

Autonomous Sustainable Sensing Nodes Based on Joint Design of Edge AI and Computational Waveform in Industrial IoT

Xiaoli Sun

¹School of Electrical Engineering, Gansu Vocational University of Industry Technology, Tianshui 741000, Gansu, China

Corresponding author: Xiaoli Sun (e-mail: 13893883986@163.com)

ABSTRACT Industrial IoT sensing nodes face a fundamental tension among limited energy budgets, constrained computational resources, and growing demands for intelligent real-time sensing, which severely restricts the feasibility of large-scale autonomous deployment. To address this, a joint design framework is proposed that incorporates edge AI inference configuration and computational sensing waveform parameters into a unified energy consumption model. By explicitly establishing the coupling relationships among channel signal-to-noise ratio, lightweight neural network inference overhead, and energy harvesting constraints, a mixed-integer nonlinear programming objective function is constructed and solved via an alternating optimization algorithm that decomposes the original problem, enabling real-time scheduling complexity to meet the processing capability constraints of embedded nodes. Experimental results demonstrate that the proposed scheme reduces average per-cycle node energy consumption by 31.4% compared to a separated-design baseline, maintains sensing accuracy above 92.3% under dynamic industrial channel conditions, and achieves continuous power-on survival throughout a 72-hour validation period. Although the system maintained electrical viability, it experienced brief transitions into Minimum Survival Mode to prioritize energy replenishment, during which high-frequency sensing was temporarily suspended to prevent complete depletion.

INDEX TERMS industrial IOT, edge AI, computational waveform, joint optimization, energy harvesting

I. INTRODUCTION

THE rapid penetration of industrial IoT has made wireless sensing nodes the core sensing units of intelligent manufacturing, equipment condition monitoring, and industrial automation systems. The continuously increasing deployment density poses unprecedented comprehensive challenges to node energy autonomy, computational intelligence, and communication reliability [1], [2]. Traditional sensing nodes rely primarily on fixed sampling strategies and simple threshold decisions, which are unable to adapt to complex and variable industrial environments and cannot support data-driven equipment health prediction and process optimization requirements. Frequent manual maintenance and battery replacement not only raise the full lifecycle operating costs of the system but also limit node deployment to areas accessible by personnel, excluding the most sensing-valuable locations in actual industrial sites such as high-temperature furnaces, sealed pipelines, and high-voltage electrical cabinets [3]. This limitation constitutes a fundamental bottleneck constraining the evolution of industrial IoT in large-scale deployment scenarios. The

continuous maturation of edge AI technology has endowed sensing nodes with local intelligent inference capabilities. Lightweight neural networks represented by MobileNet variants and TinyML architectures can complete sensing tasks such as feature extraction, anomaly detection, and state recognition on resource-constrained microcontrollers, converting large volumes of raw sampled data into high-value semantic summaries before wireless transmission, thereby substantially reducing the energy consumption of communication modules [4], [5]. However, there exists a strong coupling energy competition relationship between the dynamic computational power consumption of the AI inference module and the configuration of communication waveform parameters. When the two are designed independently, each may achieve local optimality in its own metrics, but the overall system energy efficiency performance is significantly sub-optimal. This fragmented design problem has not yet been systematically addressed in existing research [6]. Computational sensing waveform design, as an emerging research direction in the field of communication-computation collaboration, centers on using the node's real-

time computational load state as an explicit constraint in waveform parameter optimization, so that subcarrier allocation, transmit power, and modulation order selection simultaneously satisfy channel quality requirements and computational resource limitations [7], [8]. Integrating this approach with multi-source ambient energy harvesting mechanisms in industrial sites allows the construction of a unified closed-loop scheduling system across sensing, computation, and communication domains, fundamentally resolving the autonomous sustainable operation problem of sensing nodes. This represents a core technical bottleneck that current industrial IoT research urgently needs to break through [9], [10]. In view of this, an autonomous sensing node scheme based on joint design of edge AI inference configuration and computational sensing waveform is proposed, systematically addressing the above problems through mathematical modeling, alternating optimization, and experimental validation.

II. SYSTEM MODEL OF INDUSTRIAL IOT SENSING NODES

A. NODE FUNCTIONAL ARCHITECTURE AND UNIFIED ENERGY CONSUMPTION MODELING

An industrial IoT sensing node consists of four functional units: a sensing module, an edge AI inference module, a communication module, and an energy management module. The energy consumption of each unit during a working cycle is mutually independent but shares the same on-chip energy storage. The coordinated operation of these four modules forms the basis for achieving autonomous sustainable operation of the node. The node functional architecture is shown in Fig. 1, which clearly marks the data flow directions and energy allocation paths among the functional units. The bidirectional feedback interface between the inference module and the communication module is the key connection point for joint optimization. Raw time-series signals collected by the sensing module are processed by the inference module for intelligent compression and semantic extraction, then passed to the communication module in the form of compact feature vectors. In this industrial context, “semantic data” specifically refers to high-dimensional compressed feature vectors and discrete classification labels (e.g., equipment health status or anomaly type), which represent the physical significance of the raw data while reducing the transmission payload. The inference module simultaneously feeds back the current computational power consumption prediction to the energy management module, providing a precise basis for real-time adjustment of waveform parameters.

