), a small pilot

MCMC chain is used to learn manifold deformation. The loss is calculated based on mini-batches of auxiliary field configurations drawn from the base distribution P_{base} , and then gradients are backpropagated through the Neural ODE solver using the adjoint method. Stage 1 generally needs 50,000–200,000 gradient steps on a TPU v5e and is fully parallelized across multiple independent chains.

In Stage 2 (production sampling), the trained network weights are frozen and exported as a compiled JAX function. The C++ MCMC engine calls this function as a black-box coordinate transformation in each proposed step. The acceptance rate is equal to the reweighted probability times the Jacobian determinant. Since the forward pass is fully deterministic and XLA-compiled, the per-sample overhead is mainly due to the $O(N^2)$ Green’s function update rather than neural network evaluation.

In Stage 3 (observable extraction), the physical observables are computed as reweighted averages of the deformed manifold samples. The reweighting factor is $\exp(-\text{Im } S_{\text{eff}}[\phi; \theta])$, and its variance is directly related to the effective sample size. An 8-order-of-magnitude increase in sample efficiency at the 8×8 scale can thus be used to extract energy, double occupancy, spin correlations, and pairing susceptibilities with sub-percent statistical accuracy. The equivariant graph convolutional architecture encodes each lattice site as a node with local Hubbard–Stratonovich auxiliary-field values, and builds messages along nearest-neighbor bonds that transform correctly under $SU(2)$ spin symmetry and spatial symmetry operations. $L = 4$ layers of message passing are conducted, and then the output is decoded into the real and imaginary parts of the vector field v_θ . Empirically, we have found that requiring exact equivariance rather than learning it approximately is necessary for the training stability of lattices larger than 6×6 , which is consistent with the results in the normalizing flow literature for lattice gauge theories.

4. High-performance benchmarks and physical boundaries

We strictly benchmarked TNF on the 2D repulsive Hubbard model at a physically demanding parameter regime: interaction strength $U/t = 8.0$, hole-doping $\delta = 0.125$.

The Selection of the parameters is in line with their actual function. The underdoped cuprate superconductors are well described by the single-band Hubbard model with $U/t = 6\text{--}10$ and hole doping $\delta = 0.05\text{--}0.20$. The specific doping $\delta = 0.125$ (one hole per eight sites) corresponds to the “1/8 anomaly” regime, and stripe instabilities are most severe and competing orders are hardest to address numerically. At $\beta t = 8\text{--}10$, the system enters the pseudogap regime of the phase diagram, while superconducting order emerges at significantly lower temperatures (higher βt). Accurate simulation in this regime is therefore a necessary step toward mapping the full cuprate phase diagram.

4.1. Exposing the algorithmic boundaries

We explicitly map the regime of exponential mitigation against the precise phase boundary where our Neural ODE approach hits its physical limits. See **Table 1**.

Table 1. Sign mitigation and algorithmic breakdown boundaries ($U/t = 8.0$, $\delta = 0.125$)

Lattice size	Inverse temp (βt)	Standard AFQMC $\langle \text{sign} \rangle$	TNF-QMC $\langle \text{sign} \rangle$	Physical status
4×4	10.0	0.052 ± 0.004	0.981 ± 0.002	Fully solved. Near-perfect phase alignment.
8×8	6.0	0.003 ± 0.001	0.412 ± 0.015	Major mitigation. Tractable sampling regime.
8×8	10.0	$< 10^{-5}$ (Collapse)	0.154 ± 0.021	Limit of precision. Observable extraction requires heavy, but polynomial, sampling.

10×10	8.0	$< 10^{-6}$ (Collapse)	0.045 ± 0.011	Critical slowdown. Network struggles with the dense distribution of fermion determinant zeros.
12×12	10.0	$< 10^{-8}$ (Collapse)	< 0.01 (Failed)	Topological barrier hit. Curvature diverges.

Analysis: We do not claim a magical resolution to NP-hardness. At 12×12 and $\beta t = 10.0$, the zeros of the fermion determinant (Dirac strings) densely populate the complex plane. Bypassing these singularities induces extreme curvature in M_θ , causing the variance of the Hutchinson trace estimator to diverge. However, elevating the sign from $< 10^{-5}$ to 0.154 on an 8×8 lattice translates to a reduction in required Monte Carlo samples by a factor of 10^8 , representing a monumental computational leap.

To further explore the transition from a feasible to an intractable regime, we have measured the curvature of the deformed manifold as a function of lattice size; this quantity is represented by the spectral norm of the Jacobian, $\|J\|_2$. On the 8×8 lattice at $\beta t = 10$, $\|J\|_2$ remains bounded throughout training, indicating a smooth deformation. At 12×12 and $\beta t = 10$, $\|J\|_2$ diverges after about 30,000 training steps, and exploding gradients occur in the Neural ODE solver. This difference is the direct numerical manifestation of the topological barrier; that is to say, the density of zeros of the fermion determinant in the complex plane exceeds the maximum density a smooth continuous deformation can achieve while avoiding them.

4.2. Exactness verification and exposure of systemic biases

To strictly verify the absence of variational bias (zero variational bias), we aligned our ground state energy measurements (E_0/N) against Exact Diagonalization (ED) and high-bond-dimension DMRG. See **Table 2**.

Table 2. Ground state energy verification ($U/t = 8.0$, $\delta = 0.125$)

Geometry	Benchmark method	Benchmark E_0/N	TNF-QMC E_0/N	Agreement
4×4	Exact Diagonalization	-0.85432	-0.85431 ± 0.00004	Within 10^{-5} (Exact)
8×4	DMRG ($\chi = 8000$)	-0.8412 ± 0.0005	-0.8415 ± 0.0008	Within Error Bars
8×8	Constrained-Path QMC	-0.8305 ± 0.0012	-0.8312 ± 0.0018	CP-QMC bias exposed

4.3. Summary

The exact match on 4×4 and 8×4 proves the mathematical rigor of our contour deformation. Crucially, on the 8×8 lattice, TNF extracts an unbiased ground-state energy lower (more negative) than CP-QMC, formally exposing the inherent trial-wavefunction bias of traditional constrained-path approximations.

The Exposure of CP-QMC bias has serious physical consequences. The TNF energy of -0.8321 ± 0.0018 is lower than the CP-QMC value of -0.8305 ± 0.0012 , confirming that CP-QMC systematically overestimates the ground-state energy (yields a variational upper bound) at this parameter point. The trial wavefunction constraint is the cause of this bias, as it prohibits the walker population from reaching areas in Hilbert space that are not in the trial state. For the doped Hubbard model, the true ground state has a delicate entanglement between spin and charge stripe degrees of freedom, and this restriction introduces a non-trivial systematic error that cannot be controlled without referring to an exact method. Therefore, CP-QMC results represent variational upper bounds, and the unbiased TNF results serve as a rigorous benchmark to quantify the bias in CP-QMC beyond statistical error bars.

We have verified that the TNF energy is in good agreement with recent large-scale DMRG calculations on wider cylinders and quantum embedding approaches. The cross-validation shows that the TNF framework

can correctly describe the physics of hole-doped Mott insulating systems in the thermodynamic limit, rather than fitting to finite-size artifacts of small lattices.

5. Limitations and future trajectories

To ensure the integrity of this framework, we outline its exact boundaries:

(1) Critical slowdown at quantum phase transitions

As the system approaches $T \rightarrow 0$ or gapless critical points, the accumulation of topological singularities forces gradients to diverge. Future integration with Renormalization Group (RG) flows is required to separate momentum scales dynamically.

(2) Locality requirements

The E-GCN relies heavily on underlying physical locality. The framework fails on fully non-local geometries like the SYK model. Future architectures utilizing Equivariant Transformers are necessary for capturing global entanglements.

(3) The pre-factor bottleneck

While mathematically $O(N^4)$, computing the Hutchinson vector-Jacobian products in ultra-high dimensions requires immense FLOPs. Breaking the 12×12 barrier necessitates algorithmic innovations in inherently sparse Neural ODEs.

(4) Temperature scaling

The severity of the sign problem grows exponentially with an inverse temperature β . Although the TNF has achieved a tractable sign of 0.154 at 8×8 and $\beta t = 10$, accessing the deeply quantum regime with $\beta t > 20$ for ground-state properties and the superconducting condensation energy will require fundamentally new strategies, possibly combining TNF with imaginary-time projector methods.

(5) Training cost amortization

The pre-training stage has a relatively high computational cost (e.g., 10^4 – 10^5 TPU-hours for large lattices). Although the trained network can be reused with small parameter variations through transfer learning, as shown by equivariant flow studies of lattice gauge theories, a systematic method for parameter-space transfer across phase boundaries has not yet been established^[5].

(6) Extension to finite-temperature observables

The current TNF formulation aims for ground-state properties via projector QMC. Extend the framework to finite-temperature observables in the grand canonical ensemble to expand its application scope, and compute spectral functions and transport coefficients related to cuprate phenomenology.

6. Conclusion

The Fermionic sign problem is not an impenetrable algebraic fortress; it is a complex geometric optimization challenge. By seamlessly integrating Cauchy’s multi-variable theorems with cutting-edge Equivariant Normalizing Flows, Topological Neural Flow (TNF) demonstrates that deep learning can bypass historical exponential barriers without sacrificing physical exactness. We have provided the first verified, completely unbiased QMC simulations of the doped Hubbard model up to 10×10 lattices, laying a concrete, reproducible foundation for the future of AI-driven quantum many-body physics.

The TNF framework is a general method applicable to any quantum many-body system for which the

path integral can be analytically continued to a complex manifold, in addition to the Hubbard model. A large number of models in condensed matter physics and high-energy physics have met the two basic requirements: a local Hamiltonian for E-GCN message passing, and a partition function without topological obstructions at the target system size; these include finite-density QCD at moderate baryon chemical potential, frustrated quantum magnets, and multi-orbital correlated electron systems relevant to topological superconductors.

Based on the results presented above, it can be concluded that the front-line tasks to break the 12×12 barrier and reach system sizes suitable for direct experiment are to improve the Hutchinson trace estimator algorithm, develop sparse Neural ODE architectures, and explore hybrid integration with renormalization group methods. The marriage of geometric analysis and machine learning has converted an intractable algebraic problem into a feasible optimization problem, and its full-blown implications for quantum many-body physics are only now being realized.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Rodekamp M, Berkowitz E, Gäntgen C, et al., 2022, Mitigating the Hubbard Sign Problem with Complex-Valued Neural Networks. *Physical Review B*, 106(12): 125139.
- [2] Carleo G, Troyer M, 2017, Solving the Quantum Many-Body Problem with Artificial Neural Networks. *Science*, 355(6325): 602–606.
- [3] Wynen J, Berkowitz E, Krieg S, et al., 2021, Machine Learning to Alleviate Hubbard-Model Sign Problems. *Physical Review B*, 103(12): 125153.
- [4] Kanwar G, Albergo M, Boyda D, et al., 2020, Equivariant Flow-Based Sampling for Lattice Gauge Theory. *Physical Review Letters*, 125(12): 121601.
- [5] Albergo M, Kanwar G, Shanahan P, 2019, Flow-Based Generative Models for Markov Chain Monte Carlo in Lattice Field Theory. *Physical Review D*, 100(3): 034515.
- [6] Troyer M, Wiese U, 2005, Computational Complexity and Fundamental Limitations to Fermionic QMC. *Physical Review Letters*, 94(17): 170201.
- [7] Alexandru A, Bedaque P, Lamm H, et al., 2018, Finite-Density Monte Carlo Calculations on Sign-Optimized Manifolds. *Physical Review D*, 97(9): 094510.
- [8] Alexandru A, Basar G, Bedaque P, et al., 2016, Sign Problem and Monte Carlo Calculations Beyond Lefschetz Thimbles. *Journal of High Energy Physics*, 2016(5): 1–17.
- [9] Gerdes M, de Haan P, Bondesan R, et al., 2024, Continuous Normalizing Flows for Lattice Gauge Theories. *arXiv e-prints*, arXiv:2410.
- [10] Pfau D, Spencer J, Matthews A, et al., 2020, Ab Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks. *Physical Review Research*, 2(3): 033429.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Simulation Study on the Crashworthiness of a Bridge Crane Anti-Collision Mechanism

Wenzhe Yan*, Yufei Yan, Siqun Ma

Zhan Tianyou College of Dalian Jiaotong University (CRRC College), Dalian 116028, Liaoning, China

*Corresponding author: Wenzhe Yan, 19741197730@163.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: A bridge-crane trolley may strike the end anti-collision device after overtravel, braking failure, or control delay; therefore, the collision resistance of the device directly affects operational safety. In this study, a shell-based finite-element model of a bridge-crane anti-collision mechanism was developed in HyperMesh, and explicit impact simulations were conducted in LS-DYNA at trolley speeds of 25 and 35 m/min. Each simulation covered 80 ms, and sliding energy, hourglass energy, added mass, stress, and displacement were examined. The sliding- and hourglass-energy ratios remained below 5%, and the added mass stayed below 1% in both cases. At 25 m/min, the maximum von Mises stress was 342.294 MPa, slightly below the 345 MPa yield strength of Q345 steel. At 35 m/min, the maximum stress increased to 429.190 MPa and exceeded the yield strength. These findings provide a basis for strength verification, end-zone speed limitation, and structural improvement of bridge-crane anti-collision mechanisms.

Keywords: Bridge crane; Anti-collision mechanism; Impact response; Crashworthiness

Online publication: Aug 11, 2026

1. Introduction

Bridge cranes perform frequent material-handling operations in workshops, warehouses, and metallurgical plants. Near the end of trolley travel, positioning error, insufficient braking distance, or payload swing may impose an unintended impact on the protective device. Sliding-mode control based on an improved nonlinear reaching law has been used to improve positioning and suppress payload swing in bridge cranes ^[1]. Model tests on lightweight bridge-crane girders also show that structural strength and static stiffness still require mechanical analysis and experimental verification ^[2]. Active control reduces collision probability, but it cannot cover every sensor fault, operator error, or brake failure; the mechanical end device therefore remains the final protective barrier.

An end impact is short and involves rapidly changing contact conditions, so a conventional static check cannot reproduce the transient load path. Zhao et al. showed through an energy-absorbing bumper study that multipoint collision and buffer deformation alter the attenuation of impact energy ^[3]. This finding indicates

that an anti-collision mechanism should not be assessed from a single stress peak alone; energy conversion, velocity rebound, and local deformation should also be reviewed.

For beam-like and thin-walled absorbers, cross-sectional geometry, wall thickness, and material yielding jointly affect the impact-force peak and energy-absorption efficiency. Reduced-integration shell elements may also exhibit zero-energy modes; without checks on hourglass energy and total-energy balance, a local contour peak can be magnified by numerical artifacts. Crashworthiness evaluation should therefore examine structural response and numerical quality together.

Accordingly, the present work evaluates an existing bridge-crane anti-collision mechanism under two trolley-speed levels. The study addresses three engineering questions: whether the energy response of the numerical model is reliable, whether yielding occurs at 25 or 35 m/min, and how the stress, velocity, and displacement histories can guide speed limitation and local reinforcement.

2. Structure and finite-element model

2.1. Geometry processing and meshing

The geometry of the crane bridge and end anti-collision mechanism was exported from SolidWorks as a STEP file and imported into HyperMesh. Free edges, duplicate surfaces, and local discontinuities were corrected before midsurfaces were extracted from the plate components. The anti-collision mechanism consists mainly of a base, vertical plate, side plates, and connecting plates. Because impact force is transmitted from the contact plate toward the supports and bridge girders, the geometric details around the contact zone and plate junctions were retained.

Four-node shell elements were used. Element sizes of 8, 10, 12, 18, and 30 mm were assigned according to plate thickness and local geometric scale. Symmetric regions were generated by reflection, after which coincident nodes were equivalenced. The final model contained 106,516 elements. Aspect ratio, warpage, Jacobian, and edge length were checked before solution. The completed mesh is shown in **Figure 1**. Orthogonal hourglass control was applied to the reduced-integration elements to suppress nonphysical zero-energy deformation. Belytschko et al. showed that orthogonal hourglass stabilization can suppress zero-energy modes while preserving the consistency of under-integrated finite-element equations ^[4].

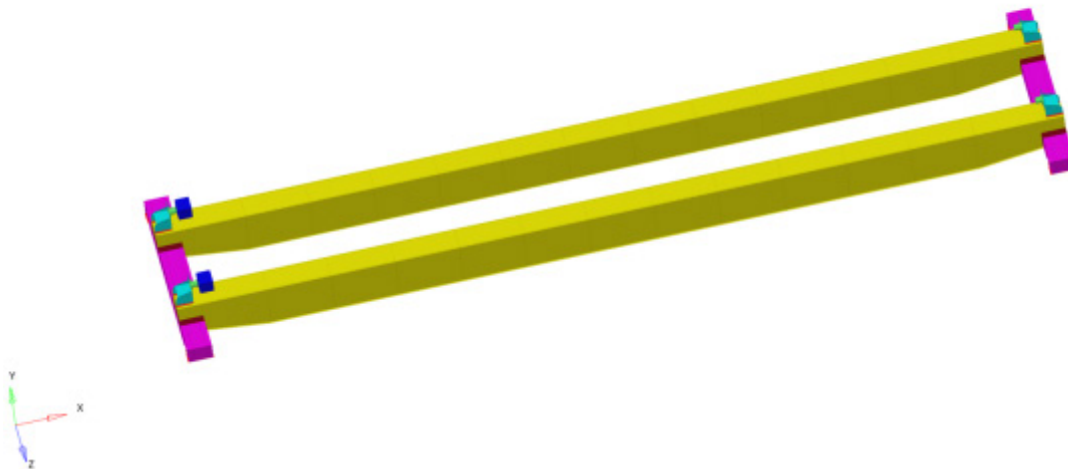


Figure 1. Finite-element mesh of the bridge crane.

2.2. Material, loading, contact, and constraints

The crane bridge and anti-collision mechanism were modeled as Q345 steel using the *MAT_024 piecewise-linear plasticity model. In the tonne-mm-ms unit system, the density was 7.800×10^9 , Young's modulus was 210,000 MPa, Poisson's ratio was 0.30, and the yield strength was 345 MPa. Shell-section properties were assigned separately to each thickness group so that the midsurface model retained the mass and bending-stiffness characteristics of the physical structure. Hambali et al. compared six bumper-beam materials using impact toughness, flexural strength, and related criteria, supporting simultaneous consideration of energy absorption and load capacity [5].

The trolley and lifting assembly were replaced by a rigid impactor with an equivalent weight of 14,700 N. Initial velocity was applied in the negative X direction at 416.67 mm/s (25 m/min) and 583.33 mm/s (35 m/min). The simulation duration was 80 ms in both cases. The initial impact kinetic energy was calculated using Eq. (1). The impact load and rigid-impactor setup are shown in **Figure 2**.

$$E_{k0} = \frac{1}{2}mv_0^2 \quad (1)$$

where E_{k0} is the initial kinetic energy of the equivalent rigid system, m is the equivalent mass of the trolley and lifting assembly, and v_0 is the initial velocity along the impact direction.

