Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Efficient Control of Mechatronic Systems Enabled by Generative AI for Single-Chip Microcomputers

Hang Xu*, Yao Mai

Zhuhai College of Science and Technology, Zhuhai 519041, Guangdong, China

**Author to whom correspondence should be addressed.*

Abstract: In recent years, research on industrial innovation and development has primarily focused on industrial automation and intelligent manufacturing. Within the field of integrating mechatronics and intelligent control, analyzing the efficient control of mechatronic systems enabled by generative AI for single-chip microcomputers can further highlight the value and significance of promoting AI technology applications. This paper examines the technical characteristics of generative AI in data generation, multimodal fusion, and dynamic adaptation, proposing lightweight model deployment strategies that compress large generative models to a range compatible with single-chip microcomputers, ensuring local real-time inference capabilities. It constructs an edge intelligent control architecture, enabling generative AI to directly participate in decision-making instruction generation, forming a new working system of perception, decision-making, and execution. Additionally, it designs a collaborative optimization training mechanism that leverages federated learning to overcome single-machine data limitations and enhance model generalization performance. At the application level, an intelligent fault prediction system is developed for early identification of equipment anomalies, an adaptive parameter optimization module is constructed for dynamically adjusting control strategies, and a multi-device collaborative scheduling engine is established to optimize production processes, providing technical support for embedded intelligent control in Industry 4.0 scenarios.

Keywords: Generative AI; Single-chip microcomputer; Mechatronic system; Efficient control

Online publication: December 16, 2025

1. Introduction

With continuous innovation in science and technology, the field of industrial automation is undergoing a crucial phase of transition from digitization to intelligence. As the core carrier, the control accuracy and response speed of mechatronic systems are closely related to production efficiency [1]. Traditional single-chip microcomputer control solutions, based on preset rules and fixed parameters, exhibit certain limitations when dealing with complex scenarios such as nonlinear operating conditions and multivariable coupling [2]. Therefore, to innovate and optimize the overall control performance of the system and enhance technological capabilities, it is essential to emphasize the introduction and application of generative AI technology. This technology achieves creative mapping of data distributions through probabilistic modeling and possesses the ability to extract deep features from vast amounts of industrial data [3].

This study aims to explore the technical pathway for the deep integration of generative AI with single-chip microcomputers, address the challenges of deploying generative models on resource-constrained devices, and establish an intelligent control framework tailored for mechatronic systems. Through innovations in lightweight model compression, edge computing architecture design, and distributed training methods, dynamic generation and real-time optimization of control strategies can be achieved. This results in a comprehensive solution encompassing fault prediction, parameter tuning, and production scheduling, effectively enhancing the control performance of mechatronic systems and facilitating their intelligent upgrade and transformation.

1.1. Analysis of the technical characteristics of generative AI

Generative AI technology can leverage deep learning models to capture underlying patterns from vast amounts of data, construct probabilistic distribution models, and generate new data and control strategies that meet specific requirements [4]. Taking Variational Autoencoders (VAEs) as an example, this model compresses input data into low-dimensional distribution parameters (mean and variance) in a latent space through an encoder. The decoder then samples from this latent distribution and reconstructs the data, as follows:

$$P(x|z) = N\left(x \middle| f\theta(z), \sigma^2 I\right) \tag{1}$$

Equation (1) represents the reconstructed data; is the latent variable; and is the decoder's nonlinear transformation function. This enables microcontrollers to dynamically generate optimized parameters for motor control, replacing traditional fixed PID adjustment methods.

Generative Adversarial Networks (GANs) achieve iterative optimization of control strategies through adversarial training between a generator and a discriminator. The objective function of the generator is as follows:

$$\min_G \max_D V(D,G) = E_{x-pdata(x)}[\log D(x)] + E_{z-p\ (z)}[\log(1-D(G(z)))] \tag{2}$$

This mechanism allows microcontrollers to autonomously generate anti-interference control instructions under complex operating conditions. In mechatronic control, for robotic arm trajectory tracking, the discriminator evaluates deviations between the generated trajectory and the ideal trajectory in real time, driving the generator to adjust joint angle outputs and achieve micrometer-level positioning accuracy, thereby enabling efficient control [5].