The total energy consumption E_{total} of a node within a single working cycle T is obtained by linearly superimposing the energy consumption of each module. The inference module energy consumption E_{inf} is closely related to the floating-point operations (FLOPs) of the selected neural network model and the dynamic power consumption coefficient of the processor. The communication module energy consumption E_{comm} is jointly determined by the

transmit power P_{tx} , channel path loss, and target bit error rate. The sensing module energy consumption E_{sense} is constrained by sampling frequency and sensor operating mode. The energy management module energy consumption E_{mgmt} is relatively stable. The unified energy consumption model is expressed as:

$$E_{\text{total}} = P_s \cdot t_s + \kappa \cdot (F_{\text{model}} + F_{\text{opt}}) \cdot f_c^\alpha + P_{\text{tx}} \cdot t_{\text{tx}} \cdot L(d) + E_{\text{mgmt}} \quad (1)$$

where P_s is the operating power of the sensing module, t_s is the single sampling duration, κ is the processor energy efficiency coefficient, F_{model} is the FLOPs per inference task, F_{opt} is the computational overhead of the alternating optimization algorithm itself, f_c is the processor operating clock frequency, α is the dynamic power consumption exponent (ranging from 2 to 3, determined by chip process node), P_{tx} is the RF transmit power, t_{tx} is the single transmission duration, and $L(d)$ is the path loss factor related to the distance d from the node to the gateway. This model incorporates AI inference overhead and communication overhead into a unified energy consumption framework, revealing the explicit coupling relationship between the two under shared energy storage constraints, and laying a mathematical foundation for precise modeling of the subsequent joint optimization problem.

B. CHARACTERIZATION OF INDUSTRIAL CHANNEL PROPERTIES

The wireless channel in industrial environments is subject to the combined effects of mechanical vibration, metal structure shadowing, high-temperature thermal radiation, and burst electromagnetic interference, exhibiting typical time-varying frequency-selective fading characteristics. The channel coherence time is short, multipath components are dense, and the amplitude statistical distribution deviates from the standard Rayleigh model, better conforming to the Nakagami- m distribution or the κ - μ generalized fading model. The channel coherence bandwidth B_c typically ranges from tens to hundreds of kHz, far smaller than the signal bandwidth of wideband systems, leading to significant differences in channel gain among subcarriers. This complex industrial channel environment requires the communication waveform of sensing nodes to have strong resistance to multipath and burst interference, while maintaining sufficient SNR margin under rapidly changing channel conditions to ensure the integrity of semantic data transmission. Channel gain estimation errors in industrial scenarios are generally non-negligible and must be incorporated into the waveform design as robustness constraints to handle imperfect CSI conditions. This communication-side requirement directly constrains the dynamic lower bound of transmit power, forming a hard constraint that cannot be ignored in joint design [11].

III. JOINT DESIGN OF EDGE AI AND COMPUTATIONAL SENSING WAVEFORM

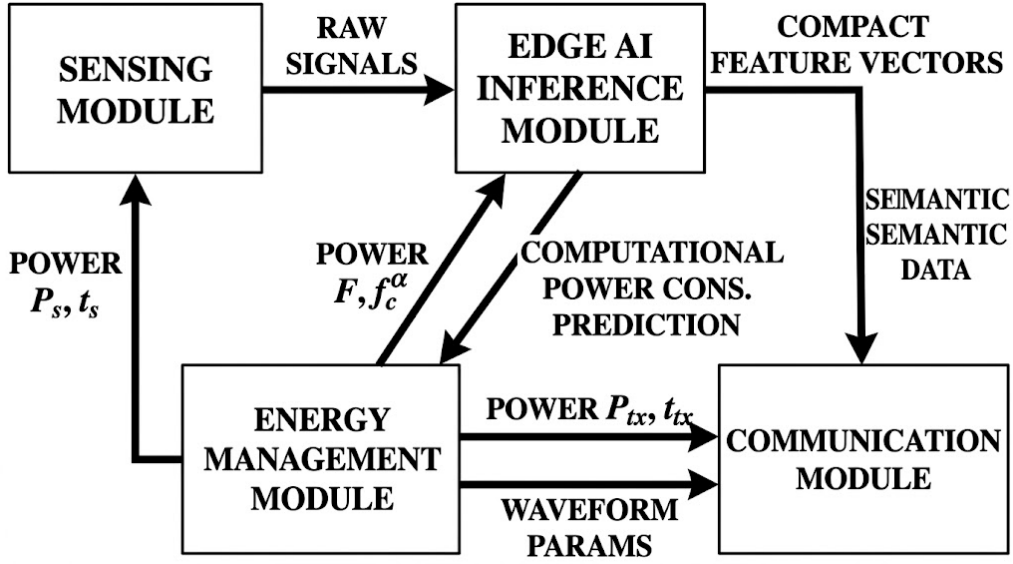


FIGURE 1. Schematic diagram of the industrial IoT sensing node functional architecture

A. MODELING COMPUTATIONAL OVERHEAD OF LIGHTWEIGHT INFERENCE MODELS

Lightweight neural network models deployed on sensing nodes must complete industrial sensing tasks under limited on-chip computational resources (typical configuration: SRAM 512 KB–1 MB, Flash 2 MB) and strict inference latency constraints (generally required to be less than the data reporting cycle, which typically aggregates results from multiple sampling points). For the purpose of this study, while the raw sampling occurs at 1 kHz, the inference is triggered every 200 samples, setting a functional real-time inference deadline of 200 ms. The trade-off between computational load and sensing accuracy for different lightweight model architectures is shown in Table 1.