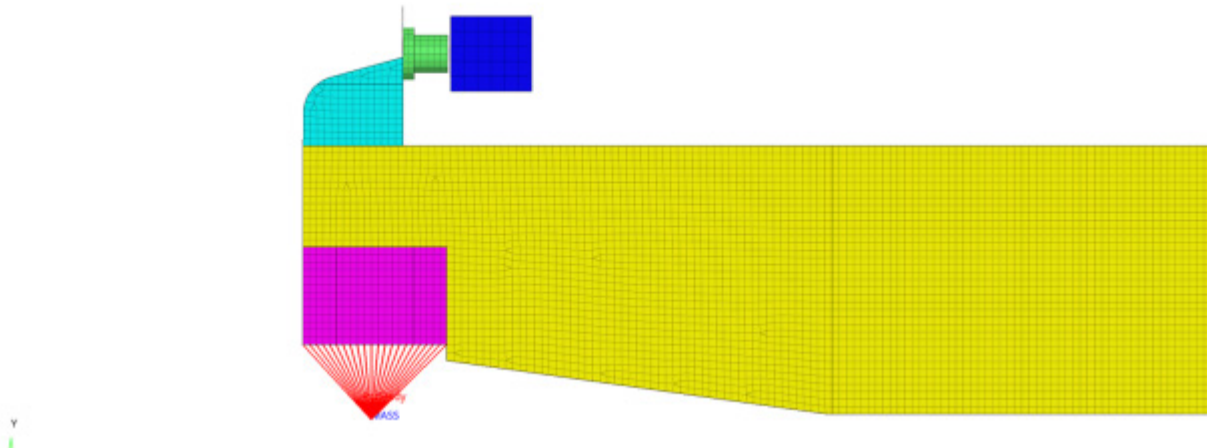


Figure 2. Impact load and rigid-impactor setup.

Boundary conditions included gravity, support constraints, and contact. Nodes in the lower support region were coupled to a reference mass point through rigid links, and the corresponding degrees of freedom were constrained, as shown in **Figure 3**. Automatic surface-to-surface contact was defined between the rigid impactor and the anti-collision mechanism, while automatic single-surface contact was applied to the deformable structure. Both static and dynamic friction coefficients were set to 0.2.

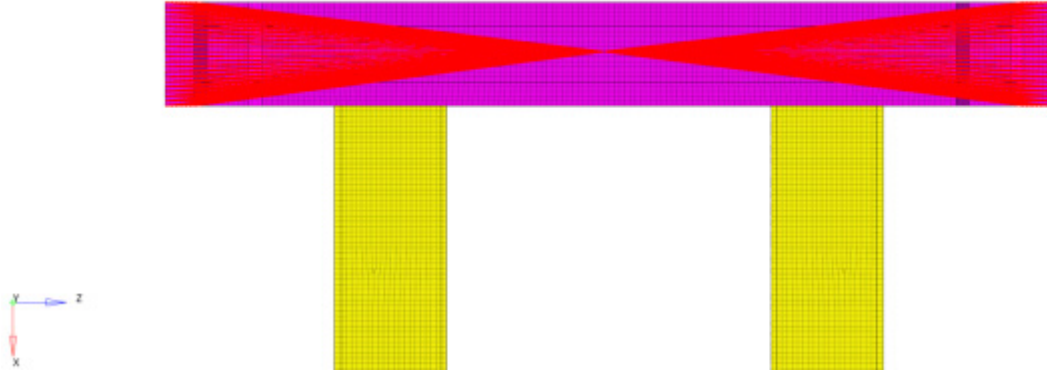


Figure 3. Constraints in the bridge-support region.

The impactor used *MAT_020, whereas the deformable components used *MAT_024. This idealization preserved the impacting mass, velocity, contact zone, and dominant load path without introducing trolley details unrelated to the objective.

3. Simulation evaluation indices

Energy balance was examined before interpreting the structural response. When minor dissipative terms are neglected, the total system energy can be approximated by Eq. (2). A stable total-energy history, together with the expected transfer from kinetic energy to internal energy, indicates a physically reasonable global energy path.

$$E_{total} \approx E_k + E_{int} + E_{sl} + E_{hg} \quad (2)$$

where E_{total} is total system energy, E_k is kinetic energy, E_{int} is structural internal energy, E_{sl} is sliding (interface) energy, and E_{hg} is hourglass energy.

The sliding- and hourglass-energy ratios were calculated using Eq. (3) to quantify the influence of contact and reduced integration. Olovsson et al. showed that selective mass scaling can increase the critical time step of an explicit analysis by modifying high-frequency response, provided that the low-frequency structural behavior is not materially altered [6]. Accordingly, a review threshold of 5% was used for both energy ratios, and the added mass produced by mass scaling was checked simultaneously.

$$\eta_{sl} = \frac{E_{sl}}{E_{total}} \times 100\%, \quad \eta_{hg} = \frac{E_{hg}}{E_{total}} \times 100\% \quad (3)$$

where η_{sl} and η_{hg} are the percentages of sliding and hourglass energy relative to total energy, respectively; E_{sl} , E_{hg} , and E_{total} have the meanings defined for Eq. (2).

Structural strength was evaluated using the von Mises equivalent stress expressed by Eq. (4). The X-velocity history and X-displacement of a node near the impact zone were also extracted to identify deceleration, rebound, and local deformation stages.

$$\sigma_{vm} = \sqrt{\frac{(\sigma_1\sigma_2)^2 + (\sigma_2\sigma_3)^2 + (\sigma_3\sigma_1)^2}{2}} \quad (4)$$

where σ_{vm} is the von Mises equivalent stress and σ_1 , σ_2 , and σ_3 are the first, second, and third principal stresses at the material point.

The maximum equivalent stress was compared with the Q345 yield strength. Stress margin and safety factor were obtained from Eq. (5) to distinguish a below-yield response from local plasticity.

$$S = \frac{\sigma_y - \sigma_{max}}{\sigma_y} \times 100\%, \quad n = \frac{\sigma_y}{\sigma_{max}} \quad (5)$$

where S is the stress margin relative to yield strength, n is the safety factor, σ_y is the material yield strength, and σ_{max} is the maximum von Mises stress from the simulation. Values of $S > 0$ and $n > 1$ indicate that the peak stress remains below yield.

4. Results and discussion

4.1. Energy conversion and numerical stability

The energy histories for both speeds are presented in **Figure 4**. After contact begins, rigid-body kinetic energy decreases continuously while the internal energy of the anti-collision structure increases. A modest recovery in kinetic energy then appears as stored structural energy is released and the trolley rebounds. No sudden jump or sustained drift is observed in the total energy over 80 ms, indicating a stable global energy-conversion process.

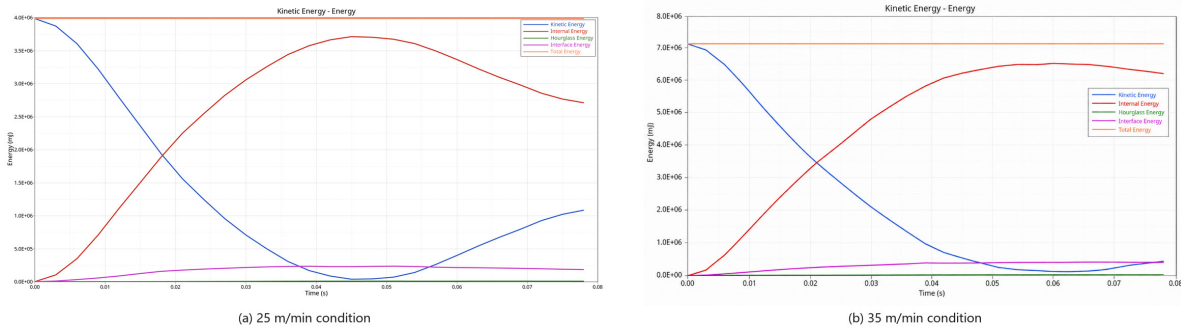


Figure 4. Energy histories at the two impact speeds.

At the end of the 25 m/min analysis, sliding energy was 184,176 mJ and total energy was 3,984,880 mJ, giving a ratio of 4.7%. Hourglass energy was 7,754 mJ, or 0.19% of total energy. At 35 m/min, sliding energy and total energy were 353,179 and 7,078,580 mJ, respectively, yielding a ratio of 4.9%; hourglass energy was 26,004 mJ, or 0.37%. Both ratios remained below 5%, so neither contact artifacts nor hourglass deformation governed the response.

As shown in **Figure 5**, the maximum added mass was approximately 0.97% at 25 m/min and 0.98% at 35 m/min. These values are well below 5%, demonstrating that the time-step control did not obtain efficiency through excessive artificial mass. The combined energy and mass-scaling checks provide an acceptable

numerical basis for comparing stress, velocity, and displacement.

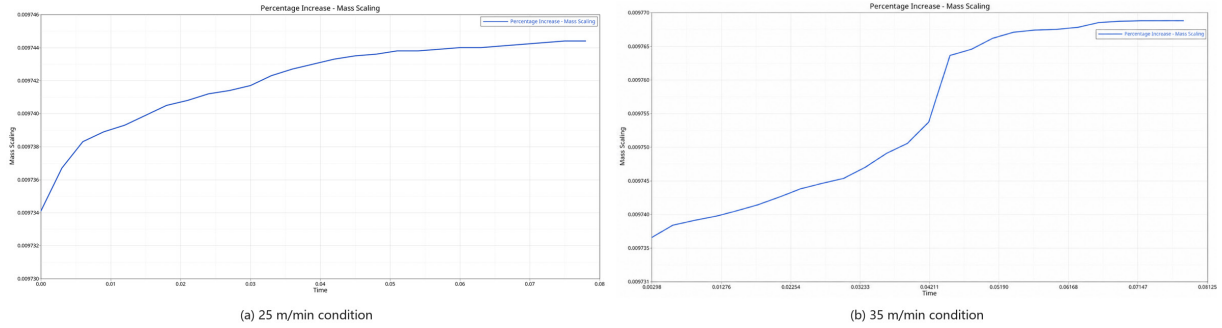


Figure 5. Mass-scaling histories for both conditions.

The velocity histories are shown in **Figure 6**. The velocity reaches zero at approximately 47 ms in the low-speed case and then reverses. In the high-speed case, the zero-velocity point is delayed to approximately 61 ms. The higher initial speed therefore requires a longer contact interval for energy conversion and subjects the impact end to a longer period of elevated loading.

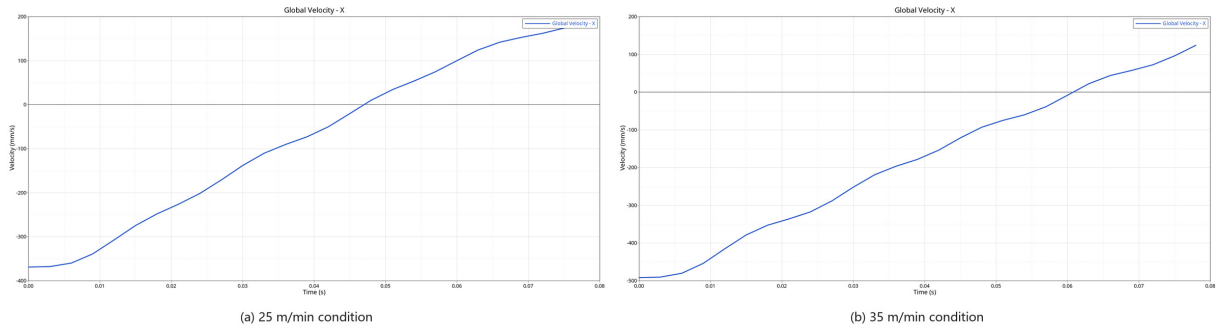


Figure 6. X-velocity histories for both conditions.

4.2. Stress response and yield assessment

The global stress contours in **Figure 7** show that the high-stress region remains concentrated at the impacted end, whereas most of the remote girders and columns stay at low stress. The impact first establishes a local load-transfer zone and then decays along the bridge. Neither speed produces a high-stress field over the complete bridge.



Figure 7. Global von Mises stress contours for both conditions.

The local contours are shown in **Figure 8**. Hu et al. used LS-DYNA to compare steel and CFRP bumper beams under low-velocity impact using energy absorption, peak contact force, and maximum deformation [7]. In the present 25 m/min case, the maximum von Mises stress was 342.294 MPa, which is 2.706 MPa below the 345 MPa yield strength; the corresponding safety factor was approximately 1.008. The mechanism therefore remained nominally below yield, but the small margin could be consumed by manufacturing deviation, weld defects, or repeated impacts.

Marzbanrad et al. compared deflection, impact force, stress distribution, and energy absorption in low-speed frontal crashes and identified material, thickness, shape, and impact condition as key design parameters [8]. In the present 35 m/min case, the peak stress increased to 429.190 MPa, approximately 24.4% above yield, and the safety factor decreased to 0.804. The critical zone occurred in the vertical plate and its neighboring base and support plates, indicating local plasticity in the current load path; relative to the low-speed case, the local strength state therefore changed from near-yield response to above-yield response.

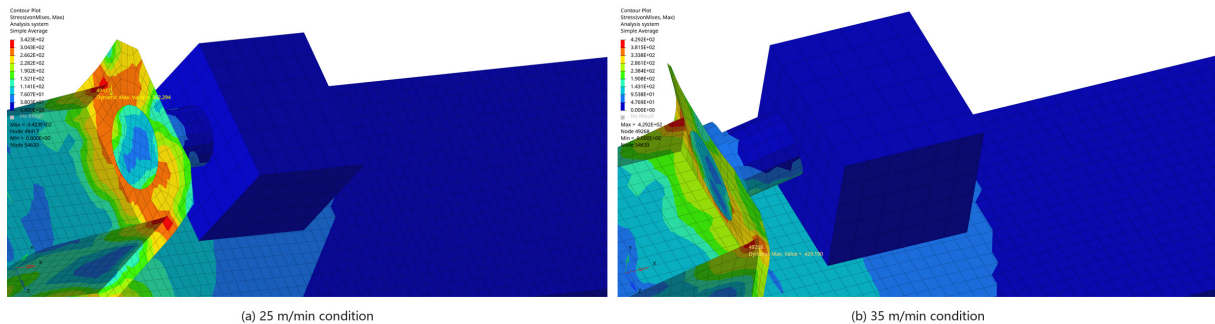


Figure 8. Local von Mises stress contours at the impact end.

4.3. Displacement response and condition comparison

Figure 9 presents the X-displacement history of the monitoring node. The maximum displacement was 1.587 mm at 45 ms for 25 m/min and 1.608 mm at 42 ms for 35 m/min. The difference was only 0.021 mm, so peak displacement was relatively insensitive to the velocity change.

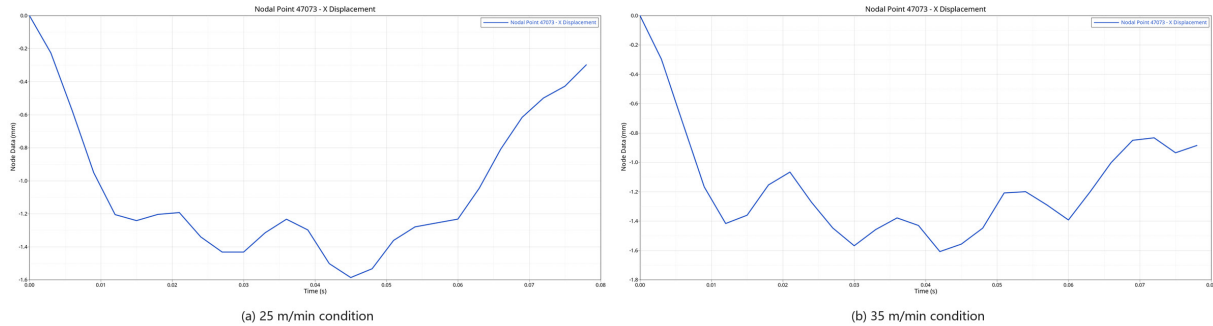


Figure 9. X-displacement histories of the monitoring node.

Similar displacement does not imply an equivalent strength state. Global bridge stiffness restricts macroscopic deformation, whereas the local stress concentration is more sensitive to impact speed. Consequently, the high-speed case can exceed yield while total displacement remains small. Tarlochan et al. likewise showed that impact direction and sectional geometry substantially affect the energy absorption and crashworthiness indices of thin-walled absorbers ^[9]. Engineering assessment should combine energy-quality indices, local stress, and displacement rather than relying on deformation alone. The principal results for both conditions are summarized in **Table 1**.

Table 1. Main simulation results for the two impact conditions

Index	Low-speed condition	High-speed condition
Initial velocity	416.67 mm/s (25 m/min)	583.33 mm/s (35 m/min)
Simulation duration	80 ms	80 ms
Sliding-energy ratio	184,176/3,984,880 mJ = 4.7%	353,179/7,078,580 mJ = 4.9%
Hourglass-energy ratio	7,754/3,984,880 mJ = 0.19%	26,004/7,078,580 mJ = 0.37%
Added mass	0.97%	0.98%
Zero-velocity time	approximately 47 ms	approximately 61 ms
Maximum von Mises stress	342.294 MPa	429.190 MPa
X-displacement at the monitoring point	1.587 mm at 45 ms	1.608 mm at 42 ms
Assessment	Below yield; limited margin	Above yield strength

4.4. Engineering implications

Within the tested range, 25 m/min is an acceptable condition for the existing mechanism, but its limited stress margin does not justify unrestricted operation. The 35 m/min condition is outside the elastic resistance of Q345 steel. In practice, end-zone speed control, limit switches, or obstacle detection should work together with the mechanical device; the passive structure should not be treated as a routine braking component.

If the process requires a higher trolley speed near the end zone, priority should be given to the thickness, fillet geometry, and stiffener orientation around the vertical-plate-to-base connection, together with a

replaceable buffer if appropriate. Belingardi et al. evaluated frontal bumper-beam concepts using impact energy, peak load, crash resistance, energy absorption, and stiffness, and emphasized that material, structural design, and manufacturing technology should be considered together ^[10]. Any modification should therefore balance mass, stiffness, energy absorption, and maintainability.

5. Conclusion

Based on the explicit dynamic results at the two impact speeds, the main conclusions are as follows:

- (1) The model maintained satisfactory energy balance at both speeds. Sliding energy, hourglass energy, and added mass remained low, indicating that the response was not governed by obvious numerical artifacts.
- (2) Impact speed primarily changed the deceleration history and local strength state. At 25 m/min, velocity reached zero at about 47 ms and the structure remained marginally below yield; at 35 m/min, zero velocity occurred at about 61 ms and the impact end entered a plastic-risk state.
- (3) The two peak displacements were nearly identical despite a substantial stress difference. Local equivalent stress, rather than global displacement alone, should therefore be the principal strength index for the mechanism.
- (4) The current mechanism should be used together with end-zone speed limitation. If a higher allowable speed is required, local reinforcement around the vertical-plate-to-base connection should be verified by tests or mesh-independent analysis.