2. Technical methods for empowering microcontrollers with generative AI

2.1. Lightweight model deployment strategies

Deploying generative AI models on microcontrollers faces strict constraints in terms of computational resources, storage space, and energy consumption, necessitating efficient operation through model lightweighting techniques. Therefore, it is essential to consider from the perspectives of algorithm optimization and hardware-software co-design. In algorithm optimization, model compression techniques can be employed to reduce parameter precision and shrink model size. For instance, converting FP32 parameters to INT8 can reduce model size by 75% and increase inference speed by 3-5 times [6].

The Infineon PSoC 6 series of microcontrollers integrates a hardware quantization accelerator, enabling direct on-chip 8-bit integer arithmetic operations. When combined with pruning techniques to remove redundant neurons, this allows the YOLO Tiny object detection model to achieve real-time processing capabilities of 15 FPS on the STM32H7 series [7]. Knowledge distillation techniques transfer knowledge from complex models to lightweight models by constructing a teacher-student model architecture. MobileNetV3 achieves an accuracy rate of up to 75% on the ImageNet dataset, with a parameter count that is only 1/32 of that of ResNet-50.

In hardware architecture design, the goal is to strengthen the physical support for model deployment. The deep integration of the RISC-V architecture with AI coprocessors is a major trend. The Canaan K210 chip integrates a KPU neural network accelerator, supporting an energy efficiency ratio of 1 TOPS/W, enabling it to run facial recognition models with a power consumption of less than 1W. The STM32N6 series of microcontrollers incorporates the TensorFlow Lite for Microcontrollers runtime library, providing optimized convolutional and pooling operators that reduce the inference energy consumption of TinyML models to the microwatt level. For 8-bit microcontrollers, algorithm engineers have developed binary neural networks (BNNs), which restrict weights and activation values to 1, enabling the model to achieve handwritten digit recognition on 51-series microcontrollers with an accuracy rate maintained above 90%.

2.2. Edge intelligence control architecture

In edge intelligence computing, its control framework reconstructs the data interaction mode between microcontrollers and the cloud, forming a closed-loop control chain that integrates perception, decision-making, and execution [8]. In industrial robot collaborative control scenarios, edge nodes adopt a hierarchical architecture design, where the underlying microcontrollers are responsible for motor driving and sensor data acquisition, the mid-level edge gateways run lightweight generative AI models, and the top-level cloud performs model training and global scheduling. The Siemens S7-1500 series PLC integrates the Edge Impulse platform, supporting the deployment of vibration anomaly detection models locally, reducing fault identification latency from 500 ms in cloud-based modes to 20 ms.

In the edge intelligence control architecture, the real-time operating system (RTOS) needs to be optimized. FreeRTOS ensures the temporal determinism of AI inference tasks and control tasks through task priority scheduling and memory partition management. In the flight control system of unmanned aerial vehicles (UAVs), the RTX5 operating system assigns the highest priority to model inference tasks and, combined with the DMA data transmission mechanism, stabilizes the attitude calculation cycle within 2 ms. For resource-constrained scenarios, μC/OS-III introduces a time-triggered architecture (TTA) to guarantee the timing of multiple tasks through static scheduling tables, achieving stable hovering of quadrotor UAVs in MSP430 microcontroller control, which is efficient control. The overall architecture is shown in **Figure 1**.

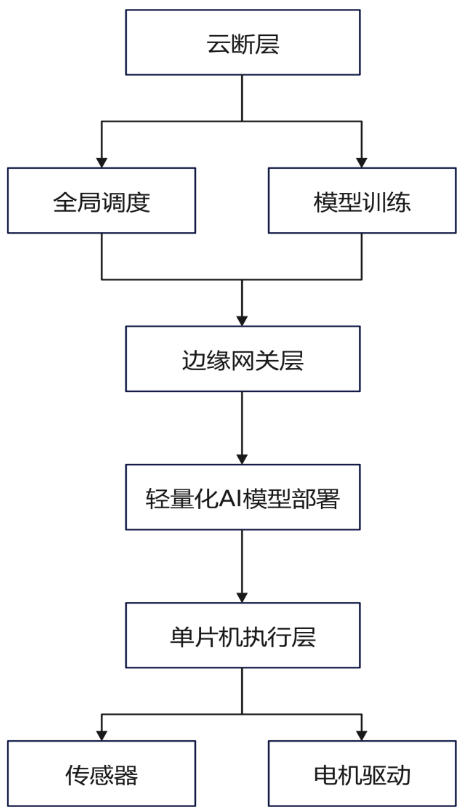

Figure 1. Edge intelligence control architecture.