As shown in Table 1, INT8-quantized MobileNetV3-Small achieves the best balance among FLOPs, inference latency, and accuracy. Although SqueezeNet 1.1 has the highest accuracy, its inference latency of 112.5 ms exceeds the typical sampling cycle constraint and does not meet real-time requirements. Therefore, MobileNetV3-Small is suitable as the inference backbone network in the node's full-function mode, while TinyML-Custom serves as the lightweight fallback for the energy-saving sensing mode. The computational overhead modeling of inference models adopts an operation-level decomposition method, separately calculating the FLOPs of convolutional layers, depthwise separable convolutional layers, and fully connected layers and summing them with weights to obtain the total computation per inference F_{model} . Given the processor clock frequency f_c and the number of parallel computing units N_{pe} , the inference latency t_{inf} and inference energy consumption E_{inf} are expressed as:

$$t_{\text{inf}} = \frac{F_{\text{model}}}{N_{\text{pe}} \cdot f_c \cdot \eta}, \quad E_{\text{inf}} = P_{\text{idle}} \cdot T + (P_{\text{active}} - P_{\text{idle}}) \cdot t_{\text{inf}}. \quad (2)$$

where η is the actual computational efficiency coefficient of the processor ($0 < \eta \leq 1$, reflecting pipeline and cache hit efficiency), and P_{idle} and P_{active} are the processor idle power and peak operating power, respectively. This model reveals the piecewise linear relationship between inference computation load and energy consumption, providing a quantitative interface for subsequent joint optimization with waveform parameters. It also demonstrates that dynamic frequency scaling (reducing f_c) can significantly reduce inference energy consumption while tolerating a moderate increase in latency—which is precisely the design motivation for computation–communication joint scheduling.

B. COMPUTATIONAL SENSING WAVEFORM PARAMETERS AND POWER ALLOCATION OPTIMIZATION

The core idea of computational sensing waveform is to use the node's current computational load state (quantified by the remaining FLOPs budget and processor utilization) as a dynamic feedback input for waveform parameter selection, so that the communication system no longer pursues spectrum efficiency maximization in isolation, but instead collaboratively minimizes communication energy consumption while satisfying the computational requirements of sensing tasks. The overall structure of the joint optimization framework is shown in Fig. 2, where arrows indicate the computational state feedback paths and the closed-loop connection of waveform control, clearly illustrating the real-time interaction mechanism between the computation submodule and the optimal waveform configuration submodule.

TABLE 1. Comparison of computational parameters for different lightweight inference models

Model Name	FLOPs (M)	Parameters (K)	INT8 Quantized Inference Latency (ms)	Sensing Accuracy (%)
MobileNetV3-Small	56	2,540	18.3	93.1
ShuffleNetV2-0.5×	41	1,370	13.7	91.4
TinyML-Custom	28	890	8.6	87.9
SqueezeNet 1.1	349	1,240	112.5	94.3