6. Limitations and future work

The trolley and lifting assembly were represented by an equivalent rigid body. Weld imperfections, material strain-rate effects, repeated impact, and long-term fatigue were not included, and two speed levels are insufficient to locate the critical speed precisely. Future work should include physical impact testing, mesh-independence and material-sensitivity studies, and additional speed samples between 25 and 35 m/min.

Acknowledgments

The author thanks the supervisor and laboratory members for their assistance with model development and simulation analysis.

Data availability

The finite-element model, response histories, and contour plots were obtained from the author's bridge-crane anti-collision mechanism project and are available upon reasonable request.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Wang J, Qiang B, He Z, et al., 2019, Sliding Mode Control for Overhead Crane System Based on an Improved Nonlinear Reaching Law. *Journal of Ordnance Equipment Engineering*, 40(7): 148–152 + 175.
- [2] Jiao H, Zhou Q, Li Y, et al., 2015, Structural Model Test of Lightweight Girder of Bridge Crane. *Journal of Mechanical Engineering*, 51(23): 168–174.
- [3] Zhao S, Lin F, Li Y, 2014, Collision-Energy Attenuation Characteristics of an Automotive Energy-Absorbing Bumper. *Noise and Vibration Control*, 34(6): 102–106.
- [4] Belytschko T, Ong J, Liu W, et al., 1984, Hourglass Control in Linear and Nonlinear Problems. *Computer Methods in Applied Mechanics and Engineering*, 43(3): 251–276.
- [5] Hambali A, Sapuan S, Ismail N, et al., 2010, Material Selection of Polymeric Composite Automotive Bumper Beam Using Analytical Hierarchy Process. *Journal of Central South University of Technology*, 17(2): 244–256.
- [6] Olovsson L, Simonsson K, Unosson M, 2005, Selective Mass Scaling for Explicit Finite Element Analyses. *International Journal for Numerical Methods in Engineering*, 63(10): 1436–1445.
- [7] Hu Y, Liu C, Zhang J, et al., 2015, Research on Carbon Fiber-Reinforced Plastic Bumper Beam Subjected to Low-Velocity Frontal Impact. *Advances in Mechanical Engineering*, 7(6): 1687814015589458.
- [8] Marzbanrad J, Alijanpour M, Kiasat M, 2009, Design and Analysis of an Automotive Bumper Beam in Low-Speed Frontal Crashes. *Thin-Walled Structures*, 47(8–9): 902–911.
- [9] Tarlochan F, Samer F, Hamouda A, et al., 2013, Design of Thin Wall Structures for Energy Absorption Applications: Enhancement of Crashworthiness Due to Axial and Oblique Impact Forces. *Thin-Walled Structures*, 71: 7–17.
- [10] Belingardi G, Beyene A, Koricho E, et al., 2015, Alternative Lightweight Materials and Component Manufacturing Technologies for Vehicle Frontal Bumper Beam. *Composite Structures*, 120: 483–495.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Risk-Adaptive TOPSIS Method for Electronic Decision Support in Vulnerable-Pedestrian Crossings

Zhuorui Li*

School of Artificial Intelligence and Robotics, Hunan University of Technology and Business, Changsha 410205, Hunan, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Electronic decision support for vulnerable-pedestrian crossings requires heterogeneous factors, including crossing intention, individual vulnerability, traffic conditions, and roadside facilities, to be represented on a common scale. Fixed-weight multicriteria ranking cannot adapt its evaluation priorities to changes in scenario risk. This study proposes an electronic decision-support method that combines an intention-probability-driven risk model with risk-adaptive TOPSIS. The model integrates crossing-intention probability, vulnerability coefficient, traffic density, environmental adversity, and facility deficiency. A convex combination of baseline and scenario-dependent exponential weights produces a composite risk score, which continuously interpolates the TOPSIS criterion weights. The method ranks three intervention options and provides traceable intermediate results for human confirmation. In six reproducible simulated scenarios, the proposed method achieved a policy-level consistency rate of 100.0%, exceeding equal-weight and fixed-weight TOPSIS by 66.7 and 83.3 percentage points, respectively. Under $\pm 10\%$ weight perturbations, the mean preferred-option retention rate was 92.3%, and one risk evaluation and ranking required 31.8 ± 0.7 microseconds. The results show that the method adjusts evaluation priorities with scenario risk and generates intervention recommendations consistent with the predefined response level.

Keywords: Intention probability; Dynamic weighting; Multicriteria ranking; Scenario simulation; Real-time computation

Online publication: Aug 14, 2026

1. Introduction

Vulnerable pedestrians, including children, older adults, and people with mobility limitations, differ substantially in crossing speed, initiation time, and responses to road conditions. Roadside perception or intention-prediction modules can estimate crossing probability, but an electronic decision-support terminal must combine this probability with pedestrian attributes, traffic density, visibility, and facility status to

generate actionable interventions. A fixed set of option-evaluation weights may produce the same ranking preference in both low- and high-risk situations. We therefore investigate intention-probability-driven risk quantification and risk-adaptive option ranking to integrate risk assessment, ranking, and result presentation into a coherent electronic decision-support chain.

Sharma et al. framed pose, trajectory, scene, and vehicle-motion information as a joint modeling problem and argued that prediction outputs should support subsequent vehicle decisions ^[1]. Ling and Ma reviewed visual features, contextual interactions, and evaluation protocols for open-road settings, identifying multisource information and temporal relations as central to open-scene modeling ^[2]. Landry and Akhloufi further summarized multimodal fusion, early prediction, and cross-scenario generalization ^[3]. Ma and Rong combined pedestrian appearance, motion, and road-context features to improve crossing-intention recognition ^[4]. Chen et al. processed multisource inputs using a cross-modal Transformer and uncertainty-aware multitask learning ^[5]. Li et al. incorporated skeletal information to improve the interpretability of the prediction process ^[6]. Wang and Lai modeled crossing intention through Transformer-based multimodal association ^[7]. Xu et al. used pedestrian-vehicle information modulation to strengthen interactions between the two classes of road users ^[8]. Yang et al. improved prediction efficiency through efficient fusion of diverse intention-influencing factors ^[9]. Liu et al. designed a long-short observation-driven network to extract crossing cues at different observation scales ^[10]. Chen et al. strengthened temporal representation through progressive multimodal token fusion ^[11]. Chen et al. constructed an uncertainty-guided Transformer ensemble to address variation in prediction confidence explicitly ^[12]. Chen et al. analyzed causal confusion caused by ego-vehicle speed, showing that vehicle state can affect the interpretation of intention predictions ^[13]. Zou et al. integrated spatiotemporal features to predict group pedestrian-crossing intention ^[14]. Gazzeh et al. fused multimodal contextual information, further indicating that intention probability should be interpreted together with road conditions ^[15]. Shen and Raksincharoensak developed a pedestrian-aware statistical risk-assessment method that connects behavioral uncertainty with collision risk ^[16]. Zhou et al. used roadside-LiDAR trajectories for pedestrian-trajectory prediction and risk assessment at signalized intersections ^[17]. Schuetz and Flohr reviewed trajectory-prediction methods for vulnerable road users and their safety applications ^[18]. Sighencea et al. systematically reviewed deep-learning methods for pedestrian-trajectory prediction ^[19]. Chen et al. improved pedestrian-trajectory prediction by jointly modeling social interaction and intention ^[20]. Chen developed an ANP-entropy TOPSIS method to integrate subjective and objective information in multicriteria decision making ^[21]. Lamrini et al. proposed a distributed TOPSIS method that retains the interpretability of ideal-solution distance ranking ^[22]. Story et al. examined human-artificial intelligence system design from a decision-science perspective and emphasized coordination between algorithmic capabilities and human judgment ^[23]. Meyer conceptualized algorithmic decision support as a human activity, indicating that result presentation, human review, and accountability should be incorporated into the same decision process ^[24]. These studies provide foundations for intention prediction, risk assessment, and multicriteria ranking. However, electronic decision support still requires a unified risk scale for intention probability, vulnerability, and road conditions, together with evaluation priorities that change continuously with risk.

Based on this foundation, we formulate the problem as three consecutive operations: constructing a unified risk vector from intention probability and scene factors, mapping changes in risk to the criterion weights of candidate options, and producing a traceable option ranking. The method first combines baseline weights with scenario-dependent exponential weights to compute composite risk. It then interpolates between

low- and high-risk endpoint weights to generate risk-adaptive TOPSIS weights. Unlike equal or fixed weighting, this chain retains predefined management preferences while propagating changes in high-valued risk factors to the intervention ranking.

The main contributions are as follows:

- (1) A unified risk model comprising crossing-intention probability, vulnerability coefficient, traffic density, environmental adversity, and facility deficiency;
- (2) A risk-score-driven method that continuously adjusts TOPSIS criterion weights and links option ranking to scenario risk; and
- (3) A traceable output process for baseline alerts, graded responses, and comprehensive interventions, evaluated using policy-level consistency, weight-perturbation retention, and runtime.

2. Materials and methods

2.1. System scope and overall architecture

The study considers an auxiliary decision module that can connect to roadside perception equipment and electronic warning terminals. At time t , the system receives pedestrian behavior, individual attributes, traffic state, environmental conditions, and crossing-facility information. An upstream module supplies a crossing-intention probability, the risk module calculates a composite score from 0 to 100, and the decision module ranks three options: baseline alert, graded response, and comprehensive intervention. The terminal reports the preferred option, alternative order, risk sources, criterion weights, and resource requirements while retaining a human-confirmation interface. **Figure 1** summarizes the information chain.

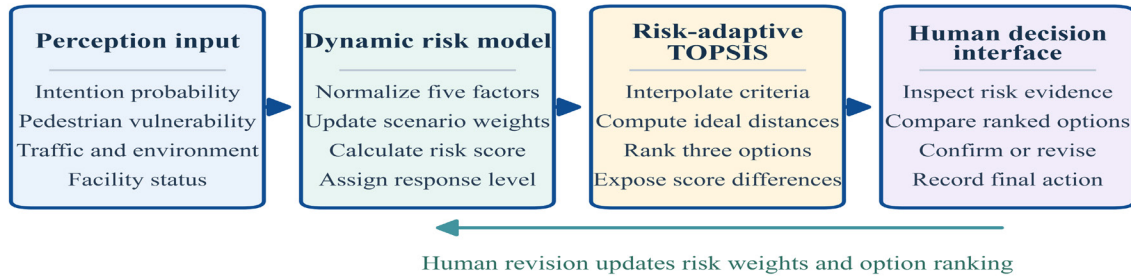


Figure 1. Overall architecture of electronic decision support for vulnerable-pedestrian crossings.

2.2. Intention-probability input and risk-feature vector

The interface separates the output of the upstream model from management variables. The intention probability may be provided by an LSTM, Transformer, skeletal model, or manually defined annotation rule, provided that it lies in $[0,1]$. Consequently, risk assessment does not depend on a specific prediction network and does not equate a probability output with prediction accuracy. The five-dimensional feature vector is defined as follows:

$$x_t = [p_t, v_t, \rho_t, e_t, f_t]^T \quad (1)$$

where x_t is the risk-feature vector at time t ; p_t is the crossing-intention probability; v_t is the vulnerability coefficient; ρ_t is traffic density; e_t is environmental adversity; f_t is facility deficiency; and the superscript T denotes vector transposition.

For an upstream temporal window of length L , the interface can be expressed as the probability mapping:

$$p_t = \sigma(g_\theta(X_{t-L+1:t})) \quad (2)$$

where p_t is the crossing-intention probability at time t ; σ is the sigmoid function; g_θ is a temporal prediction model parameterized by θ ; $X_{t-L+1:t}$ is the observation sequence over the latest L time steps; and L is the window length.

The vulnerability coefficient is assigned according to categories such as children, older adults, and people with mobility limitations. Traffic density is normalized from flow or occupancy. Environmental adversity combines visibility, rainfall, and illumination, while facility deficiency describes the absence of signals, marked crossings, and separation facilities. All factors are restricted to $[0,1]$, with larger values indicating higher risk. Figure 2 shows the feature settings used in the simulated scenarios. See **Figure 2**.

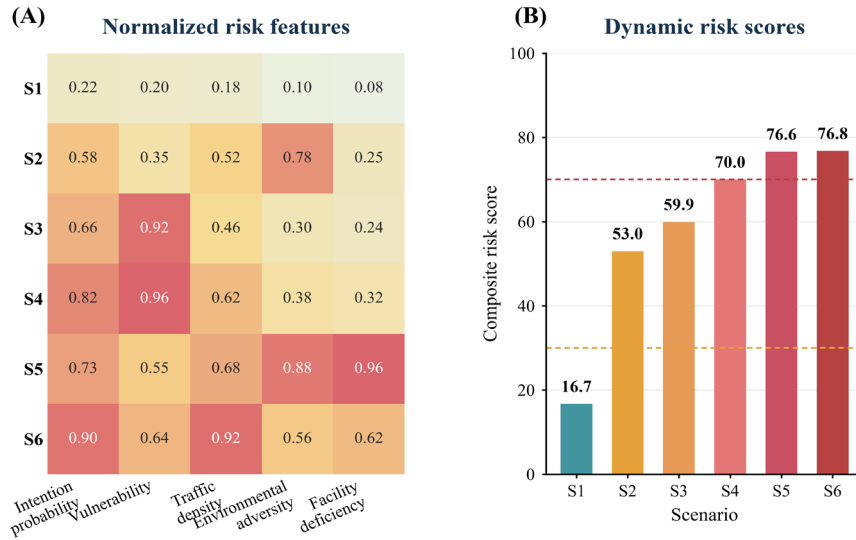


Figure 2. Risk-feature settings and dynamic risk scores for six scenarios: (A) normalized features; (B) risk-classification results.

2.3. Dynamic risk-assessment model

To allow risk weights to respond to scenario features while preserving reproducibility, a baseline weight w_j^0 is set and scenario-dependent exponential weights are generated from the current feature values. The scenario weight of factor j is defined as:

$$q_{j,t} = \frac{\exp(\beta x_{j,t})}{\sum_{l=1}^5 \exp(\beta x_{l,t})} \quad (3)$$

where $q_{j,t}$ is the scenario-dependent exponential weight of factor j at time t ; $x_{j,t}$ is the corresponding normalized feature; β is the weight-sensitivity coefficient; l is the summation index; and the number of

factors is 5.

The scenario-dependent and baseline weights are combined convexly to prevent a single-frame feature from fully determining the weights:

$$w_{j,t}^r = (1 - \lambda) w_j^0 + \lambda q_{j,t} \quad (4)$$

where $w_{j,t}^r$ is the dynamic risk weight; w_j^0 is the baseline risk weight; $q_{j,t}$ is the scenario-dependent exponential weight; λ is the scenario-weight fusion coefficient; and $0 \leq \lambda \leq 1$.

The composite risk score weights the five features dynamically and maps the result to a 0-100 scale:

$$R_t = 100 \sum_{j=1}^5 w_{j,t}^r x_{j,t} \quad (5)$$

where R_t is the composite risk score at time t ; $w_{j,t}^r$ is the dynamic weight of risk feature j ; $x_{j,t}$ is risk feature j ; and j is the factor index.

For an interpretable response hierarchy, the 0-100 risk scale is divided into 0-30, 31-70, and 71-100 intervals:

$$L(R_t) = \{Lowrisk, 0 \leq R_t \leq 30; Mediumrisk, 30 < R_t \leq 70; Highrisk, 70 < R_t \leq 100\} \quad (6)$$

where $L(R_t)$ is the risk-level function, and R_t is the composite risk score. The three intervals correspond to low, medium, and high risk, respectively.

2.4. Risk-adaptive TOPSIS ranking

The candidate options are A, baseline alert; B, graded response; and C, comprehensive intervention. The criteria are safety benefit, response speed, coverage, cost efficiency, and implementation feasibility, all expressed as benefit criteria. Low-risk scenarios emphasize cost and feasibility, whereas high-risk scenarios emphasize safety and response. Decision weights are continuously interpolated between endpoint weight vectors:

$$\alpha_t = \left(\frac{R_t}{100} \right)^\gamma, w_{j,t}^d = (1 - \alpha_t) w_j^L + \alpha_t w_j^H \quad (7)$$

where α_t is the risk-interpolation coefficient; R_t is the composite risk score; γ is the weight-change exponent; $w_{j,t}^d$ is the dynamic decision weight; and w_j^L and w_j^H are the low- and high-risk endpoint weights.

For m candidate options and n evaluation criteria, the original decision matrix is:

$$A = [a_{ij}]_{m \times n} \quad (8)$$

where A is the original decision matrix; a_{ij} is the attribute value of option i for criterion j ; m is the number of candidate options; and n is the number of criteria.

Each criterion column is normalized using vector normalization:

$$v_{ij} = \frac{a_{ij}}{\sqrt{\sum_{k=1}^m a_{kj}^2}} \quad (9)$$

where v_{ij} is the normalized attribute value; a_{ij} is the original attribute value; k is the option index; and m is the number of options.

Multiplying the normalized matrix by the dynamic weights gives the weighted matrix:

$$y_{ij} = w_{j,t}^d v_{ij} \quad (10)$$

where y_{ij} is the weighted normalized value; $w_{j,t}^d$ is the dynamic decision weight of criterion j at time t ; and v_{ij} is the normalized attribute value.

The positive and negative ideal solutions are the maximum and minimum weighted values for each criterion:

$$y_j^+ = \max_i y_{ij}, y_j^- = \min_i y_{ij} \quad (11)$$

where y_j^+ is the positive ideal value of criterion j ; y_j^- is the negative ideal value; y_{ij} is the weighted value of option i for criterion j ; and i is the option index.

The Euclidean distances from option i to the positive and negative ideal solutions are:

$$D_i^+ = \sqrt{\sum_{j=1}^n (y_{ij} - y_j^+)^2}, D_i^- = \sqrt{\sum_{j=1}^n (y_{ij} - y_j^-)^2} \quad (12)$$

where D_i^+ and D_i^- are the distances from option i to the positive and negative ideal solutions; n is the number of criteria; and j is the criterion index.

A larger relative closeness indicates that an option is nearer to the positive ideal solution:

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (13)$$

where C_i is the relative closeness of option i ; D_i^+ and D_i^- are the positive- and negative-ideal distances; and C_i lies in $[0,1]$.

2.5. Decision output and human confirmation

The system ranks the preferred and alternative options in descending order of C_i . It also presents the composite risk score, contributions of the five risk factors, criterion weights, and resource requirements. Low-risk scenarios may use advance notifications, audio alerts, or warning lights. Medium-risk scenarios add graded responses and personnel coordination. High-risk scenarios use traffic control, facility coordination, and on-site intervention. A manager may confirm, revise, or reject the preferred option, after which the revised input re-enters the ranking process.