2.3. Collaborative optimization training mechanism

In the empowerment of generative AI, a cloud-edge collaborative training framework is established to overcome the resource bottlenecks of standalone systems, constructing a complete iterative optimization system that encompasses cloud-based training, edge-based fine-tuning, and terminal deployment. Federated learning technology enables multiple edge nodes to collaboratively train a global model based on local data, avoiding privacy risks associated with uploading raw data. In the context of smart agriculture, 100 distributed soil sensor nodes train a humidity prediction model through a federated learning framework, reducing communication overhead and accelerating model convergence. The introduction of transfer learning technology accelerates model adaptation; for instance, fine-tuning the last two fully connected layers of a pre-trained ResNet-18 model achieves 97% accuracy in plant disease recognition on an ESP32-S3 chip, with training time compressed from 72 hours to just 15 minutes.

Furthermore, the continuous learning mechanism endows microcontrollers with adaptive capabilities. Online learning algorithms dynamically update models through a sliding window mechanism; in equipment condition monitoring scenarios, an LSTM-based temporal model undergoes incremental training every 1,000 new samples collected, continuously improving fault prediction accuracy over time. Hardware-friendly optimization techniques reduce the overhead of continuous learning; quantization-aware training introduces quantization noise during the model training phase, limiting the accuracy loss of INT8 models to within 1%. The NVIDIA Jetson Nano development board integrates a dual-core ARM Cortex-A57 processor with a 128-core Maxwell GPU, enabling continuous optimization of the YOLOv5 model at the edge. The size of the model update package has been compressed from several hundred megabytes to a few dozen kilobytes, facilitating efficient mechatronic control via single-board computers.

3. Application measures for efficient control enabled by generative AI

3.1. Intelligent fault prediction and maintenance

By applying generative AI technology, a deep GAN can be established to dynamically model equipment degradation processes. Taking CNC machine tools at an automotive manufacturer as an example, raw vibration sensor data undergoes joint time-frequency domain analysis before being input into an improved WGAN-GP model for feature enhancement, generating a training dataset comprising 2,000 virtual fault samples. This model achieves over 92% accuracy in diagnosing outer ring bearing faults, significantly outperforming the traditional SVM-based approach (73.6%) and enabling precise control. In practical operations, a three-stage data governance strategy is employed: environmental interference is eliminated through wavelet threshold denoising, 128-dimensional temporal features are extracted using LSTM-Autoencoder, and t-SNE dimensionality reduction maps data to a 3D space for visual anomaly detection. Case studies on SMT placement machines demonstrate that this solution extends the mean time between failures (MTBF) from 450 to 680 hours while substantially reducing maintenance costs. It achieves intelligent fault prediction and effective maintenance in single-board computer systems, enhancing operational efficiency.

3.2. Adaptive parameter optimization control

Experimental data from a precision machining center using a generative AI-based dynamic parameter optimization system is presented in **Table 1**.

Table 1. Comparison of adaptive parameter optimization

Parameter index	Traditional PID control	Generative AI optimized
Spindle speed fluctuation rate	2.1%	0.7%
Feed rate overshoot	18.5%	5.2%
Surface roughness (Ra)	0.82 μm	0.47 μm
Unit energy consumption	0.32 kW·h/cm³	0.21 kW·h/cm³
Tool wear rate	0.012 mm/h	0.0065 mm/h

Based on the data, this system employs a reinforcement learning framework that integrates the Q-learning algorithm with generative data augmentation techniques. During the training phase, 5,000 sets of virtual operating condition data were generated using a VAE, covering a cutting force range of 0–1500 N and a rotational speed range of 5000–12000 rpm. The decision-making network adopts a dual-stream architecture, where the spatiotemporal feature stream processes vibration signals using a 3D-CNN, and the temporal feature stream processes temperature data using a Transformer encoder. Experiments demonstrate that during 200 consecutive hours of machining, the system dynamically adjusted parameters 1,274 times, significantly reducing ineffective adjustments compared to fixed parameter modes.