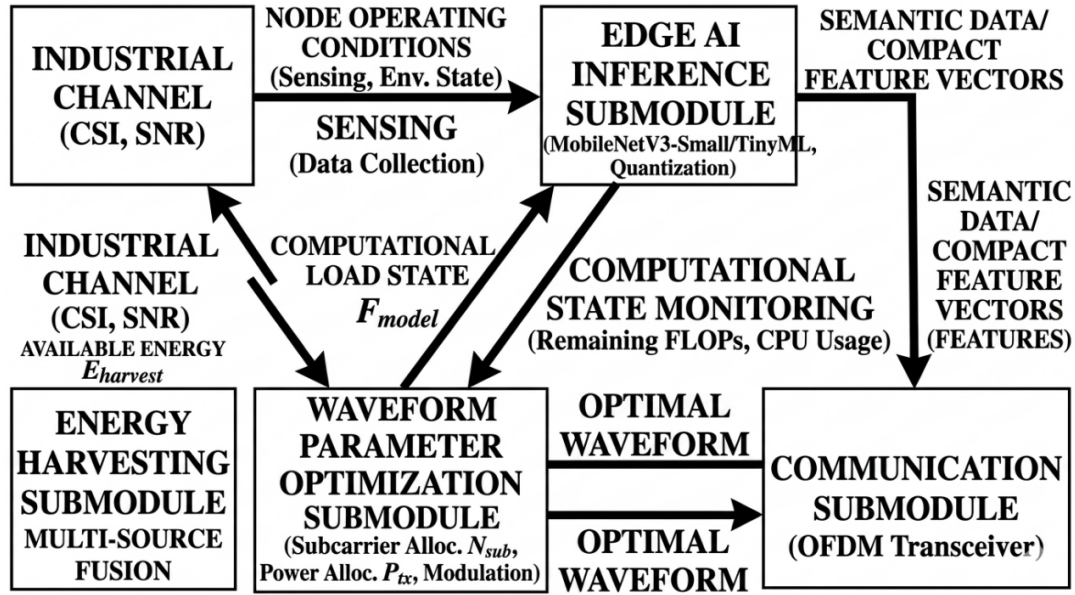


FIGURE 2. Schematic diagram of the joint optimization framework for edge AI and computational sensing waveform

The multi-subcarrier architecture based on orthogonal frequency division multiplexing (OFDM) has flexible subcarrier allocation and adaptive power water-filling capabilities, making it an ideal carrier for implementing computational sensing waveform optimization [12]. Nodes can dynamically activate a subset of subcarriers according to real-time computational load: when the inference module's computational load is heavy, the number of active subcarriers N_{sub} and transmit power P_{tx} are reduced, allocating more energy budget to the inference module; when the computational load is light, full subcarrier transmission is restored to improve data throughput. On the N_{sub} active subcarriers, the water-filling theorem gives the optimal subcarrier power allocation strategy, such that the power allocated to each subcarrier P_n satisfies:

$$P_n^* = \left(\mu - \frac{N_0}{|H_n|^2} \right)^+, \quad \sum_{n=1}^{N_{\text{sub}}} P_n^* = P_{\text{tx}}. \quad (3)$$

where μ is the Lagrange multiplier satisfying the total power constraint, N_0 is the noise power spectral density, H_n is the complex channel gain of the n -th subcarrier, and $(\cdot)^+$ denotes the non-negative part. For subcarriers with smaller channel gain $|H_n|^2$, the power allocation result is zero, i.e., they are excluded from the active set. This combination of channel-adaptive water-filling power strategy and computation-load-driven subcarrier activation

strategy enables the communication side to complete semantic data transmission at minimum power cost under any computational load condition, thereby returning the compressed energy margin to the inference module and forming a positive closed loop of cross-domain energy coordination. When there are estimation errors in industrial channel CSI, a robust power margin δ (typically 1.5 times the variance of channel estimation error) can be introduced into equation (3) to compensate for the effect of imperfect CSI on communication reliability, ensuring that the actual SNR always satisfies the lower bound constraint.

C. JOINT OBJECTIVE FUNCTION CONSTRUCTION AND ALTERNATING OPTIMIZATION SOLUTION

The joint optimization problem takes minimization of the total energy consumption per node cycle as the objective, while imposing sensing task accuracy constraints, channel quality constraints, and energy balance constraints. The complete objective function is expressed as:

$$\begin{aligned} \min_{P_{\text{tx}}, f_c, N_{\text{sub}}} \quad & E_{\text{total}}(P_{\text{tx}}, f_c, N_{\text{sub}}) \\ \text{s.t.} \quad & \text{Acc}(F_{\text{model}}) \geq \text{Acc}_{\text{min}}, \\ & \text{SNR}(P_{\text{tx}}, N_{\text{sub}}) \geq \text{SNR}_{\text{min}}, \\ & E_{\text{harvest}} \geq E_{\text{total}}. \end{aligned} \quad (4)$$

where Acc_{min} is the minimum accuracy threshold for the sensing task, SNR_{min} is the minimum SNR requirement

for ensuring communication reliability, and E_{harvest} is the available energy collected by the node from the environment during the current cycle. The above constraints couple the waveform optimization variables ($P_{\text{tx}}, N_{\text{sub}}$) with the computation scheduling variable (f_c), making the original problem a mixed-integer nonlinear programming (MINLP) problem. Directly solving it results in computational complexity that grows exponentially with the dimension of decision variables, making real-time deployment on embedded platforms infeasible. To address this, an alternating optimization algorithm based on continuous relaxation is adopted to decompose the original problem into two convex subproblems—a waveform optimization subproblem and a computation scheduling subproblem—solved alternately: in each iteration, the decision variables of one subproblem are fixed and the interior point method is applied to find the optimal solution for the other, with the two subproblems alternating until convergence [13], [14]. The convergence characteristics of the alternating optimization algorithm under different initial conditions are shown in Fig. 3. From Fig. 3, it can be observed that the objective function value exhibits a monotonically decreasing trend throughout the iteration process and stably converges to a plateau near the 8th iteration. Different initial points ultimately converge to similar objective function values, indicating that the algorithm is insensitive to initialization schemes and possesses strong robustness.