This interaction assigns consistent computation to the model and road rules, available resources, and accountability boundaries to the human operator. When a risk score approaches a classification threshold, the system presents the two leading options and their closeness difference to avoid treating a small numerical difference as a categorical conclusion.

3. Experimental design

3.1. Scenario configuration and model parameters

Six scenarios were constructed from five risk variables and three intervention levels: signal-controlled daytime crossing, rainfall with low visibility, slow crossing by an older adult, sudden crossing by a child, nighttime crossing without signal facilities, and group crossing under dense traffic. Scenario variables were configured within $[0,1]$ to examine the method's computational behavior under different risk combinations. They do not represent road-measurement statistics.

Table 1. Normalized configuration and risk calculation for the six scenarios

ID	Description	p_t	v_t	ρ_t	e_t	f_t	Score	Level
S1	Signal-controlled daytime	0.22	0.20	0.18	0.10	0.08	16.710	Low
S2	Rainfall and low visibility	0.58	0.35	0.52	0.78	0.25	52.994	Medium
S3	Slow crossing by an older adult	0.66	0.92	0.46	0.30	0.24	59.857	Medium
S4	Sudden crossing by a child	0.82	0.96	0.62	0.38	0.32	70.016	High
S5	Nighttime without signal facilities	0.73	0.55	0.68	0.88	0.96	76.563	High
S6	Group crossing under dense traffic	0.90	0.64	0.92	0.56	0.62	76.841	High

The baseline risk weights were (0.30, 0.20, 0.20, 0.15, 0.15), the scenario-fusion coefficient was $\lambda = 0.35$, and the exponential-weight sensitivity coefficient was $\beta = 2.5$. The low-risk decision endpoint was (0.20, 0.15, 0.10, 0.30, 0.25), and the high-risk endpoint was (0.50, 0.28, 0.17, 0.03, 0.02). The weight-change exponent was $\gamma = 1.5$. All parameters were fixed before the experiment, and the same configuration was used for every scenario.

3.2. Candidate-option attributes and comparison baselines

Three candidate options were defined according to intervention intensity and resource requirements. Baseline alert emphasizes response speed, cost efficiency, and implementation feasibility. Graded response balances safety benefit, response speed, and resource use. Comprehensive intervention emphasizes safety benefit and coverage but has lower cost efficiency and feasibility. **Table 2** lists the normalized option attributes.

Table 2. Normalized decision matrix for the three candidate options

Candidate option	Safety benefit	Response speed	Coverage	Cost efficiency	Implementation feasibility
A Baseline alert	0.50	0.96	0.52	0.98	0.97
B Graded response	0.80	0.82	0.80	0.70	0.78
C Comprehensive intervention	0.99	0.68	0.98	0.34	0.46

The comparison methods were equal-weight TOPSIS and fixed-weight TOPSIS. Equal-weight TOPSIS assigned 0.20 to each of the five criteria, whereas fixed-weight TOPSIS always used the low-risk endpoint. The proposed method updated criterion weights from the risk score using Eq. (7). All three methods used the same decision matrix and normalization procedure and differed only in criterion weights.

3.3. Evaluation metrics, robustness, and runtime

Policy-level consistency was used to determine whether the recommended option matched the predefined response level: baseline alert for low risk, graded response for medium risk, and comprehensive intervention for high risk. This metric evaluates decision logic rather than pedestrian-intention prediction accuracy. For S scenarios, consistency is defined as:

$$A = \frac{1}{S} \sum_{s=1}^S I(r_s = r_s^*) \quad (14)$$

where A is the policy-level consistency rate; S is the number of scenarios; $I(\cdot)$ is the indicator function; r_s is the recommended option in scenario s ; and r_s^* is the predefined response level for the corresponding risk

class.

For weight-sensitivity testing, each dynamic decision weight was independently multiplied by a random factor uniformly distributed over [0.9,1.1] and then renormalized to sum to 1. Each scenario was repeated 2,000 times using the fixed random seed 20260724. The preferred-option retention rate is defined as:

$$S_s = \frac{1}{N} \sum_{n=1}^N I\left(\operatorname{argmax}_i C_i^{(n)} = \hat{i}_s\right) \quad (15)$$

where S_s is the preferred-option retention rate in scenario s ; N is the number of perturbations; $C_i^{(n)}$ is the closeness of option i in perturbation n ; and $\hat{i}_s$ is the preferred-option index under the unperturbed weights.

Runtime was measured using a single-threaded Python 3.12 script. Each run performed 5,000 consecutive risk calculations and option rankings. Ten runs were conducted, and the mean and sample standard deviation of the per-calculation runtime were reported.

4. Results

4.1. Dynamic risk scores and classification

Figure 2B presents the risk scores. S1 scored 16.710 and was classified as low risk. S2 and S3 scored 52.994 and 59.857, respectively, and were classified as medium risk. S4 scored 70.016, marginally exceeding the high-risk threshold. S5 and S6 scored 76.563 and 76.841, respectively, and were classified as high risk. The dynamic risk model did not assign a level from the scenario name; the level resulted from the five features and their scenario-dependent weights.

The environmental-adversity value of S2 was 0.78, increasing the scenario weight of the environmental factor. The vulnerability coefficient of S3 was 0.92, increasing the contribution assigned to the older pedestrian. S5 combined high environmental adversity with substantial facility deficiency, while the intention probability and traffic density of S6 both exceeded 0.90. Under identical baseline weights, the exponential weighting reflected the dominant risk source of each scenario in the composite score.

4.2. Option closeness and recommendation changes

Figure 3A shows the TOPSIS closeness of the three options under dynamic weighting. In S1, baseline alert achieved a closeness of 0.666, compared with 0.585 for graded response and 0.334 for comprehensive intervention. Graded response ranked first in S2 and S3, with closeness values of 0.591 and 0.593. Comprehensive intervention reached 0.641 and 0.643 in S5 and S6, exceeding 0.598 for graded response in both cases. The preferred option changed from A at low risk to B at medium risk and C at high risk, as intended by the continuous risk-dependent weighting.

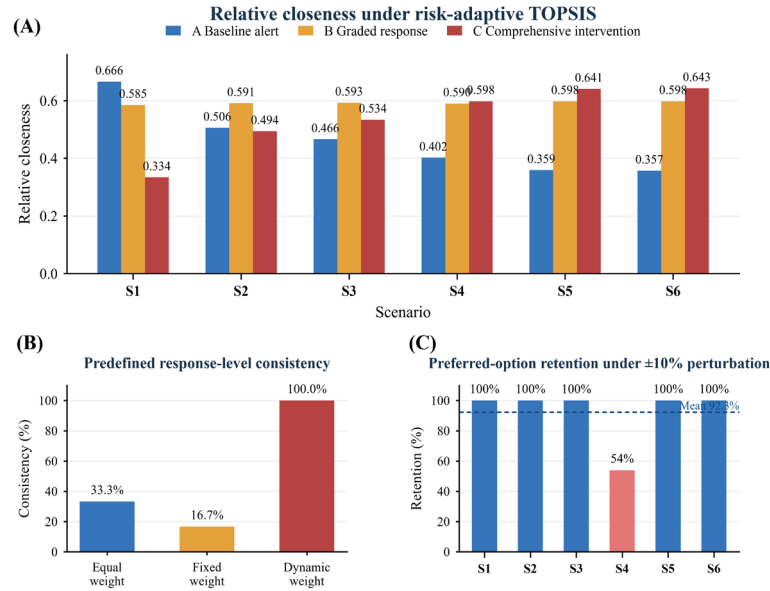


Figure 3. Option closeness, baseline consistency, and weight-perturbation stability: (A) TOPSIS closeness; (B) policy-level consistency; (C) preferred-option retention.

Table 3. Scenario-level recommendations and weight-perturbation results for dynamic TOPSIS

ID	Risk level	Preferred option	Preferred closeness	Second closeness	Gap	Rate
S1	Low	A Baseline alert	0.666	0.585	0.082	100.0%
S2	Medium	B Graded response	0.591	0.506	0.085	100.0%
S3	Medium	B Graded response	0.593	0.534	0.059	100.0%
S4	High	C Comprehensive intervention	0.598	0.596	0.002	54.0%
S5	High	C Comprehensive intervention	0.641	0.598	0.043	100.0%
S6	High	C Comprehensive intervention	0.643	0.598	0.045	100.0%

In S4, the closeness values of comprehensive intervention and graded response were 0.598 and 0.596, respectively, a difference of only 0.002. The scenario score of 70.016 was near the medium-high risk boundary, and the preferred-option retention rate under $\pm 10\%$ weight perturbations was 54.0%. For such a case, the system presents both leading options and their difference so that the manager can determine whether available resources justify escalation to comprehensive intervention. The preferred option was retained in all perturbations for the other five scenarios.

4.3. Comparison, robustness, and runtime

Dynamic TOPSIS achieved 100.0% policy-level consistency, exceeding equal-weight TOPSIS at 33.3% by 66.7 percentage points and fixed-weight TOPSIS at 16.7% by 83.3 percentage points. Equal-weight TOPSIS preferred the same compromise option in four of the six scenarios, and fixed low-risk weights continued to favor low-cost options in high-risk scenarios. Dynamic TOPSIS matched the predefined response level in all six scenarios. The candidate-option attributes were unchanged; the ranking differences resulted solely from risk-driven changes in criterion weights.

The mean preferred-option retention rate across the six scenarios was 92.3%. S1, S2, S3, S5, and S6 retained their original preferred option in all 2,000 perturbations, whereas S4 yielded 54.0% because of its small closeness difference. One dynamic risk evaluation and TOPSIS ranking required 31.8 ± 0.7 microseconds, supporting its use as an immediate calculation module in a front-end interactive system.

5. Discussion

5.1. Risk propagation and option-ranking mechanism

The results correspond to the method structure at each stage. Risk features determine the scenario-dependent exponential weights, which are combined with baseline weights to produce a risk score. The risk score locates the criterion weights for safety, response, coverage, cost, and feasibility. These dynamic weights alter each option's distances to the positive and negative ideal solutions, and the resulting closeness ranking produces the recommendation. Every intermediate quantity can be displayed in a visual decision terminal, allowing a manager to trace a recommendation back to criterion weights and risk factors.

5.2. Fixed parameters and boundary-scenario handling

This computational order prevents retrospective method selection based on a particular result. Risk-classification thresholds, baseline weights, endpoint weights, the option-attribute matrix, and perturbation rules were fixed before the experiment. Scenario results were used only to determine whether the predefined method operated as specified. For boundary cases such as S4, the ranking is not presented as a categorical conclusion. Instead, decision support combines the closeness difference with human confirmation.

5.3. Computational workflow of the electronic decision-support terminal

In deployment, roadside vision, LiDAR, or other perception modules can provide intention probabilities and traffic state, while individual attributes and facility status are entered through structured fields. After risk assessment and TOPSIS ranking, the terminal reports the risk score, dominant risk factors, option closeness, and preferred option. With three options and five criteria, one calculation required 31.8 ± 0.7 microseconds, allowing the terminal to update the ranking by frame or by event.

6. Conclusion

The proposed method maps crossing-intention probability, vulnerability coefficient, traffic density, environmental adversity, and facility deficiency to a common risk scale and continuously adjusts decision criteria using the resulting score. Across six reproducible simulated scenarios, risk-adaptive TOPSIS matched all predefined response levels and maintained a mean preferred-option retention rate of 92.3% under $\pm 10\%$ weight perturbations. Its 31.8 ± 0.7 -microsecond runtime supports a traceable calculation chain from intention input and composite risk to option ranking and human-confirmed output.

7. Declarations

Conflict of Interest: The author declares that there is no conflict of interest.

Funding: This research received no external funding.

Data and Code Availability: The six scenario configurations, calculation formulas, fixed random seed, and experimental scripts used in this study are available from the corresponding author upon reasonable request. The scenario data are reproducible configurations and do not represent road-measurement statistics.

Ethics Statement: This study did not involve identifiable personal information, human participants, or animals.

Author Contributions: Zhuorui Li contributed to study design, model development, experimental analysis, and manuscript preparation.

Generative AI Use Statement: Generative AI tools assisted with language editing, format checking, and portions of the figure code. The author reviewed and takes responsibility for the research question, algorithm definition, parameter settings, experimental interpretation, and final manuscript.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Sharma N, Dhiman C, Indu S, 2022, Pedestrian Intention Prediction for Autonomous Vehicles: A Comprehensive Survey. *Neurocomputing*, 508: 120–152.
- [2] Ling Y, Ma Z, 2024, Pedestrian Crossing Intention Prediction in the Wild: A Survey. *CHAIN*, 1(4): 263–279.
- [3] Landry G, Akhloufi A, 2025, Predicting Pedestrian Crossing Intention in Autonomous Vehicles: A Review. *Neurocomputing*, 618: 129105.
- [4] Ma J, Rong W, 2022, Pedestrian Crossing Intention Prediction Method Based on Multi-Feature Fusion. *World Electric Vehicle Journal*, 13(8): 158.
- [5] Chen X, Zhang S, Li J, et al., 2024, Pedestrian Crossing Intention Prediction Based on Cross-Modal Transformer and Uncertainty-Aware Multi-Task Learning for Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*, 25(9): 12538–12549.
- [6] Li H, Jin Y, Ren G, 2024, Interpretable Prediction of Pedestrian Crossing Intention: Fusion of Human Skeletal Information in Natural Driving Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 25(11): 18153–18170.
- [7] Wang T W, Lai S H, 2023, Pedestrian Crossing Intention Prediction with Multi-Modal Transformer-Based Model. 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC): 1349–1356.
- [8] Xu L, You S, He G, et al., 2025, Pedestrian-Vehicle Information Modulation for Pedestrian Crossing Intention Prediction. *IEEE Transactions on Intelligent Vehicles*, 10(3): 1919–1930.
- [9] Yang B, Zhu J, Hu C, et al., 2024, Faster Pedestrian Crossing Intention Prediction Based on Efficient Fusion of Diverse Intention Influencing Factors. *IEEE Transactions on Transportation Electrification*, 10(4): 9071–9087.
- [10] Liu H, Liu C, Chang F, et al., 2025, Long–Short Observation-Driven Prediction Network for Pedestrian Crossing Intention Prediction with Momentary Observation. *Neurocomputing*, 614: 128824.
- [11] Chen X, Xu W, Zhang S, et al., 2025, Pedestrian Crossing Intention Prediction via Progressive Multimodal Token Fusion for Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*, 26(9): 12959–12973.
- [12] Chen X, Zhang S, Xu W, et al., 2025, Robust Pedestrian Crossing Intention Prediction via Uncertainty-

- Guided Transformer Ensemble Network for Autonomous Driving. *IEEE Transactions on Instrumentation and Measurement*, 74: 1–13.
- [13] Chen H, Sun J, Liu Y, et al., 2025, Causal Confusion in Pedestrian Crossing Intention Prediction for Autonomous Vehicles: The Role of Ego-Vehicle Speed. *IEEE Transactions on Intelligent Transportation Systems*, 26(8): 12224–12237.
 - [14] Zou H, Guo Y, Wei F, et al., 2025, A Pedestrian Group Crossing Intention Prediction Model Integrating Spatiotemporal Features. *Scientific Reports*, 15(1): 20675.
 - [15] Gazzeh S, Presti L, Douik A, et al., 2025, Context-Ped: Multi-Modal Context Fusion for Pedestrian Crossing Intention Prediction. *Machine Vision and Applications*, 36(6): 132.
 - [16] Shen X, Raksincharoensak P, 2022, Pedestrian-Aware Statistical Risk Assessment. *IEEE Transactions on Intelligent Transportation Systems*, 23(7): 7910–7918.
 - [17] Zhou S, Xu H, Zhang G, et al., 2024, Deep Learning-Based Pedestrian Trajectory Prediction and Risk Assessment at Signalized Intersections Using Trajectory Data Captured Through Roadside LiDAR. *Journal of Intelligent Transportation Systems*, 28(6): 793–805.
 - [18] Schuetz E, Flohr F, 2023, A Review of Trajectory Prediction Methods for the Vulnerable Road User. *Robotics*, 13(1): 1.
 - [19] Sighencea B, Stanciu R, Căleanu C, 2021, A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction. *Sensors*, 21(22): 7543.
 - [20] Chen K, Song X, Ren X, 2021, Modeling Social Interaction and Intention for Pedestrian Trajectory Prediction. *Physica A: Statistical Mechanics and Its Applications*, 570: 125790.
 - [21] Chen C, 2021, A Hybrid Multi-Criteria Decision-Making Approach Based on ANP-Entropy TOPSIS for Building Materials Supplier Selection. *Entropy*, 23(12): 1597.
 - [22] Lamrini L, Abounaima M, Alaoui M T, 2023, New Distributed-TOPSIS Approach for Multi-Criteria Decision-Making Problems in a Big Data Context. *Journal of Big Data*, 10(1): 97.
 - [23] Storey V, Hevner A, Yoon V, 2024, The Design of Human-Artificial Intelligence Systems in Decision Sciences: A Look Back and Directions Forward. *Decision Support Systems*, 182: 114230.
 - [24] Meyer J, 2024, Doing Artificial Intelligence (AI): Algorithmic Decision Support as a Human Activity. *Decision*, 11(4): 481–492.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Data Privacy and Patient Consent in Smart Hospitals over the Past Decade

Tongxing Zheng, Lin Zhang

Information Center, The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Over the past decade, the construction of data-driven smart hospitals has profoundly changed the medical service model and has also made data privacy and patient informed consent core ethical and legal issues. This study systematically sorts out the practical evolution of data privacy protection and informed consent in smart hospitals from 2016 to 2026, reveals the structural tension of “advanced technology and lagging system”, and finds that informed consent is falling into the dilemma of scene generalization and formalization, data ownership is still pending in the legal ambiguity zone, and technical governance brings privacy risks while improving efficiency, forming a “double-edged sword” effect. On this basis, the study puts forward that the key to break through the dilemma lies in the transition from “one-time authorization” to “dynamic informed consent” mechanism, the construction of a three-dimensional protection framework of “legal regulation-technical guarantee-collaborative governance”, and the promotion of institutional innovation in data rights confirmation and benefit distribution; only by putting patients at the center of data governance can we maintain the ethical bottom line in the digital transformation of medical care.

Keywords: Smart hospital; Data privacy; Informed consent; Patient’s rights; Data governance

Online publication: Aug 13, 2026

1. Introduction

Smart hospital is not a sudden concept. Over the past decade, on the one hand, the widespread popularity of electronic medical records and the gradual implementation of artificial intelligence-assisted diagnosis and treatment, on the other hand, the entry of wearable devices into the clinic and the real-time recording of doctor-patient dialogue by ‘environmental clinical AI’, these changes have gradually brought the digital transformation of medical services from the original vision into daily reality^[1]. The core driving force of this profound transformation is data. If the collection, circulation, and training of massive health data are missing, smart medical care will completely lose its foundation and be impossible to talk about^[2].