3.3. Multi-device collaborative production scheduling

To address the challenges of scheduling in flexible manufacturing systems, a cloud-edge collaborative architecture based on generative AI is proposed. An improved Transformer model is deployed in the cloud, featuring a multi-scale time window design for its attention mechanism, enabling it to capture minute-level equipment state changes and analyze hourly-level order fluctuation patterns [9]. On the edge side, a lightweight LSTM prediction module is implemented, with the cloud model compressed to $1/15^{th}$ of its original size through knowledge distillation techniques, ensuring a response speed of 50 milliseconds.

In terms of energy optimization, the system trains scheduling strategies using Generative Adversarial Imitation Learning (GAIL), reducing unit product energy consumption while meeting delivery deadlines. Test data from a semiconductor packaging plant shows that when handling six different types of process products simultaneously, the system maintains a 98.7% plan completion rate. Thus, with the empowerment of generative AI, single-chip microcomputer control systems are reshaping the control forms of mechatronic systems, effectively driving the manufacturing industry towards intelligence and flexibility [10].

4. Conclusion

This study addresses the limitations of traditional single-chip microcomputer control performance by introducing innovative generative AI technology. It analyzes the technical characteristics of generative AI and explores the innovations in single-chip microcomputer-controlled mechatronic systems enabled by generative AI. Additionally, it examines the technical pathways and methodologies for achieving efficient control. By leveraging features such as data-driven approaches, adaptive learning, and multimodal integration, this research breaks through the limitations of traditional technologies in single-chip microcomputer control. It proposes technical methods, including lightweight model deployment, edge intelligent control architecture, and collaborative optimization training mechanisms to achieve efficient operation. At the application level, intelligent fault prediction enhances equipment reliability, adaptive parameter optimization reduces processing errors and energy consumption, and collaborative scheduling of multiple devices enhances production flexibility. The research demonstrates that generative AI empowerment significantly improves system control precision and energy efficiency, driving the manufacturing industry toward deep transformation towards intelligence and flexibility, as well as high-quality development.

Funding

Single-Chip Microcomputer and Interface Technology Project (Project No.: SYSJ2025032)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Chen Z, Gao B, 2025, Automated Control System for Brushless DC Motors Based on STM32 Microcontroller. Electronic Design Engineering, 33(3): 58–62.

[2] Dang J, Zhao R, Dai J, et al., 2024, Control System for Liquid Zoom Lenses Based on ATmega32U4 Microcontroller. Electronic Measurement Technology, 47(22): 25–30.

[3] Ma W, 2024, Application of Mechatronics Technology in Agricultural Machinery Design. Chinese Journal of Agricultural Resources and Regional Planning, 45(8): 253–266.

[4] Wu H, 2023, Intelligent Mechatronics Empowers Modernization of Agriculture and Rural Areas and Rural Revitalization. China Fruits, 2023(7): 155.

[5] Lin F, 2023, Design of Hydraulic Transmission Control System based on AT89S52 Microcontroller. Hydraulics Pneumatics & Seals, 43(3): 46–49.

[6] Fan C, Tong Y, Ma Z, et al., 2024, TSN Function Design for Gateway Board Combining Microcontroller and TSN Chip. Integrated Circuits and Embedded Systems, 24(3): 67–71.

[7] Li L, 2025, Design of Greenhouse Temperature Control System Based on Microcontroller. Southern Agricultural Machinery, 56(17): 73–75.

[8] Zhuang Q, 2025, Efficient Control of Mechatronic Systems Using Microcontrollers. Time Automobile, 2025(18): 130–132.

[9] Li Y, Zhan H, 2025, Control Method for Electronically Controlled Mechanical Automatic Transmission Using STM32 Microcontroller. Machinery Design & Manufacture, 1–7.

[10] Zhang J, 2024, Automated Control System for Mechatronic Devices Based on Laser Sensors. Automation & Instrumentation, 39(3): 103–106.