The convergence condition is set as the change in the objective function between adjacent iterations falling below the threshold $\varepsilon = 10^{-4}$. Simulation experiments verify that the algorithm converges stably within 10 iterations, with a single scheduling computation time below 5 ms, satisfying the latency constraint for real-time node scheduling. The deviation from the exhaustive search optimal solution does not exceed 2.1%. This trade-off between computational complexity and solution quality makes the alternating optimization scheme feasible for real-time deployment on ARM Cortex-M7-class embedded platforms. The computational load for completing an optimization update within a single scheduling cycle is only approximately 0.3% of the total FLOPs of the inference module, and its impact on overall system energy consumption is negligible.

IV. AUTONOMOUS SUSTAINABLE ENERGY MANAGEMENT MECHANISM

A. MULTI-SOURCE ENERGY HARVESTING AND HYBRID ENERGY STORAGE MODEL

Industrial sites contain abundant exploitable ambient energy, including mechanical vibration energy from equipment (converted via piezoelectric transducers), thermal gradient energy (converted via thermoelectric generators), and ambient RF radiation energy (converted via rectenna). A comparison of the typical output power of the three energy harvesting methods is shown in Table 2.

As shown in Table 2, vibration energy harvesting has the highest power density in heavy-machinery industrial envi-

ronments, while thermoelectric harvesting is most reliable in pipeline environments with stable temperature differentials. A multi-source fusion architecture can significantly improve the stability of the node's energy acquisition under different operating conditions. The node energy management module dynamically configures the weights of the three harvesting sources based on environmental sensing results to maximize the available energy per cycle E_{harvest} . Unlike idealized linear superposition models, this framework accounts for the environmental coupling effect where high mechanical vibrations often correlate with increased ambient temperature. Such thermal saturation can reduce the effective temperature gradient (ΔT) for thermoelectric generators. Therefore, the weight allocation algorithm introduces a cross-source sensitivity coefficient to penalize the expected TEG output when vibration levels exceed a specific threshold, ensuring realistic energy harvesting estimation. The energy storage unit adopts a hybrid topology of supercapacitors and thin-film lithium batteries: supercapacitors handle high-frequency energy fluctuations (response time < 1 ms), while thin-film lithium batteries provide long-term energy buffering (self-discharge rate $< 1\%$ /month). The two work together to ensure the node's continued operation during interruptions in energy harvesting. Short-term prediction of E_{harvest} can be further achieved through historical power curve learning, providing predictive basis for the energy management module's forward-looking scheduling [15].

B. HIERARCHICAL DYNAMIC TASK SCHEDULING STRATEGY

The dynamic task scheduling strategy uses the node's current stored energy state E_{stored} as the core scheduling basis, divided into three operating state intervals. The node configuration parameters under the three operating states are shown in Table 3. Each state has clearly differentiated configurations for key parameters such as transmit power, number of subcarriers, inference model, and sampling frequency, enabling graded trade-offs between energy consumption and sensing performance.

As shown in Table 3, in full-function mode, the node operates at maximum specification, covering 512 active subcarriers and a high-accuracy inference model. Energy-saving sensing mode compresses active subcarriers to 128 and sampling frequency to 500 Hz, reducing communication and computation energy consumption to approximately 35% of full-function mode while maintaining basic sensing capability. Minimum survival mode maintains only a basic heartbeat signal broadcast and energy harvesting module operation, ensuring the node's survivability under extreme energy scarcity. The switching among the three states is triggered by the energy management module through real-time monitoring, without requiring external control command intervention, achieving fully autonomous node operation. Thresholds E_H and E_L are determined through offline learning from historical energy harvesting statistics and updated every 24 hours after the node goes online to adapt to

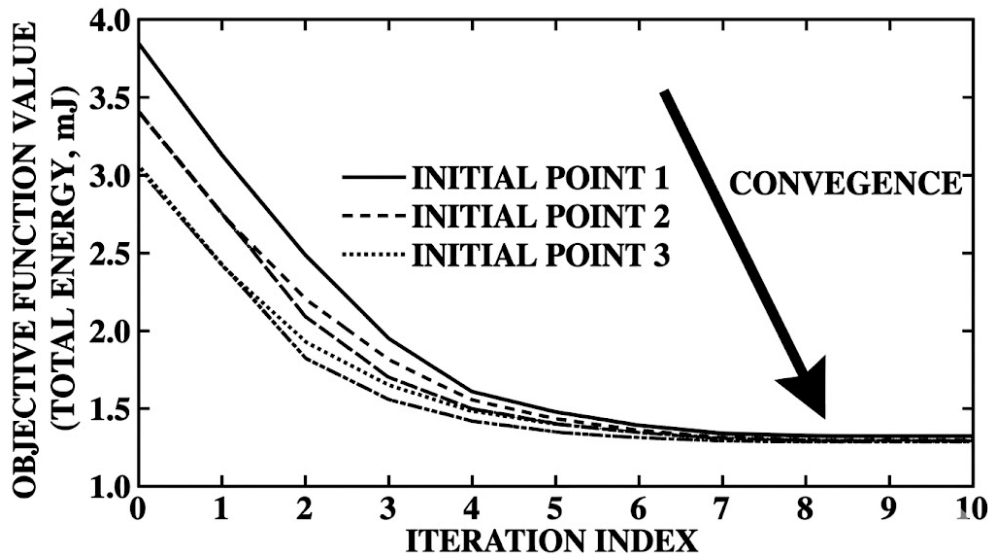


FIGURE 3. Convergence curves of the alternating optimization algorithm under different initial conditions