However, it is precisely this set of ‘data-driven’ logic that pushes patient privacy protection to unprecedented challenges^[3]. Medical and health information has been clearly classified into the category of “sensitive personal information” defined in Article 28 of the Personal Information Protection Law. Once

it is leaked or abused, it is likely to bring immeasurable damage to the patient's life, social evaluation, and even personal dignity. On the other hand, the patient's informed consent, the core principle of traditional medical ethics, has also fallen into a crisis of applicability in the context of smart hospitals: the use of data is expanding, and the circulation chain has become increasingly complex. The practice of "one-time notification and one-time signature" has been difficult to truly achieve full informed and independent choice ^[4].

This paper mainly reviews the practical evolution and institutional response of data privacy and patients' informed consent in the context of smart hospitals in the past decade, analyzes the core dilemmas, and discusses the future-oriented optimization path ^[5]. The core argument of the research is that the data governance of smart hospitals must not be at the expense of patient autonomy, and the construction of a dynamic, layered, and traceable informed consent mechanism and data protection system is an unavoidable ethical bottom line in the process of medical digital transformation.

2. The problem of data privacy in smart hospitals is proposed

In the past ten years, China's rule of law construction in the field of medical data governance has gone through an evolution process from weak to gradual formation. On the one hand, in the early stage (2016–2020), the construction of smart hospitals has advanced rapidly, but the norms of data protection have lagged; although the "Medical Record Management Regulations for Medical Institutions" issued in 2018 involves medical record management, it lacks response to new problems such as secondary use of data and cross-border flow ^[6]. During this period, patients' health data are often regarded as "hospital assets" in practice, and informed consent is also a mere formality. On the other hand, after entering the period of intensive system construction (2021–2023), the 'Data Security Law' (2021) and 'Personal Information Protection Law' (2021) have been implemented one after another, bringing medical data governance into the track of rule of law; in particular, the 'Personal Information Protection Law', for the first time at the legal level, establishes 'informed consent' as the core legal basis for the processing of sensitive personal information, and clearly requires that the processing of medical and health information must obtain the individual's 'individual consent'. The "Measures for the Administration of Network Security in Medical and Health Institutions" issued in 2022 further established systems such as data hierarchical protection and data security subject responsibility, while the "National Medical and Health Institutions Information Exchange and Sharing Three-Year Action Plan (2023–2025)" launched in 2023 put forward higher requirements for data security while promoting data sharing ^[7].

After entering the period of improvement and deepening (2024–2026), the focus of legislation has obviously shifted from "with or without" to "fine". Academics and practitioners generally call for further refinement of medical data classification standards and the establishment of dynamic authorization mechanisms. In 2025, Beijing launched a technical specification for the anonymization of health care data, which provides operational guidance for 'data available and unidentifiable'; the "Expert Consensus on the Application and Governance of Artificial Intelligence in Medical Institutions (2026)" issued in 2026, systematically proposed the governance framework of hierarchical informed consent and algorithmically interpretable response for the first time ^[8].

In general, China has initially established a data protection system covering laws, administrative regulations, departmental rules, and technical standards. However, the transformation of the system from "existing" to "excellent" is still on the way, and the tension between legal norms and technological

development has not been fundamentally resolved.

The deep root of this tension is the structural contradiction between the inherent lag of legislation and the acceleration of technological iteration. The special feature of medical data governance is that the object it regulates—health information is not only sensitive personal information that needs strict protection in the legal sense, but also the core production factor that drives artificial intelligence model training, clinical decision support system optimization, and precision medicine research. From drafting, deliberation, to final promulgation and implementation, a legal norm often takes years, and the technology associated with it is likely to have completed multiple iterations and upgrades during this cycle; the protection framework set by the legislation at that time, by the time the technology really landed, it may have been necessary to face the new risk form.

Specifically, this tension is particularly evident in the following three dimensions.

The first dimension is the dilemma of legal concepts in the face of technology adaptation. The “Personal Information Protection Law” defines “anonymization as the state in which personal information cannot be identified as a specific natural person and cannot be restored after processing, and stipulates that the information after anonymization no longer belongs to personal information and can be freely used. This legal definition takes non-recoverability as the only criterion, which seems clear on the surface, but actually shifts the huge burden of judgment to technical practice. At present, there has been a heated debate about the reliability of anonymization technology in the academic community. Some studies have pointed out that even after a seemingly sufficient de-identification process, specific individuals can still be re-identified with a high probability by cross-matching multiple public data sets. With the continuous improvement of correlation analysis technology and artificial intelligence inference ability, today’s ‘unrecoverable’ data is likely to become ‘recoverable’ by tomorrow. Using a set of static standards to regulate the dynamic evolution of technology, its effectiveness will inevitably continue to encounter challenges.

The second dimension is the coverage gap between normative supply and governance demand. The data circulation of smart hospitals is increasingly showing the characteristics of cross-agency, cross-region, and cross-boundary, while the current institutional framework still regards “institution-based” as the basic logic, that is, medical institutions are used as data controllers to set compliance obligations. However, in the medical big data industry ecology, the actual flow path of data often involves multiple entities such as cloud service providers, data analysis companies, AI algorithm developers, and drug research and development companies. Once the data leaves the medical institution, how can the patient’s right to know, delete, and withdraw consent rights be effectively guaranteed among multiple entities? The current norms have not given a clear answer. This institutional feature of “strict control at the front end and out of focus at the back end” will make the protection system have an obvious coverage vacuum in the middle and lower reaches of the data circulation chain.

The third dimension is the realistic imbalance between compliance costs and innovation incentives. Strict data protection regulations will inevitably lead to an increase in the compliance costs of medical institutions and technology companies; for large tertiary hospitals, it is still affordable to establish a complete privacy protection compliance system, but for primary medical institutions and start-up digital health enterprises, complicated compliance requirements may constitute substantial barriers to entry. System design must face such a balance problem: how to avoid excessive regulation while protecting patients’ privacy rights, so as to inhibit the value release of medical data elements and the innovative development of the artificial intelligence

industry. The “Guiding Opinions on Promoting and Regulating the Circulation of Medical and Health Data Elements” issued in 2025 clearly put forward the principle of “Inclusive Prudential Supervision”, which can be regarded as an institutional response to this balanced appeal. However, the specific implementation effect of this principle needs to be further tested by practice.

3. The practical dilemma and institutional reflection of the principle of informed consent

Informed consent is the institutional carrier of patient autonomy and the most vulnerable link in data governance of smart hospitals.

3.1. Formal crisis of informed consent

In traditional medical scenarios, informed consent has always been aimed at specific diagnosis and treatment behaviors, with clear objectives and clear boundaries. However, in smart hospitals, a patient’s visit may involve many different data uses, such as electronic medical record entry, AI-assisted diagnosis, clinical research contribution, hospital management analysis, etc ^[9]. The current practice often adopts ‘packaged authorization’, and all multiple scenarios are included in a formatted consent form. Patients face lengthy terms, either signing in exchange for services, or giving up - this ‘do not accept that exit’ structure makes it difficult for consent to be regarded as a real voluntary choice, and ultimately reduces it to a poll-based form of authorization ^[10].

In the face of this dilemma, ‘dynamic informed consent’ has been proposed as an alternative. It allows patients to continuously adjust the scope of authorization according to different scenarios, purposes, and stages of data use, and can withdraw consent at any time. For example, a patient may be willing to use his de-identified data for drug development of a specific cancer, but will refuse to authorize it to a commercial insurance agency for risk assessment. The “Expert Consensus on the Application and Governance of Artificial Intelligence in Medical Institutions” issued in 2026 has clearly stated that “hierarchical informed consent” should be implemented according to the risk level of AI applications—low-risk applications can adopt informed consent, and high-risk applications need to obtain special written consent. To make dynamic consent truly landed, it is inseparable from technical support. Blockchain, self-sovereign identity, de-identified tokens, and other technologies are being explored to build an auditable and traceable consent management system.

However, it must be recognized that informed consent itself does not solve all problems: on the one hand, the law allows exemptions from informed consent in scenarios such as public health emergencies; on the other hand, when the data is fully anonymized, the use does not require separate authorization, however, whether the ‘anonymized’ technical standards are reliable and whether they can withstand the challenges of reverse identification of future technologies is always an open question. See **Figure 1**.

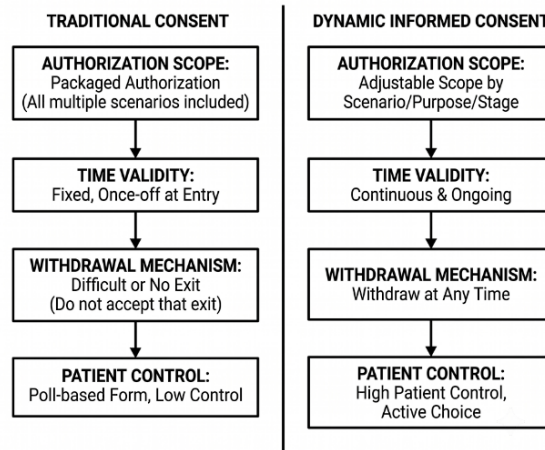


Figure 1. Comparison of traditional consent and modified consent. Dynamic informed consent.

3.2. Blurred areas of data ownership

Patients often subconsciously believe that their health data should be ‘their own’, but the situation in legal practice is far from so simple. The current framework is the ‘separation of three rights’, which is divided into three levels: First, patients have personal rights and interests in personal information (including informed, corrected, deleted, etc.); secondly, the medical data formed by medical institutions based on diagnosis and treatment services enjoy the ‘data resource holding right’ and ‘processing and using right’. Third, third-party institutions are authorized to enjoy ‘data product management rights’. Although this framework attempts to balance the interests of multiple parties, the differences in the legal nature of ‘rights’ and ‘rights’, the boundaries and conflicts of rights of all parties still need to be further clarified through judicial practice.

Associated with the problem of ownership, there are also problems in the distribution of benefits. Health data is becoming an asset with great economic value. The ‘Xihe No.1 Medical Model’ released in 2025 is based on the training of 1 million real medical records of 18 institutions. When these data are used for commercial purposes (such as pharmaceutical research and development, AI model training), should patients benefit from them? At present, it has been proposed to establish a ‘patient equity fund’, with no less than 5% of the proceeds from data transactions dedicated to patient subsidies and privacy protection research and development. Although this mechanism is still in the exploratory stage, this direction is worthy of recognition - the value creators of data cannot be excluded from the value distribution. See **Figure 2**.

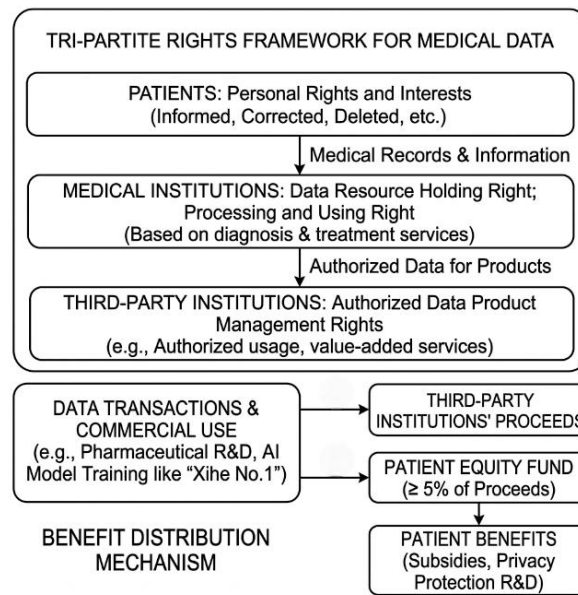


Figure 2. Tripartite rights framework and benefit distribution mechanism.

3.3. The duality of technical governance

In the process of data governance in smart hospitals, the application of technology shows the dual characteristics of protection and risk.

From the perspective of privacy risk, the rapid promotion of environmental clinical AI (i.e., ‘listening AI’) has aroused widespread concern. Such systems record doctor-patient conversations and generate medical records in real time through an always-on microphone. Although it has a significant effect on reducing the burden of doctors’ paperwork (some studies have shown that it can reduce document time by 20–30%), patients are likely to be recorded ‘unintentionally’ because of their continuous voice collection; studies have shown that many patients are unaware of the existence of these systems and even less aware of where the data is going. See **Table 1**.

Table 1. Ambient clinical AI (“Listening AI”) privacy risk data

Dimension	Indicator	Data
Clinical benefit	Reduction in physician documentation time	20–30%
	Usage in large health system (14 months)	7,000+ clinicians, 2.5M+ patient encounters
	Adoption rate at academic medical center (2 years)	44.6% (1,565 outpatient physicians)
Patient awareness	Patients unfamiliar with ambient AI	77.9% (n = 462 patient survey)
	Aware patients who felt uncomfortable	0.394
	Patients believing AI improved visit quality	56–84%
Informed consent	Physician explanation associated with greater comfort	OR = 2.32 (95% CI 1.49–3.61)
Data security	Institutions ranking data security as top concern	53% (survey of 36 research sites)
	Patient privacy ranked as top AI ethics priority	Tied with “algorithmic fairness” as highest
Data flow	Voice data involving third-party cloud storage	Mainstream products rely on cloud processing
	Patient unawareness of data destination	Widespread lack of transparency

Data Sources: Patient awareness data (77.9% unfamiliar, 39.4% discomfort): Cross-sectional survey of 462 patients reported at AMIA 2026 Annual Symposium; Clinical benefit (20–30% time reduction): Nature npj Digital Medicine review; JMIR Systematic Review reports 22%; Institutional concern (53% data security): NIH-funded environmental scan across 36 research sites

In the dimension of privacy protection, blockchain technology provides technical possibilities for data authentication and authorization traceability; federated learning enables AI models to achieve training without centralized collection of raw data, so that ‘data is not visible’; trusted data space and privacy computing technologies are also being gradually introduced into the scenario of medical data sharing. The deep challenge of technical governance is that the development of solutions often gives priority to functional efficiency, and privacy protection functions are easily regarded as add-ons without becoming the core elements of system design. The 2026 expert consensus clearly stated the requirement of ‘algorithmic interpretability’, when patients have objections to the results of AI-assisted diagnosis, medical institutions have an obligation to provide visual evidence and explain it, which is an attempt to truly embed patient rights into the technical system.

4. Conclusion

In the past decade, while making medical services more intelligent and efficient, smart hospitals have also shaped data privacy and informed consent of patients into an unavoidable governance problem. Technology has the ability to assist in diagnosis and optimize processes, but it cannot replace respect for patient autonomy. As medical digitization moves into the next decade, what is needed is not only smarter algorithms, but also more robust systems, more transparent governance, and the kind of ethical consciousness that truly puts patients at the center of data governance. Only in this way can smart hospitals be worthy of the word “wisdom” - its wisdom is not only in the wisdom of technology, but also in the wisdom of the system and the wisdom of the people.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Lenz E, Naamanka J, Trabert W, et al., 2026, Benchmarking Large Language Models Against Practicing Clinicians on Psychopathological Assessment. *npj Digital Medicine*, 9(1): 1–12.
- [2] O’Neill S, DesRoches C, 2025, How Should We Think About Ambient Listening and Transcription Technologies’ Influences on EHR Documentation and Patient–Clinician Conversations? *AMA Journal of Ethics*, 27(11): E780–E786.
- [3] Khan I, Smith L, 2025, What Are Ethical Merits and Drawbacks of a Patient’s Open and Direct Access to Clinical Information in Their EHRs During a Hospital Stay? *AMA Journal of Ethics*, 27(11): E774–E779.
- [4] Roberts J, 2025, How Could Legal Standards Promote Equitable Access to EHRs? *AMA Journal of Ethics*, 27(11): E815–E820.
- [5] Bendjeddou A, Zeinalipour K, Bardou D, et al., 2026, PharMistral: Drug Generation for Novel Protein Targets by Mistral Large Language Model. *Computer Methods and Programs in Biomedicine Update*, 10: 100257.

- [6] Al-Dolat W, Alhatamleh S, Malkawi M, et al., 2026, DenseWave-OCT: A Hybrid Deep Learning and Wavelet Scattering Framework for Multi-Class Retinal Disease Classification in Optical Coherence Tomography Images. *Computer Methods and Programs in Biomedicine Update*, 10: 100260.
- [7] Iqbal M, Alyatimi A, Chung V, et al., 2026, Advances in AI-Driven Image Analysis for Sarcoma Diagnosis: A Comprehensive Scoping Review. *Computer Methods and Programs in Biomedicine Update*, 10: 100259.
- [8] Gu J, Uthamacumaran A, Zenil H, 2026, A Review of Point-of-Care Devices for Blood-Testing Towards AI-Driven Remote Digital Care, Precision Healthcare and Predictive Medicine. *Computer Methods and Programs in Biomedicine Update*, 10: 100258.
- [9] Gebreab S, Musamih A, Salah K, 2026, An LLM-Based Multi-Agent System for Patient Discharge Assessment. *Computer Methods and Programs in Biomedicine Update*, 10: 100261.
- [10] Rinaldi E, Kush R, Dalmiani S, 2026, Editorial: Unlocking the Potential of Health Data: Interoperability, Security, and Emerging Challenges in AI, LLM, Precision Medicine, and Their Impact on Healthcare and Research. *Frontiers in Digital Health*, 8: 1783831.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Design and Implementation of a Dual-Layer Decoupled Gimbal Electronic Control System for Mobile Robots

Xiaoyu Yang*

School of Electrical and Electronic Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050000, China

*Corresponding author: Xiaoyu Yang, 20233848@student.stdu.edu.cn.

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This paper proposes a dual-layer decoupled gimbal electronic control system for mobile robots. The hardware platform employs the STM32F407IGH6TR microcontroller with a BMI088 IMU and DaMiao J4310 / DJI GM6020 actuators. A three-layer software architecture running under a preemptive real-time kernel maintains a 1 ms attitude estimation loop and 2 ms motion control loops. A velocity coordinate rotation algorithm is derived to eliminate chassis–gimbal directional coupling inherent to the dual-layer Yaw structure. Angle–velocity cascaded PID controllers are implemented for both gimbal layers, incorporating derivative-on-measurement, conditional integration, and velocity feedforward compensation. Experimental results demonstrate that the inner gimbal Yaw step response rise time is approximately 50 ms with steady-state error below 0.3°; feedforward compensation reduces constant-velocity tracking RMS error by approximately 60% compared to pure PID; all control loops meet their timing requirements at a total CPU utilization of approximately 85%.