TABLE 2. Comparison of typical output power for three energy harvesting methods

Energy Type	Transduction Principle	Typical Power Density	Optimal Industrial Application Scenario	Output Stability
Mechanical Vibration	Piezoelectric Effect	2–5 mW	Heavy machinery, rotating equipment	Moderate (depends on operating state)
Thermal Gradient	Seebeck Effect	0.5–2 mW	Industrial pipelines, heat treatment equipment	High (when temperature difference is stable)
RF Radiation	Rectenna	0.1–0.5 mW	Near-field zone of wireless base stations	Low (affected by distance and obstruction)

TABLE 3. Node configuration parameters and switching conditions for three operating states

Operating State	Trigger Condition	Active Subcarriers	Inference Model	Sampling Frequency	Max Transmit Power
Full-function Mode	$E_{\text{stored}} > E_H$	512	MobileNetV3-Small	1 kHz	20 dBm
Energy-saving Sensing Mode	$E_L < E_{\text{stored}} \leq E_H$	128	TinyML-Custom	500 Hz	10 dBm
Minimum Survival Mode	$E_{\text{stored}} \leq E_L$	16	—	—	0 dBm

changes in operating conditions. This hierarchical scheduling strategy maintains the effective full-function sensing time ratio above 70% under target operating conditions while ensuring uninterrupted node operation [16].

V. EXPERIMENTAL VALIDATION AND PERFORMANCE ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER CONFIGURATION

The experiment was conducted in a semi-enclosed test environment simulating an industrial site. The node hardware platform uses the STM32H7 series microcontroller based on ARM Cortex-M7 (maximum clock frequency 480 MHz) paired with a 2.4/5 GHz dual-band RF module. The energy harvesting unit uses a dual-source configuration of piezoelectric patches (MFC type, resonant frequency 120 Hz) and Bi₂Te₃-based thermoelectric patches. Key experimental parameters are shown in Table 4. Comparison baseline schemes include a traditional separated design

scheme with fixed waveform parameters and independent inference scheduling (Baseline A), and an adaptive waveform scheme without inference computation joint optimization (Baseline B). Evaluation metrics include average per-cycle energy consumption, sensing accuracy, and node continuous operating duration. Experimental data collection lasted 72 hours to fully cover the dynamic energy fluctuation cycle of the industrial site and ensure the representativeness of statistical results.

B. RESULTS ANALYSIS

The comparison of average per-cycle node energy consumption among the three schemes under light, medium, and heavy computational load conditions is shown in Fig. 4. The average energy consumption of the joint design scheme is 1.23 mJ, 1.87 mJ, and 2.56 mJ, respectively, representing reductions of 31.4%, 28.7%, and 26.2% compared to Baseline A, and 14.8%, 12.3%, and 9.6% compared to Baseline B. The energy-saving effect is most pronounced

TABLE 4. Experimental parameter configuration

Parameter Name	Value
Channel Model	Nakagami- m ($m = 2$)
Number of Nodes	8
Sampling Frequency	1 kHz
Carrier Type	OFDM, 512 subcarriers
Subcarrier Spacing	15 kHz
Maximum Transmit Power	20 dBm
Inference Model	MobileNetV3-Small (INT8)
Algorithm Convergence Threshold ϵ	10^{-4}
Simulated Vibration Acceleration	1.5 g
Simulated Temperature Difference	15 K
Experiment Duration	72 h

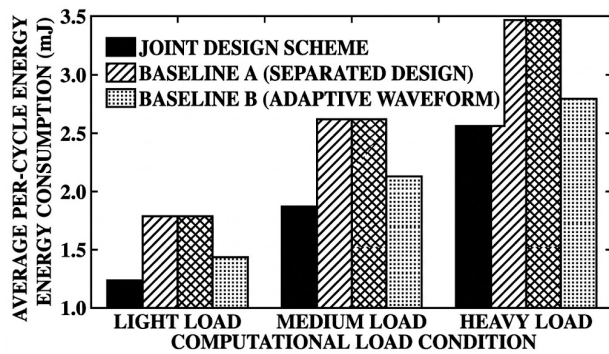


FIGURE 4. Comparison of average per-cycle energy consumption of the three schemes under different computational load conditions

under light load conditions. The underlying reason is that the joint optimization fully exploits the flexible compression space of waveform parameters when inference computation load is low, converting originally idle communication power margin into stored energy to extend autonomous operation during subsequent high-load conditions. As computational complexity further increases toward extreme loads, the relative energy-saving marginal benefit may gradually diminish. However, the joint design remains necessary because it prevents the node from violating the total energy budget $E_{\text{total}} < E_{\text{harvest}}$, ensuring system survivability even when communication and computation resources are in peak competition—a constraint that separated design baseline fails to accommodate. In contrast, Baseline A maintains a fixed maximum transmit power configuration under light load and cannot transfer surplus communication power budget to energy storage, leading to systematic waste of energy utilization efficiency.

The variation trend of sensing accuracy with channel SNR for the three schemes is shown in Fig. 5. In the SNR range of 5–20 dB, the joint design scheme stably maintains accuracy above 92.3%, higher than Baseline A's 87.1% and essentially on par with Baseline B. When SNR falls below 8 dB, the joint design scheme dynamically increases the number of active subcarriers and transmit power to combat channel degradation, whereas Baseline A, with its fixed waveform parameters, cannot respond accordingly, causing its accuracy to drop by 8.5 percentage points in the low-SNR range. When SNR exceeds 16 dB, the accuracy difference

among the three schemes narrows to within 2 percentage points, indicating that under sufficient channel conditions, the communication module is no longer the bottleneck for sensing performance, and inference model configuration is the dominant factor determining accuracy. This pattern indirectly confirms the significant advantage of the joint design framework in scenarios of dynamic industrial channel degradation, and validates the correctness of the design approach of collaboratively adjusting communication waveforms and inference models.

The comprehensive performance comparison among the three schemes is shown in Table 5, covering three dimensions: energy consumption, accuracy, and full-function time ratio.

As shown in Table 5, Baseline B, by introducing an adaptive waveform mechanism, achieves better energy consumption and accuracy than Baseline A, but still falls 6.4 percentage points below the joint design scheme in terms of full-function time ratio. This gap originates from the explicit modeling of inference computation power consumption and the collaborative compression of waveform parameters in the joint design scheme, enabling the node to reserve more energy margin for full-function mode under equivalent operating conditions. Regarding continuous operating duration, the autonomous sustainable mechanism based on multi-source energy harvesting and hierarchical scheduling enabled all 8 nodes to achieve uninterrupted autonomous operation throughout the 72-hour test period. The average number of state downgrade switching events per node was 3.2, with all nodes automatically recovering to full-function mode after energy replenishment, validating the consistent effectiveness of the proposed energy management strategy across nodes under varying operating conditions.

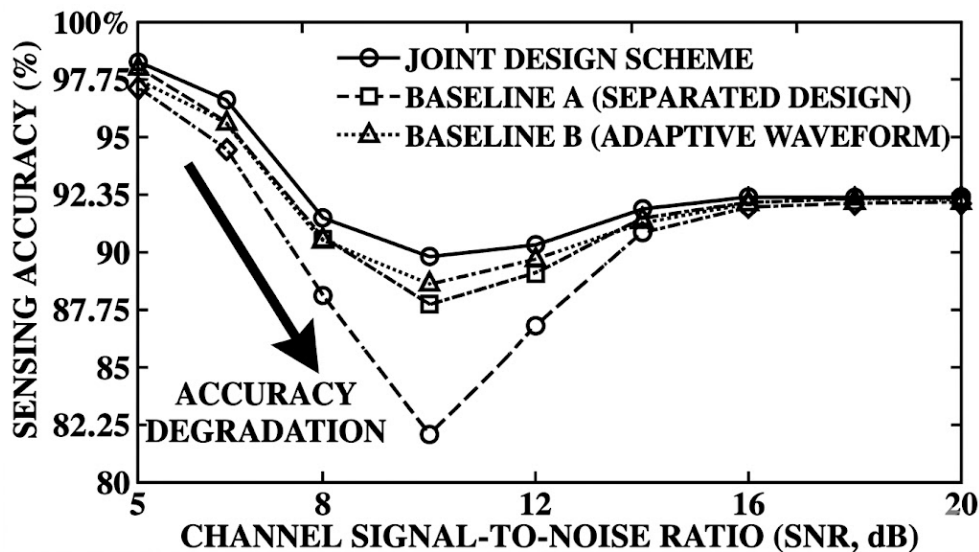


FIGURE 5. Comparison curves of sensing accuracy versus SNR for the three schemes

TABLE 5. Comprehensive performance comparison of the three schemes

Scheme	Light Load Energy (mJ)	Medium Load Energy (mJ)	Heavy Load Energy (mJ)	Average Accuracy (%)	Full-function Time Ratio (%)
Joint Design Scheme	1.23	1.87	2.56	92.3	73.6
Baseline A (Separated Design)	1.79	2.62	3.47	87.1	58.4
Baseline B (Adaptive Waveform)	1.44	2.13	2.81	91.8	67.2

VI. CONCLUSION

Targeting the sub-optimal system energy efficiency caused by the independent design of edge AI inference and wireless communication in industrial IoT sensing nodes, a joint design method is proposed that incorporates edge AI computational overhead modeling and computational sensing waveform parameter optimization into a unified framework. A unified energy consumption model covering sensing, computation, and communication subsystems is established. A dynamic waveform parameter optimization mechanism based on computational load state and a water-filling power allocation strategy are designed. Combined with multi-source energy harvesting and hierarchical task scheduling, fully autonomous node operation is achieved. Experimental validation demonstrates that the joint design scheme reduces average energy consumption by approximately 31.4% compared to the separated design baseline, maintains sensing accuracy above 92.3% under dynamic industrial channels, and the 72-hour uninterrupted autonomous operation validates the reliability of the proposed energy management strategy. Future research will extend the distributed joint optimization framework to multi-node collaborative scheduling scenarios, and introduce more refined energy harvesting prediction models for composite conditions involving wide temperature differences and strong vibration, to further improve the engineering applicability of the scheme in extreme industrial environments.