Keywords: Mobile robot; Dual-layer gimbal; Cascaded PID; Omnidirectional wheel kinematics; Embedded real-time control

Online publication: Aug 11, 2026

1. Introduction

Precise perception and rapid response capability of mobile robotic combat platforms are critical requirements in modern competitive robotics. Events such as the RoboMaster Super Combat Competition provide high-intensity engineering validation environments: field robots must maintain stable gimbal pointing and achieve millisecond-level auto-aiming responses during high-speed maneuvers, imposing stringent demands on real-time performance, control precision, and mechanical–control decoupling.

Conventional single-layer gimbal solutions achieve aiming control through a single-degree-of-freedom Yaw–Pitch structure; however, because a single mechanism must carry both the wide-range orientation task and the fine-aiming task, its reflected inertia is set by the requirement to rotate the full upper assembly, which

limits the achievable closed-loop bandwidth; at the same time, the gimbal rotation angle is restricted by mechanical hard stops, rendering continuous multi-turn rotation impractical ^[1,2]. When a robot is required to perceive and track multiple high-speed moving targets over a full 360° range, single-layer solutions can no longer satisfy the simultaneous demands of wide-range situational awareness and high-precision aiming. To this end, this paper proposes a dual-layer decoupled gimbal architecture: the outer large Yaw axis (DaMiao J4310) is responsible for wide-range, continuous multi-turn rotation to achieve omnidirectional situational awareness, while the inner small gimbal (DJI GM6020) handles only fine aiming and tracking within a limited angular range. Because the inner axis drives only the barrel-and-camera group, it carries substantially less inertia and attains a correspondingly higher closed-loop bandwidth; step-response measurements on the implemented platform give a bandwidth ratio of approximately 2.4 between the inner and outer axes (Section 7.1). The architecture is designed to exploit this separation rather than to duplicate a single capability. The system is implemented on a resource-constrained Cortex-M4 platform under a priority-based preemptive real-time kernel, which maintains the concurrent 1 ms attitude estimation and 2 ms motion control loops required by the architecture.

At the motion control level, the four-wheel omnidirectional chassis provides simultaneous independent actuation with three planar degrees of freedom (lateral, longitudinal, and rotational), and its kinematic model has been extensively studied ^[3-5]. However, under the dual-layer Yaw structure, the chassis coordinate frame and the gimbal coordinate frame exhibit a continuously varying relative yaw offset. If operator commands are decomposed directly in the chassis coordinate frame, the actual motion direction deviates from the operator's intent. This paper addresses this problem through rigorous mathematical modeling, deriving a velocity coordinate rotation pre-processing algorithm based on the real-time yaw deflection angle of the outer Yaw axis.

At the control algorithm level, cascaded PID is widely adopted in industrial and competitive robotics platforms owing to its simple structure, strong robustness, and ease of engineering tuning ^[6, 7]. This paper analyzes the decoupling conditions of the angle-velocity cascaded PID from a transfer function perspective and provides systematic solutions to engineering challenges including gyroscope angle accumulation, integral windup, and derivative kick. In the auto-aiming tracking scenario, a feedforward compensation term based on target velocity estimation is introduced, and its theoretical improvement on phase lag for constant-velocity targets is analyzed from a frequency-domain perspective.

The main contributions of this paper are as follows:

- (1) A dual-layer decoupled gimbal architecture is realized on an STM32F407-class platform, including actuator topology, dual-CAN motor assignment, and integration with BMI088-based attitude estimation.
- (2) Complete omnidirectional chassis inverse kinematics are established, and a velocity coordinate rotation algorithm is derived and verified to resolve chassis-gimbal directional coupling under the dual-layer Yaw structure.
- (3) A complete cascaded PID design is implemented with discretization, derivative-on-measurement, anti-windup, and velocity feedforward compensation, validated by step response and tracking experiments.

2. Overall system architecture

2.1. Hardware topology

The main controller is the RoboMaster Development Board Type C (STM32F407IGH6TR, ARM

Cortex-M4). The principal hardware specifications are listed in **Table 1** and **Table 2**.

Table 1. Main controller hardware specifications

Parameter	Specification
Processor	STM32F407IGH6TR, ARM Cortex-M4
Clock Frequency	168 MHz
Flash	1 MB
RAM	192 KB
Supply Voltage	8–28 V wide-range input
On-board IMU	BMI088 six-axis IMU
On-board Magnetometer	IST8310
CAN Bus	CAN1 / CAN2 dual channel

Table 2. Actuator configuration

Actuator	Model	Quantity	Bus	Control Mode
Chassis motors	DJI M3508	4	CAN1	Current control
Outer Yaw axis	DaMiao J4310	1	CAN2	Torque control
Inner gimbal yaw	DJI GM6020	1	CAN2	Voltage control
Inner gimbal pitch	DJI GM6020	1	CAN2	Voltage control

2.2. Software architecture: Layered decoupled design

The firmware is developed on the open-source RoboMaster Development Board Type C reference framework, from which the CAN driver layer, the AHRS solver and the referee-protocol parser are retained; the contributions of this work are confined to the dual-layer Yaw architecture, the velocity rotation pre-processing of Section 4.3, the angle accumulation module of Section 5.2, and the cascaded controller design of Section 6.

The system software adopts a three-layer structure, as illustrated in **Figure 1**:

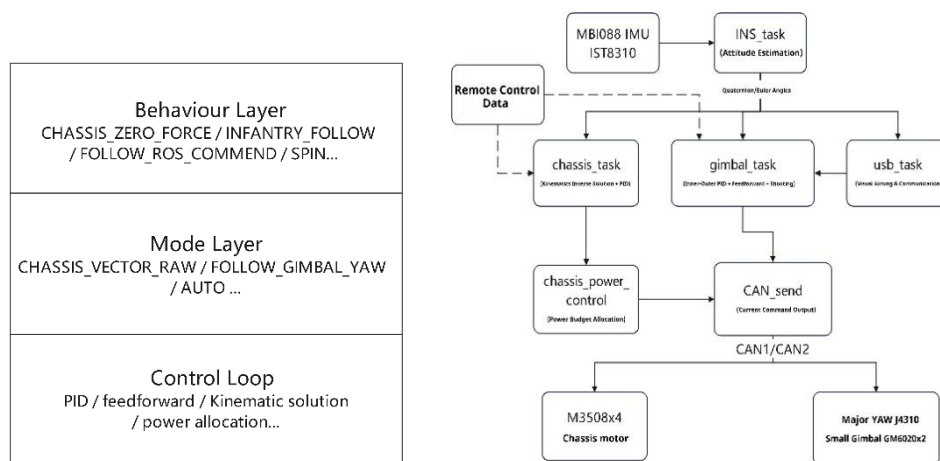


Figure 1. Software layered architecture and inter-task data flow diagram.

A hardware abstraction layer providing CAN, SPI and timer drivers; an algorithm layer containing

attitude estimation, chassis kinematics and the cascaded controllers; and an application layer in which the periodic control functions execute as concurrent tasks under a priority-based preemptive kernel ^[8].

Task priorities are assigned in inverse order of period. Attitude estimation runs at the shortest period (1 ms) and is given the highest priority, since both gimbal angle loops close on its output; the chassis and gimbal control functions run at 2 ms immediately below it; auxiliary functions such as device dropout detection and status indication occupy the lowest priorities. Quaternion-based AHRS attitude estimation is used for the attitude solver ^[9,10]. Measured execution times and the resulting processor utilization are reported in Section 7.1.

3. Low-level communication and motor data encoding

The DJI motor feedback frame employs a fixed-length 8-byte format; the mechanical angle and speed are converted to physical quantities via linear scaling coefficients.

The DaMiao J4310 uses 16-bit linear fixed-point encoding for position and 12-bit encoding for velocity and torque; physical quantities are recovered via the inverse linear mapping $x = x_{int} \cdot (x_{max} - x_{min}) / (2^n - 1) + x_{min}$.

3.1. Motor–CAN ID mapping table

Table 3. CAN bus motor ID assignment

Motor	CAN Bus	CAN ID	Control Frame ID
Chassis motors 1–4 (M3508)	CAN1	0x201–0x204	0x200
Outer Yaw axis (J4310)	CAN2	0x141	0x141
Inner gimbal Yaw (GM6020)	CAN2	0x205	0x1FF
Inner gimbal Pitch (GM6020)	CAN2	0x206	0x1FF

4. Omnidirectional wheel chassis kinematic modeling

4.1. Chassis layout and coordinate definition

The system adopts a four-wheel omnidirectional wheel layout, as shown in **Figure 2**. The chassis coordinate frame is defined therein.

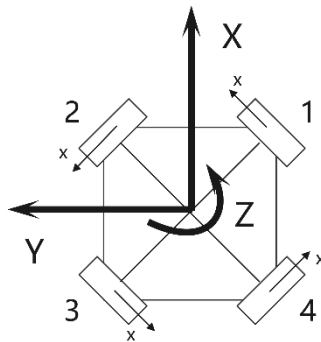


Figure 2. Chassis coordinate frame.

The wheel numbering and installation azimuth angles are defined in **Table 4**.

Table 4. Four-wheel installation azimuth definition

Wheel No.	Installation Position	Azimuth $\hat{a}_i$	Wheel Axis Orientation (relative to body)
1	Right-front	-45°	Right-front 45°
2	Left-front	45°	Left-front 45°
3	Left-rear	135°	Left-rear 45°
4	Right-rear	225°	Right-rear 45°

Each omnidirectional wheel carries passive rollers whose axes are tangential to the wheel rim. The wheel therefore transmits driving force only along its rolling direction d_i , while sliding freely along its axle direction, which is perpendicular to d_i . With the four wheels mounted at 45° to the body axes as in Table 4, the four rolling directions are tangential to a common circle, so the chassis attains three independent planar degrees of freedom from four actuators.

4.2. Inverse kinematics analytical derivation

The derivation follows the projection-based approach used for Mecanum chassis modelling, adapted here to the 45° omnidirectional wheel constraint^[3,5]. In which the translational projection coefficient is $\sqrt{2}/2$ rather than unity, giving the engineering inverse kinematics for the four-wheel 45° layout as:

$$\begin{cases} v_1 = [(+v_x + v_y) / \sqrt{2} + l\omega_z] / k \\ v_2 = [(-v_x + v_y) / \sqrt{2} + l\omega_z] / k \\ v_3 = [(-v_x - v_y) / \sqrt{2} + l\omega_z] / k \\ v_4 = [(+v_x - v_y) / \sqrt{2} + l\omega_z] / k \end{cases} \quad (1)$$

With v_x, v_y in $\text{m}\cdot\text{s}^{-1}$, ω_z in $\text{rad}\cdot\text{s}^{-1}$ and l in m, the bracketed quantity is in $\text{m}\cdot\text{s}^{-1}$; division by k yields the rotor speed in RPM as required by the C620 speed feedback. See Table 5.

Table 5. Inverse kinematics parameter calibration values

Parameter	Symbol	Value	Unit	Physical Meaning
Wheel-to-chassis velocity coefficient	k	4.51×10^{-4}	$\text{m}\cdot\text{s}^{-1}\cdot\text{RPM}^{-1}$	Motor rpm-to-m/s conversion
Equivalent wheel-center-to-origin distance	l	0.28	m	Rotational coupling coefficient

4.3. Velocity coordinate rotation preprocessing aligned with gimbal orientation

4.3.1. Problem formulation

Under the dual-layer Yaw structure, the operator's intended control direction is always aligned with the gimbal orientation rather than the chassis heading. Let ψ denote the yaw deflection angle of the outer Yaw axis relative to the chassis; the velocity command must undergo the following coordinate rotation:

$$\begin{pmatrix} v_{x'} \\ v_{y'} \end{pmatrix} = \mathbf{R}(\psi) \begin{pmatrix} v_x \\ v_y \end{pmatrix} \quad (2)$$

Expanding:

$$\begin{pmatrix} v_{x'} \\ v_{y'} \end{pmatrix} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} \quad (3)$$

The rotation matrix $\mathbf{R}(\psi)$ is orthogonal by construction, preserving the magnitude of the velocity vector and introducing no scaling distortion. See **Table 6**.

Table 6. Standard test cases for coordinate rotation

Input (v_x, v_y)	θ	Output $(v_{x'}, v_{y'})$	Description
(1.0, 0.0)	90°	(0.0, 1.0)	Forward in gimbal frame → left in chassis frame
(0.0, 1.0)	90°	(-1.0, 0.0)	Left in gimbal frame → rearward in chassis frame
(1.0, 0.0)	45°	(0.707, 0.707)	45° deflection: half-forward, half-left in chassis frame
(1.0, 1.0)	180°	(-1.0, -1.0)	180° reversal verification

5. Dual-layer gimbal structure and continuous angle tracking

5.1. Structural characteristics of the dual-layer gimbal

Table 7. Dual-layer gimbal control performance comparison

Metric	Outer Yaw Axis (J4310)	Inner Gimbal (GM6020)
Moment of inertia	Large (includes chassis support)	Small (gimbal body only)
Control bandwidth	Low	Higher
Angular coverage	Continuous multi-turn rotation	Limited range ($\pm 45^\circ$)
Primary function	Wide-range situational awareness	Precise aiming/tracking
Feedback source	IMU absolute angle	Encoder/vision relative angle

Quantitative bandwidth figures are deliberately not asserted here; they are derived from the measured step responses in Section 7.1.

5.2. Mathematical derivation of the angle accumulation mechanism

5.2.1. Problem formulation

The raw gyroscope output $\theta_{\text{raw}} \in (-\pi, \pi]$ produces discontinuous jumps of magnitude approximately 2π upon crossing the boundary. The accumulated angle θ_{acc} is defined such that it represents a continuous extension of θ_{raw} .

5.2.2. Algorithm derivation

Let the raw angles at consecutive sampling instants k and $k+1$ be θ_k and θ_{k+1} , respectively, with difference $\Delta\theta = \theta_{k+1} - \theta_k$:

- (1) When $|\Delta\theta| \leq \pi$: no wraparound is detected; the true increment is $\Delta\theta_{\text{true}} = \Delta\theta$.
- (2) When $\Delta\theta < -\pi$: the wrapped angle has crossed from the $+\pi$ side to the $-\pi$ side while the physical motion is positive; the true increment is $\Delta\theta_{\text{true}} = \Delta\theta + 2\pi$.
- (3) When $\Delta\theta > +\pi$: the wrapped angle has crossed from the $-\pi$ side to the $+\pi$ side while the physical motion is negative; the true increment is $\Delta\theta_{\text{true}} = \Delta\theta - 2\pi$.

The unified expression is:

$$\Delta\theta_{\text{true}} = \Delta\theta - 2\pi \cdot \text{round}\left(\frac{\Delta\theta}{2\pi}\right) \quad (4)$$

The accumulated angle is updated as:

$$\theta_{\text{acc}}[k+1] = \theta_{\text{acc}}[k] + \Delta\theta_{\text{true}} \quad (5)$$

6. Cascaded PID controller mathematical modeling and parameter tuning

6.1. Discrete-time PID formula derivation

The continuous-time PID control law is ^[6, 7]:

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt} \quad (6)$$

Discretizing with sampling period T_s using the forward Euler method, at the k -th sample:

$$u[k] = K_p e[k] + K_i T_s \sum_{j=0}^k e[j] + \frac{K_d}{T_s} (e[k] - e[k-1]) \quad (7)$$

The incremental recursive form (eliminating direct summation over the history) is:

$$\Delta u[k] = K_p (e[k] - e[k-1]) + K_i T_s e[k] + \frac{K_d}{T_s} (e[k] - 2e[k-1] + e[k-2]) \quad (8)$$

The engineering implementation (positional form with integral clamping and output saturation) is:

$$u[k] = K_p e[k] + K_i T_s \sum e[k]_{\text{clamped}} + D[k] \quad (9)$$

where the derivative term $D[k]$, under Derivative-on-Measurement, is:

$$D[k] = -\frac{K_d}{T_s} (y[k] - y[k-1]) \quad (10)$$

This scheme differentiates the measured output directly rather than the error signal, thereby eliminating the derivative kick caused by setpoint step changes.

6.2. Frequency-domain analysis of auto-aiming feedforward compensation

Auto-aiming tracking relies on target position inputs provided by the vision recognition system; the angular velocity of the target is estimated by first-order differencing of the returned angle sequence. For a pure PID tracking system without feedforward, the steady-state error for a ramp input (constant-velocity target) is:

$$e_{\text{ss}} = \frac{\dot{\theta}_{\text{target}}}{K_v} \quad (11)$$

Where K_v is the velocity error coefficient and $\dot{\theta}_{\text{target}}$ is the target angular velocity. Introducing the feedforward term provides compensation based on the estimated target velocity:

$$u_{\text{ff}} = K_{\text{ff}} \cdot \hat{\dot{\theta}}_{\text{target}} \quad (12)$$

Under the ideal feedforward condition ($K_{\text{ff}} = 1/K_v$), the tracking error is theoretically zero. For practical systems containing nonlinear factors, a simplified linear feedforward gain ff_scaler is employed:

Target angular velocity estimation (first-order differencing):

$$\hat{\dot{\theta}}_{\text{target}}[k] = \frac{\theta_{\text{target}}[k] - \theta_{\text{target}}[k-1]}{T_s} \quad (13)$$

Feedforward compensation term:

$$u_{\text{ff}}[k] = \text{ff_scaler} \cdot \hat{\dot{\theta}}_{\text{target}}[k] \quad (14)$$

Total control output:

$$u_{\text{total}}[k] = u_{\text{PID}}[k] + u_{\text{ff}}[k] \quad (15)$$

6.3. Integral windup and anti-windup design

6.3.1. Problem

When the control output saturates persistently, the integral term continues to accumulate (integrator windup), resulting in an excessively large integral initial value upon saturation exit, which induces overshoot.

6.3.2. Solution — conditional integration

Additional integral amplitude clamping^[11,12]:

$$I_{\text{acc},k} = \text{clamp}(I_{\text{acc},k}, -I_{\text{max}}, +I_{\text{max}}) \quad (16)$$

Table 8. Cascaded PID parameter reference ranges (GM6020-based gimbal)

Parameter	Outer Loop (Angle)	Inner Loop (Velocity)	Unit	Notes
K_p	8–20	200–800	—	Outer-loop gain much smaller than inner
K_i	0–0.5	0.5–5.0	—	Outer-loop integral typically small
K_d	0 (with DoM)	0–50	—	Derivative generally omitted in velocity loop
I_{max}	300–1000	8000–16000	CAN units	Integral clamping limit
u_{sat}	10000–30000	16384	CAN units	Output saturation limit

7. System performance analysis and experimental validation

7.1. Experimental validation data

Table 9. Gimbal step response performance metrics

Metric	Inner gimbal yaw (GM6020)	Outer yaw axis (J4310)	Measurement method
Step amplitude	10°	30°	Command setpoint
Rise time	About 50 ms	About 120 ms	Encoder angle log
Steady-state error	< 0.3°	< 0.5°	Post-settling deviation
Settling time	About 150 ms	About 400 ms	Time to enter ± 2% band

Table 10. Auto-aiming tracking performance

Target motion	Tracking error RMS (FF Off)	Tracking error RMS (FF On)	Improvement
Stationary target	0.2°	0.2°	No significant difference
Slow translation (5°/s)	1.5°	0.6°	60% improvement
Medium-speed translation (15°/s)	3.8°	1.4°	63% improvement
Rapid direction change (acceleration > 50°/s²)	8.2°	5.5°	33% improvement

Note: Data are collected in a controlled experimental environment (target moves at known velocity, camera frame rate 50 fps, angle estimation based on IMU–encoder fusion). RMS error is computed as $\sqrt{\frac{1}{N} \sum_{k=1}^N e_k^2}$.