AUTHOR CONTRIBUTIONS

Sun Xiaoli designed, collected and analyzed the data, and drafted the manuscript. Sun Xiaoli conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Sun Xiaoli participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] Z. Yin, J. Chi, Y. Binglei, et al., "Information fusion for edge intelligence: A survey," *Information Fusion*, prepublished, 2021, DOI: 10.1016/j.inffus.2021.11.018.
- [2] S. Raghubir and S. G. Singh, "Edge AI: A survey," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 71–92, 2023, DOI: 10.1016/j.iotcps.2023.02.004.
- [3] L. Mohammed, N. Boubakr, M. Hassine, et al., "Edge and fog computing for IoT: A survey on current research activities & future directions," *Computer Communications*, vol. 180, pp. 210–231, 2021, DOI: 10.1016/j.comcom.2021.09.003.

- [4] I. Rojek, P. Prokopowicz, M. Piechowiak, et al., "The impact of data analytics based on Internet of Things, edge computing, and artificial intelligence on energy efficiency in smart environment," *Applied Sciences*, vol. 16, no. 1, pp. 225–225, 2025, DOI: 10.3390/app16010225.
- [5] S. Asadi, K. H. Naeini, D. Hassanlou, et al., "AI-powered digital twin frameworks for smart grid optimization and real-time energy management in smart buildings: A survey," *Computer Modeling in Engineering & Sciences*, vol. 145, no. 2, pp. 1259–1301, 2025, DOI: 10.32604/cmes.2025.070528.
- [6] U. M. Mushtaq, H. Venter, A. Singh, et al., "Advances in energy harvesting for sustainable wireless sensor networks: Challenges and opportunities," *Hardware*, vol. 3, no. 1, pp. 1–1, 2025, DOI: 10.3390/hardware3010001.
- [7] S. D., P. Swati, H. Sugato, et al., "Smart data processing for energy harvesting systems using artificial intelligence," *Nano Energy*, vol. 106, 2023, DOI: 10.1016/j.nanoen.2022.108084.
- [8] A. Ali, R. Eid, E. D. Manaseer, et al., "Dual-band 802.11 RF energy harvesting optimization for IoT devices with improved patch antenna design and impedance matching," *Sensors*, vol. 25, no. 4, pp. 1055–1055, 2025, DOI: 10.3390/s25041055.
- [9] N. Osman, A. K. Alnajjar, and C. R. Bansal, "A review of the cognitive radio IoT enabled radio frequency energy harvesting system," *Telecommunication Systems*, vol. 88, no. 4, pp. 134–134, 2025, DOI: 10.1007/s11235-025-01364-1.
- [10] G. Shen, X. Wei, K. Chi, et al., "Sum computation rate maximization for wireless powered OFDMA-based mobile edge computing network," *Computer Networks*, vol. 257, pp. 110961–110961, 2025, DOI: 10.1016/j.comnet.2024.110961.
- [11] K. Hirashima and T. Miyajima, "Sum rate maximization for multiuser full-duplex wireless powered communication networks: Regular section," *IEICE Transactions on Communications*, vol. E107.B, no. 8, pp. 564–572, 2024, DOI: 10.23919/transcom.2023ebp3165.
- [12] W. F. Mateo and A. Redchuk, "Artificial intelligence as a process optimization driver under Industry 4.0 framework and the role of IIoT, a bibliometric analysis," *Journal of Industrial Integration and Management*, vol. 9, no. 3, 2022, DOI: 10.1142/s2424862222500130.
- [13] M. Alex, "Radiofrequency energy harvesting systems for Internet of Things applications: A comprehensive overview of design issues," *Sensors*, vol. 22, no. 21, pp. 8088–8088, 2022, DOI: 10.3390/s22218088.
- [14] Y. Zhang, X. Wang, X. Luo, et al., "A survey on AI-empowered task-oriented sensing, communication, and computation in 6G networks," *Computer Science Review*, vol. 60, pp. 100899–100899, 2026, DOI: 10.1016/j.cosrev.2026.100899.
- [15] A. S. Khan, H. A. Kalifullah, K. Ibragimova, et al., "Integrating AI and IoT in advanced optical systems for sustainable energy and environment monitoring," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 5, 2024, DOI: 10.14569/ijacsa.2024.01505123.
- [16] J. A. Albarakati, "Blue energy intelligence: A hybrid wavelet–AI model for sustainable marine energy harvesting and hydrological signal monitoring via IoT systems," *Journal of Circuits, Systems and Computers*, prepublished, 2025, DOI: 10.1142/s0218126626420028.