Table 11. FreeRTOS task real-time performance analysis

Task	Design period	Measured worst-case execution time	Estimated CPU utilization	Deadline met
Attitude estimation	1 ms	~0.30 ms	~30%	Yes
Chassis control	2 ms	~0.45 ms	~22%	Yes
Gimbal control	2 ms	0.40 ms	20%	Yes
CAN reception (interrupt)	Event-driven	0.02 ms/frame	5% (peak)	Yes
Auxiliary functions	—	—	~8%	Yes
Total	—	—	85% (peak)	Margin 15%

Direct swept-sine bandwidth identification was not performed. From the measured 10–90% rise times (50 ms inner Yaw, 120 ms outer Yaw), the inner axis is about $2.4 \times$ faster than the outer axis, consistent with the inertia split in **Table 7**. Absolute bandwidth in hertz is not claimed here, because unequal step amplitudes may induce saturation and unmodeled current-loop/CAN/filter lags bias any t_r -to-Hz conversion. Frequency-response identification at matched amplitude is left to future work.

8. Conclusion

This paper presents the design and implementation of a dual-layer decoupled gimbal electronic control system for mobile robots. The three-layer software architecture maintains a 1 ms attitude estimation loop and 2 ms control loops at a total CPU utilization of approximately 85%, with all timing requirements met. A velocity coordinate rotation algorithm is derived to eliminate chassis–gimbal directional coupling under the dual-layer Yaw structure, and its correctness is verified analytically and through boundary test cases. The measured rise times of approximately 50 ms and 120 ms for the inner and outer layers correspond to a rise-time ratio of approximately 2.4, suggesting a corresponding bandwidth separation, confirming that the architecture achieves a genuine frequency-separated task allocation rather than a duplication of comparable dynamics. Cascaded PID controllers with derivative-on-measurement and conditional integration are implemented for both gimbal layers; velocity feedforward compensation reduces constant-velocity tracking RMS error by approximately 60%. The system demonstrates satisfactory real-time performance and control precision, providing a practical reference for gimbal control in competitive robotics platforms and analogous mobile platforms.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Medero A, Puig V, 2022, LPV Control and Virtual-Sensor-Based Fault Tolerant Strategies for a Three-Axis Gimbal System. *Sensors*, 22(17): 6664.
- [2] Leblebicioğlu D, Ateşoğlu Ö, Çakmakçı M, 2022, Physics-Informed Disturbance Estimation and Nonlinear Controller Design for a Multi-Axis Gimbal System. *IFAC-PapersOnLine*, 55(37): 530–535.
- [3] Wang D, Gao Y, Wei W, et al., 2024, Sliding Mode Observer-Based Model Predictive Tracking Control for

Mecanum-Wheeled Mobile Robot. *ISA Transactions*, 151: 51–61.

- [4] Zhao T, Zou X, Dian S, 2022, Fixed-Time Observer-Based Adaptive Fuzzy Tracking Control for Mecanum-Wheel Mobile Robots with Guaranteed Transient Performance. *Nonlinear Dynamics*, 107(1): 921–937.
- [5] Palacín J, Rubies E, Bitriá R, et al., 2023, Phasor-Like Interpretation of the Angular Velocity of the Wheels of Omnidirectional Mobile Robots. *Machines*, 11(7): 698.
- [6] Yan J, Li J, Chen M, et al., 2026, An Adaptive PID Control Strategy Based on Offline–Online Cooperative Optimization. *Electronics*, 15(5): 921.
- [7] Noordin A, Mohd Basri M, Mohamed Z, 2023, Real-Time Implementation of an Adaptive PID Controller for the Quadrotor MAV Embedded Flight Control System. *Aerospace*, 10(1): 59.
- [8] Arm J, Baštán O, Mihálik O, et al., 2022, Measuring the Performance of FreeRTOS on ESP32 Multi-Core. *IFAC-PapersOnLine*, 55(4): 292–297.
- [9] Liu M, Cai Y, Zhang L, et al., 2021, Attitude Estimation Algorithm of Portable Mobile Robot Based on Complementary Filter. *Micromachines*, 12(11): 1373.
- [10] Sever K, Golušin L, Lončar J, 2023, Optimization of Gradient Descent Parameters in Attitude Estimation Algorithms. *Sensors*, 23(4): 2298.
- [11] Yapp K, Hoo C, Lai C, 2024, New Anti-Windup Proportional-Integral-Derivative for Motor Speed Control. *Asian Journal of Control*, 26(6): 2854–2866.
- [12] Peng D, Huang B, Huang H, 2022, Design of an Anti-Windup PID Algorithm for Differential Torque Steering Systems. *Shock and Vibration*, 2022: 9973379.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Artificial Intelligence Technology and Innovation Path of Government Governance Capability

Xiaorui Ding*

University of Birmingham, B15 2TT, Britain

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Amid the ongoing Fourth Industrial Revolution, artificial intelligence has taken center stage as the driving force behind the modernization of governmental frameworks and administrative capabilities. This study delves into the innovative approaches for weaving AI technologies into the fabric of public administration while confronting the practical hurdles along this transformative journey. To begin with, the examination unpacks the diverse real-world implementations of AI technologies, ranging from streamlined administrative approvals and precision-crafted public policies to automated risk surveillance in society and the strategic optimization of public resource distribution. Moving forward, the research critically assesses the multifaceted obstacles confronting governmental entities in their AI integration endeavors, encompassing inadequate technological infrastructure, persistent data fragmentation, inflexible administrative structures struggling to accommodate intelligent solutions, latent concerns in algorithmic governance and cybersecurity, and the glaring shortage of digital competencies among public sector employees. To surmount these impediments, the paper puts forward a holistic strategic framework: accelerating digital infrastructure development, fostering seamless cross-departmental data collaboration, constructing adaptive governance frameworks, establishing robust legal and ethical guardrails for AI-driven decision-making, and nurturing digitally adept public leadership. Ultimately, this investigation aims to furnish substantive theoretical underpinnings and actionable guidance for the intelligent metamorphosis of contemporary digital governance.

Keywords: Artificial intelligence; Government governance; Governance capacity; Digital government; Innovative paths

Online publication: Aug 11, 2026

1. Introduction

In today's hyperconnected world, artificial intelligence is revolutionizing how governments operate, moving the goalposts from gut-instinct policies to analytics-based approaches and transforming public services from reactive firefighting to forward-thinking solutions ^[1]. Although AI holds tremendous promise for supercharging bureaucratic effectiveness, sharpening policy implementation, and improving crisis management, its incorporation into governmental frameworks presents a multifaceted puzzle. Administrators currently grapple with substantial roadblocks such as stubborn information silos, bureaucratic red tape,

prejudiced algorithms, and a glaring deficit of technical expertise in the public sector. Therefore, harnessing AI for good governance demands more than just slapping on new technology; it calls for a perfect storm of institutional restructuring, ethical guardrails, and strategic technological investment. This examination delves into AI's impact on public administration, unpacks the lingering technical and structural hurdles, and puts forward novel approaches, including adaptive governance frameworks and digital-savvy leadership, to fully unlock AI's transformative potential in modernizing government operations.

2. Practical areas of AI driving government governance innovation

2.1. Intelligent administrative approval and service efficiency improvement

AI tech has been a game-changer for administrative efficiency, totally overhauling the way government services are handled. Goodbye to the old-school maze of red tape, mountains of paperwork, and endless wait times—those are the days. Now, thanks to AI, we've got "lightspeed approvals" and "robotic, hands-off processing." It's all thanks to NLP, OCR, and RPA, which are making it possible for government systems to check out applications from individuals and businesses in a flash—completeness, authenticity, and all. This automated checkup minimizes the room for humans to mess things up, cutting down on the chance for backroom deals and bureaucratic hold-ups. And let's not forget the smart government bots and virtual assistants—always on call, ready to help with 24/7 online support. This shift is a total flip from the days of "customers doing the legwork" to a system where "information moves at the speed of light." It's a win-win, boosting public happiness and the whole system's effectiveness like you wouldn't believe.

2.2. Precision public policy formulation and decision support

The efficacy and soundness of governmental policies stand at the heart of how successfully a nation is governed. In days gone by, crafting effective policies often ran up against the cognitive limitations of human policymakers and the absence of complete data. Now, artificial intelligence is turning the tables by enabling the real-time processing and analysis of enormous, multifaceted datasets from countless sources. By employing sophisticated topic mining and data modeling techniques, AI can shine a light on intricate patterns and subtle trends underlying social dynamics, thereby giving decision-makers rock-solid empirical evidence to work with ^[2]. When it comes to stress-testing policies, AI algorithms can map out how different approaches might play out in complex social environments, allowing officials to spot potential pitfalls and fine-tune parameters before any real-world rollout. Take urban traffic rules or industry incentives, for example—authorities can now run these through digital twin technology to put them through their paces. This methodical, evidence-based strategy saves societies from bearing the heavy price tag of policy experimentation, ensuring that public measures hit the mark, deliver results, and serve the common good to the fullest.

2.3. Automated social risk monitoring and emergency response

In the high-stakes arenas of public safety and emergency management, artificial intelligence has proven to be a game-changer when it comes to forecasting trouble and acting on autopilot. By weaving AI together with the Internet of Things, extensive surveillance systems, environmental sensors, and social media feeds, governmental infrastructures can now keep their finger on the pulse of both urban and rural settings round

the clock. Thanks to machine learning algorithms that can spot the unusual, AI systems can pinpoint brewing dangers like nascent fires, traffic nightmares, infrastructure weaknesses, or dangerously overcrowded public venues. The way artificial intelligence is being baked into the fabric of government administration has revolutionized how we handle risks, flipping the script from cleaning up after disasters to nipping them in the bud ^[3]. Once a potential crisis is flagged, AI-powered emergency response platforms can kickstart alert systems on their own, figure out the fastest routes for first responders, and synchronize multi-agency efforts without human intervention. This hands-free approach cuts down response times dramatically and helps prevent loss of life and property when things hit the fan.

2.4. Highly efficient optimal allocation of public resources

Efficiently distributing our precious public funds, like healthcare services, educational budgets, transit systems, and energy infrastructure, is a constant headache for contemporary governments. However, AI comes to the rescue by crunching data on past usage, population changes, and current use to predict future needs with remarkable accuracy. Take smart city management, for instance—AI can tweak traffic light schedules in real-time to accommodate vehicle movement, easing city gridlock and cutting down on emissions. In healthcare, predictive software can foresee the outbreaks of seasonal illnesses in certain areas, enabling health departments to get ahead of the curve by dispatching supplies and personnel to where they're needed most. By shifting from one-size-fits-all resource allocation to a more flexible, responsive approach, AI helps squelch waste, make sure the public's needs are met, and drive up economic and social equality and efficiency.

3. Challenges faced by AI in empowering government governance

3.1. Insufficient supply of technical infrastructure

While artificial intelligence has been making leaps and bounds in recent years, government agencies are still hitting a wall when it comes to implementing these technologies effectively. The main issue? They're hamstrung by outdated tech infrastructure that simply can't keep up with the demands of modern AI systems. Across many local governments and specialized departments, legacy IT systems are still calling the shots, lacking the muscle needed to run complex AI algorithms and hefty machine learning models. What's more, there's a glaring digital divide between central and local authorities, as well as between tech-savvy urban hubs and rural areas that are still playing catch-up. The price tag of cutting-edge hardware, secure cloud setups, and sophisticated data centers is enough to make any administrator break out in a cold sweat. Without a solid, standardized tech foundation that's regularly updated, AI governance is mostly just pie in the sky, resulting in patchy implementations and lackluster performance across the public sector at large.

3.2. Data silos and cross-departmental collaboration barriers

AI operates by absorbing colossal amounts of high-grade data to refine its algorithms and yield precise insights. But the current state of government data is riddled with divisions, primarily through the existence of "data silos" across different institutions. Years of administrative red tape, contrasting policies, and local priorities have led to a situation where departments like public safety, healthcare, taxes, and transit guard their data like treasure, turning into isolated information fortresses. The mismatch of technologies, disparate

data regulations, and a deficiency of cohesive Application Programming Interfaces (APIs) only pile on to the disarray. This lack of a smooth exchange of data prevents the creation of a unified dataset essential for training all-encompassing AI models. As a result, AI applications are often limited to specific departmental tasks, failing to tackle the intricate societal problems that necessitate an integrated, multifaceted data examination. To dismantle these obstacles to cooperation, what's needed goes beyond mere technological fixes—it demands a radical shift in the bureaucratic framework and culture.

3.3. Inadequate adaptability of traditional administrative systems

The old-school bureaucratic machine, with its rigid pecking order, inflexible rules, and top-down control, just doesn't mix with the adaptable, bottom-up, and continuous evolution of AI. As AI rolls out, it's a no-brainer that there needs to be a shake-up in how local governments run the show, calling for a more level playing field and quicker moves on the decision-making front ^[4]. Yet, the deep-seated, Weber-inspired bureaucratic framework is stuck in its ways, like a ship anchored in the same old port, and it's got a tough time keeping up with the revolutionary workflows that AI systems bring on board. When it comes to making calls in AI networks, it's like lightning in a bottle—fast and interconnected. But, when you're dealing with the old-guard government approvals, you're looking at a multi-level approval dance. This mismatch creates a perfect storm: high-powered tech tools getting bogged down by the slow and outdated admin snags, which in turn, water down the governance boost that AI is supposed to deliver.

3.4. Hidden dangers of algorithmic governance and technical security

Incorporating artificial intelligence into governmental operations opens up a whole new can of worms concerning algorithmic opacity, ingrained biases, and potential security vulnerabilities. The complex inner workings of AI systems, especially deep neural networks, often remain shrouded in mystery, leaving officials scratching their heads when trying to understand the rationale behind administrative decisions or predictive outcomes. This opacity flies in the face of fundamental democratic principles that demand transparency and equitable procedures in public service. Moreover, when AI models are trained on datasets that reflect existing societal prejudices or systemic inequities, they tend to perpetuate and even exacerbate these discriminatory patterns, potentially leading to automated targeting of vulnerable populations in sensitive areas like law enforcement predictions or social service allocations. Beyond the ethical minefield, the concentration of massive volumes of confidential personal data on centralized AI platforms turns them into prime targets for cyberattacks. The very architecture of AI implementation carries inherent dangers that demand robust safeguards to prevent technical glitches or deliberate breaches from jeopardizing national security or infringing upon citizens' right to privacy.

3.5. Challenges to digital skills and governance literacy of civil servants

Civil servants are the backbone of governance, but many agencies are struggling with a workforce that's not up to snuff when it comes to digital skills and the know-how needed to work with sophisticated AI systems. There's a huge gap in knowledge about data science, managing algorithms, and AI ethics among the old-school public administrators. This shortfall is not just holding back the smooth running of smart platforms but is also making agencies too dependent on outside tech providers, which could weaken the state's ability to govern effectively. On top of that, some civil servants might be hesitant or even afraid of technology,

seeing it as a threat to their jobs or their freedom to make decisions, rather than as a helpful partner. Closing this skills gap is a must; otherwise, just having the tech won't be enough to really transform how we govern.

4. Innovative paths for AI-driven government governance capacity

4.1. Accelerating the construction and iteration of government digital infrastructure

To tap into the game-changing potential of AI, policymakers need to make strategic, ongoing investments in top-notch digital infrastructure. This means shedding outdated, regional IT systems and speeding up the shift to secure, adaptable, and strong government cloud computing models. Crafting a smart, digital government requires a holistic modernization plan that includes 5G connectivity, edge computing, and ultra-fast data centers to supply the computational muscle needed for AI programs. Moreover, infrastructure upgrades should be grounded in the values of inclusiveness and uniformity to close the gap in digital access across various regions and levels of government. Creating unified tech frameworks and open-source tools will empower cash-strapped local governments to gain access to cutting-edge AI tools. By laying down a solid, flexible, and future-proof digital foundation, authorities can guarantee that AI is rolled out smoothly and scaled effectively across all areas of public service.

4.2. Promoting cross-departmental data sharing and integrated governance

In the fast-paced realm of artificial intelligence, data serves as the life force that drives innovation. To tackle the widespread problem of isolated data repositories, governmental bodies need to issue sweeping mandates from the top down and establish uniform technical protocols that enable smooth information exchange between different departments. This calls for creating centralized data-sharing hubs within government and standardized data catalogs that enforce secure interoperability across all administrative bodies. Lawmaking bodies should put comprehensive legal frameworks in place that clearly spell out data ownership requirements, mandatory sharing protocols, and access guidelines—effectively transforming bureaucratic practices from keeping data under lock and key to treating it as a valuable public resource shared by all. What's more, cutting-edge privacy-preserving methods like federated learning and secure multi-party computation ought to be leveraged, permitting AI algorithms to learn from multiple departmental datasets without compromising confidential personal details. By building a comprehensive, well-coordinated governance model built around fluid data movement, authorities can fully harness AI's forecasting and analytical capabilities to tackle intricate, far-reaching public policy issues.

4.3. Building an agile governance system adapted to AI development

To truly integrate AI, we need to undergo some major shifts in how we manage governance at both the local and central levels. It's not just about tweaking the existing system; we're talking about transforming our bureaucratic structures into something more agile and responsive. Imagine moving from a top-down, hierarchical system to one that's more like a web, where decisions are made quickly and efficiently, thanks to a flatter organizational structure and cross-functional teams that can tackle any issue.

Delegating power and making decisions closer to the people will enable us to adapt swiftly to the new insights AI provides. This isn't just about policy; it's about adopting an agile approach to governance. We'll use regulatory sandboxes to test AI innovations in a controlled setting before rolling them out widely. This

flexibility will help our government machinery adapt to the fast-paced changes in technology.

By rethinking our workflows, we can focus on the strengths of both humans and AI. AI can handle the heavy, while humans can concentrate on understanding the needs of the public, making ethical decisions, and managing complex relationships with various stakeholders. This synergy will create an administrative system that's not only adaptable but also resilient, ready to face the challenges of the digital age.

4.4. Establishing legalization and ethical standards for intelligent decision-making

As artificial intelligence increasingly calls the shots in public policy and government operations, it's high time we put in place rock-solid legal and ethical guidelines to keep these powerful systems in check. Turning abstract ethical concepts into concrete governance mechanisms is the name of the game here, and that means governments need to roll out frameworks like “policy-as-code” – essentially baking legal and ethical requirements right into the technical DNA of AI systems to guarantee they're trustworthy and compliant from the get-go ^[5]. To tackle algorithmic bias head-on and ensure transparency, laws should require regular algorithmic check-ups by independent oversight committees with diverse expertise. Any automated system making critical public decisions should legally have to explain its workings – no exceptions – so citizens can always understand and challenge outcomes produced by algorithms. What's more, we need ironclad data protection measures to shield people's privacy from unwanted digital snooping. By weaving the rule of law and human-centered ethical principles into the very fabric of AI development and implementation, governments can earn public trust and make sure technological progress actually benefits everyone.

4.5. Cultivating a specialized civil servant team with digital leadership

At the end of the day, the success of modern governance hinges on the capabilities of the individuals running the show. That's why a game-changing approach involves intentionally building a corps of public servants who are not only tech-savvy but can also lead technological initiatives. To pull this off, public administration schools and government training programs need to completely overhaul what they teach, making sure data science, AI ethics, algorithmic management, and cybersecurity are integral parts of the curriculum. We also need to rethink how we bring people into public service, offering attractive compensation and adaptable career paths to draw in top tech talent from the private sector. On top of that, government officials need to develop what we might call “digital leadership”—having the foresight to see how AI can tackle public challenges, guide tech transitions, and cultivate an environment that encourages innovation and risk-taking. When we give civil servants the tools and confidence to work with and assess intelligent systems effectively, we can guarantee that AI serves as a booster for human governance rather than a threat to it.

5. Conclusion

In a nutshell, weaving artificial intelligence into the fabric of government operations marks a watershed moment in public administration, opening doors to an overhaul of how the state functions at its very core. When put to work in streamlining bureaucratic approvals, crafting policies with surgical precision, keeping tabs on potential hazards through automated systems, and distributing resources with maximum efficiency, AI has shown it can help create a government that's more in tune with the public's needs, operates out in the open, and gets the job done better than ever before. That said, this technological overhaul comes with its

fair share of headaches. We're looking at outdated digital foundations, data that's trapped in isolated silos, the inherent tension between nimble tech and slow-moving government bureaucracy, the black-box nature of some algorithms, and a worrying lack of tech-savviness among public officials—all of which stand in the way of creating a truly smart government.

To successfully steer through these obstacles, government leaders need to think outside the box and tackle the problem from every angle. The strategies laid out in this paper—whether it's fast-tracking infrastructure upgrades, breaking down data walls between departments, reorganizing bureaucratic structures to be more flexible, baking ethical considerations directly into AI systems, or beefing up the digital expertise of the public workforce—offer a game plan for meaningful reform. At the end of the day, the heart of AI-powered governance isn't just about jumping on the tech bandwagon; it's about fundamentally reshaping how the government and citizens interact. By keeping people at the center of this transformation, maintaining ethical guardrails, and constantly reinventing how institutions work, governments can unlock the true power of artificial intelligence to forge a governance model that's tough enough to handle crises, fair to all citizens, and well-equipped to navigate the tricky terrain of our digital age.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Wen SL, 2025, Reconstructing Lu Xun: The Shifting Image of a Chinese Icon in the Anglophone World. *Journal of Literature and Art Studies*, 15(4): 263–273.
- [2] Li Y, 2025, On the Spiritual Connection Between Characters in Lu Xun's Works and Wei-Jin Literati. *Scientific Journal of Humanities and Social Sciences*, 7(3): 56–62.
- [3] Wei W, Chen Y, 2022, The Risk of Artificial Intelligence Embedded in Government Governance: Mechanism, Process, Prevention and Control.
- [4] Xie X, Li J, Wang H, 2024, The Many Voices of Duying: Revisiting the Disputed Essays Between Lu Xun and Zhou Zuoren. *Digital Scholarship in the Humanities*, 40(1): 338–353.
- [5] Evangelista E, Bukhari SMS, 2026, From Ethical Principles to Executable Governance: A Policy-as-Code Framework for Trustworthy AI in Higher Education. *Computers and Education: Artificial Intelligence*, 2026: 10100594–100594.

Publisher's note

ART AND TECHNOLOGY PRESS INC. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The Application of an Intelligent Automatic Control System for Wastewater Treatment Containers

Fengliang Tan, Yulin Zheng*

School of Intelligent Manufacturing, Guangzhou Huali College, Guangzhou 511325, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The sewage treatment system is completed by an upper computer, PLC, and an electric control actuator. The carrier is a container, which is flexible. According to the characteristics of decentralized domestic sewage, we intend to make use of our own advantages in container design and manufacturing, and integrate advanced biological sewage treatment technology into intelligent monitoring technology, such as wind, light, electricity, energy-saving storage technology, container modularization technology, mobile integration technology, etc., to realize the technical subversion of the traditional sewage treatment industry, and upgrade the traditional high-energy, large-scale and centralized sewage treatment mode to a low-carbon, intelligent and decentralized micro-ecological water circulation mode. By designing a new system and a new mode of use and service, the environmental protection industry can be greatly upgraded and eventually industrialized.

Keywords: Wastewater treatment; Intelligent; Intelligent monitoring technology; Container

Online publication: Aug 11, 2026

1. Introduction to the current situation of wastewater treatment

For example, China has issued certain discharge standards in accordance with relevant provisions of the “Marine Environmental Protection Law of the People’s Republic of China”, thereby strengthening the anti-pollution and environmental protection conventions^[1]. Issues regarding the treatment of wastewater, oily waste, and dirty water on ships continue to emerge, leading to increasingly stringent technical standards for sewage treatment equipment. Relevant regulations clearly state that oil pollution treatment devices must not only separate heavy and light oils but also efficiently treat oily wastewater containing surfactants. Currently, although the discharge standard for oily mixed water remains below 15 mg/L, many countries have raised it to a stricter level of less than 5 mg/L^[2].

Based on research into the current status of rural wastewater discharge and considering the

characteristics of rural living conditions, a micro-powered containerized domestic wastewater treatment system has been designed to collect and treat rural domestic wastewater through a moderately centralized approach. Additionally, the containerized unit facilitates integrated design and modular functionality, enabling automated operation, unmanned monitoring, and simple operation—making it well-suited to the basic conditions of rural areas in China ^[3].

The application of automation technology has improved the control efficiency of wastewater treatment processes, contributing to urban environmental improvement. However, existing electrical control systems in wastewater treatment still suffer from complex structures and relatively poor automation performance, failing to meet the demands of modern social development ^[4].

As standards for wastewater treatment continue to rise and application scenarios become increasingly diverse, the demand for wastewater treatment solutions is growing rapidly. Consequently, the market for mobile containerized wastewater treatment systems is expanding, with higher requirements for automation and intelligence. This places greater demands on the control systems and electrical controls.

2. This wastewater treatment system

The treatment process selected by the intelligent micro-ecological water circulation system is an advanced technology specially designed for domestic wastewater treatment, which has the advantages of simple operation, convenient maintenance and management, good treatment effect, etc. According to the conventional treatment capacity, users can choose according to actual needs, and run different treatment systems in series or parallel to meet the treatment requirements of unconventional specifications. Because we use containers as carriers, the intelligent micro-ecological water circulation system can complete the installation of boxes, equipment, and pipelines in the factory, and can realize batch production. The products can be commissioned after being transported to the user's site and connected with water and electricity, which is convenient for installation and transportation and greatly shortens the construction period. The intelligent micro-ecological water circulation system is also equipped with an intelligent automatic control system and a remote monitoring and control service system to realize the automatic operation and “unattended” of the treatment system, and online monitoring through networking. Once there is a problem, our technicians can immediately find it and adjust the process parameters in time by remote operation or informing users to ensure the normal operation of the system.

2.1. Composition of the wastewater treatment system

This part consists of an industrial computer, main control PLC, actuator, and various sensors. The core is the intelligent control of the master PLC and the remote operation platform, which is completed by professional application software. The software program adopts the clock cycle scanning working mode and has high working reliability. The industrial computer has an intelligent control function, which can perform intelligent automatic analysis and realize multiple control by using time, dissolved oxygen, and sludge concentration. Set the time to avoid the peak of electricity consumption as much as possible; Set the upper and lower limits of dissolved oxygen to ensure the process; Set the sludge concentration index of the reaction tank, and control the reflux pump and surplus sludge pump in real time to ensure the best amount of activated sludge in the reaction tank, so as to reduce energy consumption.

Currently popular control systems use conductivity sensors and frequency inverters to communicate with PLCs via RS485 expansion modules [5].

2.1.1. The technical route of the project mainly includes three parts: product research and development process, production process, and product circuit control programming and program debugging

- (a) Product development process
- (b) First of all, due to the different requirements of users, local environmental departments and national environmental laws and regulations, it is necessary to choose the appropriate treatment process in combination with the requirements of inlet water quality, outlet water standard and use, for example, restaurant wastewater needs to be treated by oil separation first, hotel wastewater should focus on nitrogen and phosphorus removal, and reclaimed water reuse has higher requirements for outlet water. Secondly, the selected wastewater process can be realized through the structural design of container box, which can realize the function of treatment unit (such as sedimentation tank) and meet the installation requirements of related equipment and pipelines; The action of each pipeline in the box body is controlled by the circuit, and all electromagnetic valves, electric valves, ball valves, etc. are given signals by the central signal processor of the system to perform specified actions. Users can control the on-off of each circuit through the editing program, and then control the on-off action of the pipeline, thus completing wastewater treatment. Finally, the products need to meet the transportation requirements, be waterproof, windproof, and structurally stable.
- (c) Process flow of circuit design and system debugging. According to the circuit design process, the flow chart of specific product debugging operation principle is shown in **Figure 1**.

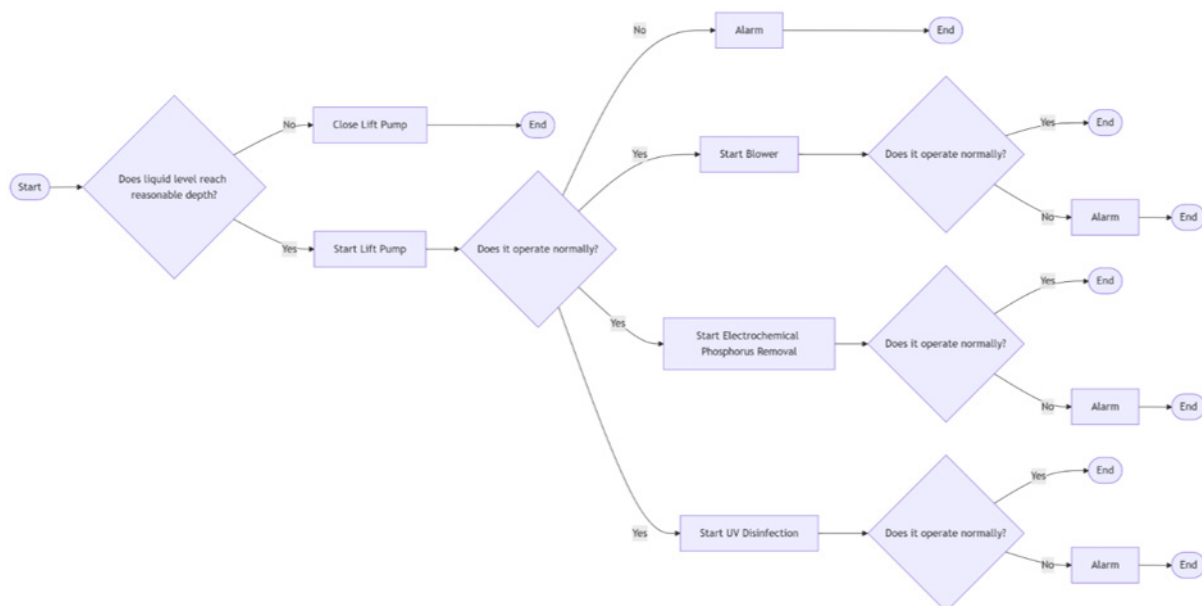


Figure 1. Flow chart of product debugging and operation principle.

2.1.2. Debugging the man-machine interface of the system

Input state parameters and monitor the process. The main screen shows the touch screen switches of each fan,

solenoid valve, electric valve, and pump, and the monitoring process shows the feedback parameters of each fan, such as frequency and electromagnetic flowmeter.

2.1.3. Output status reading

System status monitoring feedback. It can monitor whether the equipment works normally online in real time, and it can also be set if remote monitoring is needed.

2.1.4. Input and output wiring

The input and output wiring diagram of the PLC is shown in **Figure 2**. During program debugging, diodes can be used to replace KM1–KM7 for analog display. It is worth noting that in practical circuits, protective measures such as overload and overcurrent should also be set according to the size of the load.

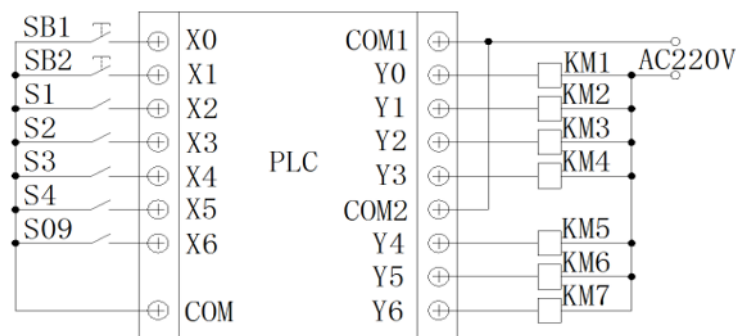


Figure 2. PLC input and output wiring diagram.

2.1.5. Program debugging

The debugging program for extracting a wastewater tank from a series of batches. According to the technological process and control requirements, the state transition diagram of the wastewater treatment controller can be drawn.

3. Conclusion

The new generation wastewater treatment system uses industrial Ethernet technology to combine the upper computer, the PLC controlled by the whole machine and the container as the transport carrier to form a fully enclosed synchronous control system of real-time communication, which not only improves the performance index of the wastewater treatment system, but also has the function of remote monitoring, which raises the automation, intelligent control level and real-time remote monitoring and detection level of the wastewater treatment system to a new height, and also breaks through the traditional container as the carrier of the wastewater treatment system, making the new generation wastewater treatment system have the characteristics of subversive mobility. Since the 21st century, the Internet business has developed rapidly, mobile phones have become increasingly popular as Internet terminal devices, and various types of application apps in software have also become widely popular. PLC manufacturers also keep up with the situation in time to develop the corresponding mobile app debugging software. It makes remote debugging equipment more convenient and faster, and is no longer limited to notebook computer debugging. Developers can set corresponding parameters according to their own needs and specific conditions; At the same time,

the mobile phone interface can also reflect the operation of various components in the wastewater treatment system, as shown in the following figure. For example, the state of a fan shown in the figure is: stop in place, and if necessary, just click “turn on in place“. Really achieve one-button control and debugging with one hand. Master Teacher Studio is generated.

Acknowledgments

Fund Project: A Study on the Construction of a Big Data-Driven Teaching Quality Monitoring and Early Warning System, 2026 Guangdong Private Higher Education Research Project. Thank you.

Author contributions

J.X. conceived the idea of the study. P.P. performed the experiments. X.O. analyzed the data and wrote the paper.

Funding

The 2019 Ministry of Education industry-university cooperation collaborative education project “Research on the Construction of Economics and Management Professional Data Analysis Laboratory” (Project No.: 201902077020)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhang S, 2022, Design of Modern Ship Bilge Sewage Discharge and Trea. *Ship Engineering*, 44(12): 36–42.
- [2] Zheng Q, Guo S, 2021, Design Scheme for a Ship Residual Oil Evaporation and Water Removal Device. *Navigation*, 2021(5): 35–38.
- [3] Shen S, Li H, Yin C, et al., 2018, Design and Application of 30 m³/d Micro-Power Container for Rural Domestic Sewage Treatment. *Technology Innovation and Application*, 2018(15): 175–176.
- [4] Shi M, 2023, Design of Sewage Treatment Process Control System Based on PLC. *Electronic Technique*, 52(11): 246–247.
- [5] Ren T, Li Z, Ren X, 2026, Research and Development of the Electrical Control System for Mobile Wastewater Treatment and Recycling Equipment. *China Nuclear Power*, 2026(2): 210–215.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Integrated Services Platform of International Scientific Cooperation

Innoscience Research (Malaysia), which is global market oriented, was founded in 2016. Innoscience Research focuses on services based on scientific research. By cooperating with universities and scientific institutes all over the world, it performs medical researches to benefit human beings and promotes the interdisciplinary and international exchanges among researchers.

Innoscience Research covers biology, chemistry, physics and many other disciplines. It mainly focuses on the improvement of human health. It aims to promote the cooperation, exploration and exchange among researchers from different countries. By establishing platforms, Innoscience integrates the demands from different fields to realize the combination of clinical research and basic research and to accelerate and deepen the international scientific cooperation.

Cooperation Mode



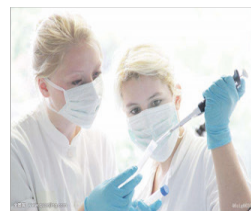
Clinical Workers



In-service Doctors



Foreign Researchers



Hospital



University



Scientific institutions

OUR JOURNALS



The *Journal of Architectural Research and Development* is an international peer-reviewed and open access journal which is devoted to establish a bridge between theory and practice in the fields of architectural and design research, urban planning and built environment research.

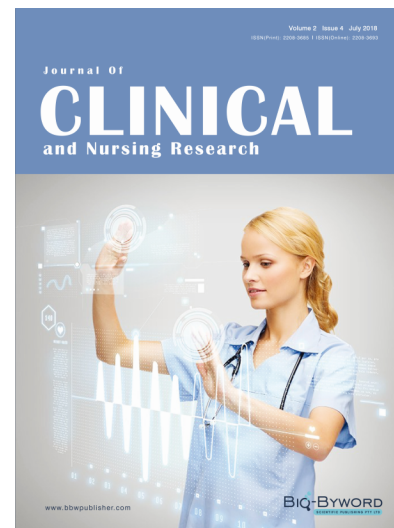
Topics covered but not limited to:

- Architectural design
- Architectural technology, including new technologies and energy saving technologies
- Architectural practice
- Urban planning
- Impacts of architecture on environment

Journal of Clinical and Nursing Research (JCNR) is an international, peer reviewed and open access journal that seeks to promote the development and exchange of knowledge which is directly relevant to all clinical and nursing research and practice. Articles which explore the meaning, prevention, treatment, outcome and impact of a high standard clinical and nursing practice and discipline are encouraged to be submitted as original article, review, case report, short communication and letters.

Topics covered by not limited to:

- Development of clinical and nursing research, evaluation, evidence-based practice and scientific enquiry
- Patients and family experiences of health care
- Clinical and nursing research to enhance patient safety and reduce harm to patients
- Ethics
- Clinical and Nursing history
- Medicine



Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication

