



# A Survey of CPU Scheduling Performance in Real-Time Systems

<sup>1</sup>Sandeep Limbure, <sup>2</sup>Shlok Lawand, <sup>3</sup>Shreyash Pasalkar, <sup>4</sup>Atharva Patil, <sup>5</sup>Minal Deshmukh

<sup>1</sup>Mr, <sup>2</sup>Mr, <sup>3</sup>Mr, <sup>4</sup>Mr, <sup>5</sup>Prof

<sup>1</sup>Electronics and Telecommunications Department ,

<sup>1</sup>Vishwakarma Institute of Information Technology, Pune, India

**Abstract :** This study elaborates in detail on the performance of CPU scheduling algorithms in real time systems with an emphasis on the two environments: uniprocessor as well as multiprocessor environments. The algorithms to be discussed in this paper are rate monotonic, Earliest Deadline First and least laxity first. The sophisticated techniques that would be discussed in this paper involve the Overrun Server Method (OSM) and the Isolation Server Method (ISM). This study grades these algorithms based on those performance parameters which indicate the utilization of the CPU resources or the processing. Such end concerns involve schedulability, deadline miss rates, and context-switching overheads. Other than these, some more related issues of the task migration, handling overload in the multiprocessor system, as well as energy efficiency in the context of the multiprocessor, are discussed in further detail. Beyond that, the open research areas of dynamic scheduling, mixed-criticality systems, and the integration of machine learning for adaptive scheduling have been identified as a direction for increasing the efficiency, scalability, and safety of real-time systems in autonomous vehicle applications or in industrial automation.

**Keywords -** CPU scheduling, Real-time systems, scheduling methods like Rate Monotonic Scheduling (RMS), Earliest Deadline First (EDF), and Least Laxity First (LLF), multiprocessor scheduling, overload management, task migration, energy efficiency, a machine learning, adaptive scheduling, mixed-criticality systems.

## INTRODUCTION

Real-time systems are used for applications, ranging from mission-critical safety systems to multimedia and gaming environments[1]. An additional requirement of these systems is that, besides the logical correctness of computations, timing constraints are also very strict. In real-time systems, managing CPU scheduling is crucial for ensuring tasks are completed within their specified deadlines. In hard real-time systems, missing specified deadlines can result in serious failures, while soft real-time systems can usually tolerate occasional deadline misses to make better use of resources[2]. CPU scheduling in real-time systems focuses on managing processor resources to ensure tasks are completed within their required time limits[3]. The use of real-time systems is applied in complex application areas with increasing diversity, where there is a greater need for scheduling strategies to become even more apparent. These strategies must balance the need for high resource utilization with a critical requirement of meeting task deadlines[4].

Two common techniques in real-time scheduling algorithms are Rate Monotonic Scheduling and Earliest Deadline First[5]. Rate Monotonic Scheduling organizes tasks based on their periodicity, while Earliest Deadline First prioritizes them according to their deadlines[6]. However, multiprocessor environments pose additional problems since the system has to automatically handle resource allocation in the several CPUs, usually with sophisticated end algorithms of types Least Laxity First (LLF) or slightly modified variants of generic schedulers to deal with more complex workloads and to minimize the occurrence of missing soft deadlines.

This survey evaluates the performance of various CPU scheduling algorithms in real-time systems, considering both hard and soft real-time scenarios, to evaluate their efficiency for these two cases[8]. It gathers information about the most significant performance metrics, which include the following ones: CPU utilization, context switching overhead, and the ability to meet deadlines under varying system loads. Thus, the paper, by comparing different approaches, tries to provide insight into the most suitable scheduling strategies for different real-time system requirements.

## I. BACKGROUND AND KEY CONCEPTS

Real-time systems form the core of many safety-critical applications from different domains: automotive, communication, and health-related applications[10]. Its key characteristic is that the correctness of its operations relies not just on the logical accuracy of outputs, but also on delivering them on time. Real-time systems can be generally divided into two categories based on timing

constraints: hard real-time systems[11], which cannot afford any missed deadlines due to potentially disastrous consequences, and soft real-time systems, that can tolerate some deadline misses without severe implications but rather focus on general performance[12].

## II. KEY CONCEPTS

### Task Scheduling in Real-Time Systems

Task scheduling forms a very important part of RTS as it has defined how tasks are to be carried out on the system given their time and resource constraints. Effective scheduling is paramount to ensure all tasks are finished within their required time constraints. Scheduling algorithms may be split into the following categories:- Static Priority Scheduling: The RMS, RMA, and MV-RMS algorithms give a fixed priority for every task upon the basis of a period it takes. Its priority can thus be precalculated and then predictable when which one is to be executed.- Dynamic Priority Scheduling: Algorithms like the EDF change priorities at runtime. Its scheduling focuses on the task whose deadline would be nearest; this optimizes timely execution.

### Performance Evaluation

The performance of scheduling algorithms is evaluated by a few key metrics , such as: - CPU Utilization: The percentage of usage of CPU to execute the tasks, the higher utilization means higher efficiency.- Deadline Miss Rate: It represents how often the tasks get missed by their deadline, which is the most fundamental parameter to measure hard real-time systems.[12]-Response Time: Time elapsed between task activation and its completion. A shorter response time means a better system responsiveness and user satisfaction.

### Handling Overload and Overrun Scenarios

In real time systems, sometimes, tasks may exceed its expected execution time,which can create overload conditions. Techniques of overload control can be efficiently used to manage them by dropping tasks or by gracefully degrading. These techniques do not disturb the vital operation,while maintaining stability in the system..

### Scheduling in Multiprocessor Systems

With increasing levels of requirements of applications for more and more computational power, multiprocessor systems are an integral part of RTS. But the problem is further compounded by task scheduling across multiple processors with factors of resource sharing and overhead due to communication. Techniques used in multiprocessor scheduling include trying to increase “CPU utilization” while at the same time all tasks will be completed before their deadlines. Techniques such as load balancing and affinity scheduling are employed.

### Comparative Analysis of Algorithms

The comparative analysis highlights the multiple scheduling algorithms along with their relative strengths and weaknesses under different scenarios. For instance, although RMS is very simple and effective under low loads, it may undergo a loss in high-load scenarios as compared to EDF, which is though complex, exhaustively utilizing processor capacity under varied task loads. This necessity makes it important to recognize such dynamics for the appropriate selection of the scheduling approach against[2] the requirements of a real-time system.Here is a step-by-step organization for your questionnaire regarding the performance of CPU Scheduling in Real-Time Systems, along with a table to be used in summarizing findings

## III. COMPARATIVE ANALYSIS

### Scheduling

RMS achieves scheduling for periodic tasks where the CPU utilization is less than a certain threshold[13]. But its scheduling decreases sharply with the increase in the task load. EDF scheduling is the highest because it can even tolerate up to 100% CPU utilization, which makes EDF an appropriate algorithm for dynamic task sets in systems[14]. LLF With laxity, LLF provides flexibility as tasks are dynamic and set their priorities in the order of laxity. While excessive preemption may dominate its scheduling in real-time situations.

### CPU Utilization

EDF uses CPU better than other algorithms, with higher efficiency in using the system under dynamic workloads[16], while RMS has an excellent performance for static and periodic workloads but bad at higher utilization.OSM and ISM also surrender some utilization to go for overruns but maximize greater system stability with such conditions at high loads[18].

### Context Switching Overhead

RMS has lesser context switching overhead compared to the other “two” as it is static in nature, and while EDF and LLF have more overhead compared to others due to dynamic priority assignments. Since recalculation of laxity is involved, LLF is at the maximum context switching[21] .

### Overrun/Overload Handling:

RMS as well as EDF suffers from an overload problem since they cannot cope with tasks whose execution time exceeds execution time[22].OSM and ISM have specifically tackled overruns by segregating or even rescheduling tasks, hence making them stronger in overload situations. ISM fares better than OSM as Isolate the overrunning tasks completely, such that those do not interfere with other jobs within the system[24] .

Algorithm	Context Switching Overhead	Handling Overrun/Overload
Rate Monotonic Scheduling (RMS)	Low overhead since priorities are fixed, and context switching occurs less frequently	Struggles with overloads and fails to meet deadlines if CPU utilization exceeds the threshold.
Earliest Deadline First (EDF)	Moderate overhead due to dynamic priority assignment, as task priorities change over time	Handles overloads poorly. Missing one deadline can cause cascading failures in other tasks.
Least Laxity First (LLF)	Very high overhead due to frequent context switching as task laxity constantly changes.	Handles overload better than "RMS" but suffers from high overhead during overload conditions.
Overrun Server Method (OSM)	Moderate to high overhead, depending on how frequently tasks get reassigned to servers.	Performs well in systems prone to overruns, preventing cascading task failures.

Table 1.a. Comparison of overheads and Overload Handling

Table 1.a. compares real-time scheduling algorithms based on context switching overhead and their ability to handle overrun or overload conditions. For example, RMS has low overhead but cannot consider overload scenarios, while for instance, EDF and LLF have higher overhead because of the changing of priority too frequently and therefore more lenient with high system loads. Overrun Server Method and Isolation Server Method are better suited for overloads, but ISM, indeed, provides excellent performance by isolating tasks from cascading failures. However, slight overhead might rise with these methods in comparison to traditional scheduling techniques.

Algorithm	Schedulability	CPU Utilization
Rate Monotonic Scheduling (RMS)	Static priority-based, works best for periodic tasks. Feasibility guaranteed if CPU utilization $\leq 69.3\%$ for large task sets.	[16] Utilization is limited by the Liu and Layland bound of 69.3% for periodic tasks.
Earliest Deadline First (EDF)	Optimal for uniprocessor systems. Ensures scheduling as long as CPU utilization $\leq 100\%$ .	Can achieve up to 100% utilization if the task set is schedulable.
Least Laxity First (LLF)	The dynamic priority assignment based on laxity (remaining time before the deadline minus execution time).	Can achieve high CPU utilization, similar to EDF.
Overrun Server Method (OSM)	Designed for systems where tasks may overrun their execution time	Utilization depends on how efficiently overrun tasks are handled

Table.1.b.Comparison of Schedulability and Cpu Utilization

Table 1.b. displays the schedulability and CPU utilization of various real-time scheduling algorithms. RMS is used on periodic tasks and had a limit of 69.3% in CPU utilization, whereas EDF scheduling algorithm was able to reach 100% utilization for schedulable sets of tasks. The other algorithm with high utilization like that of LLF relies on dynamic priority adaptation. Both the Overrun Server Method and the Isolation Server Method handle overruns. However, ISM is better suited for utilization in systems that quite frequently overrun their tasks for its mechanism of isolation.

#### IV.CASE STUDIES

##### Case Study 1: Rate Monotonic Scheduling for Embedded Automotive Systems

Examples of embedded systems in cars include anti-lock braking systems and the engine[25]. Thus, control units typically are real-time systems, and timing correctness must be enforced strictly. This will include processing sensor data and executing components like Third, brakes or engines in very short, predictable timeframes to avoid accidents on the passengers. Rate Monotonic Scheduling is the most widely used in automotive embedded systems because the task cycle of sensor data collection and processing can so perfectly fit into[26]. Static Priority Tasks like monitoring the speed of the wheel through the ABS system have short time intervals

that follow a predictable pattern and are located above those with longer periods like engine diagnostics, are considered third priority. Using RMS ensures that the high priority tasks-most of them are still related to safety-always is completed on time. However, RMS has a restriction of the CPU utilization,  $\approx 69.3\%$  means that system designers should design the workload very carefully to avoid Overloading the Processor Too much since more functionalities will be included in new automobiles. In cases where system loads approach the RMS utilization bound, some runs of jobs fail to meet deadlines, it causes degradation in performance. These constraints makes the designers to the use of more complex algorithms, like EDF or an RMS together with Preventive measures in new automotive systems.

### Case Study 2 Earliest Deadline First (EDF) in Multimedia Streaming Systems

Real-time video streaming applications or video conferencing applications while live processing is critical to ensure low latency smooth playback, and these applications often consist of dynamic task sets with changing deadlines; as such, static scheduling is impractical[27]. Earliest Deadline First is used often in the multimedia systems because it has the ability to Dynamically it updates priorities based on deadlines for tasks[28]. Suppose for a video streaming system in which render and audio are synchronized with a deadline for instance[8], depending on the frame rate (e.g., 30 frames per second), the nearest deadline is always most important of all. EDF maximizes CPU utilization by scheduling tasks with tight deadlines, such as rendering. This dynamic flexibility results in this technology improves the user experience of video streaming because it avoids frame drops and ensures Audio-video synchronization. Despite its merits, EDF may become erratic in cases of system overload where the deadline is not met, jitter or dropped frames can occur. In such implementations without overload cases, resilience needs supplementary mechanisms that should include overloads. Control or resource reservation, so critical tasks can still meet their deadlines.

### Case Study 3: Overrun Server Method in Industrial Robotics

An RTCS in industrial robotics is used primarily for control of movements of robotic arms, sensors, and actuators used in automated manufacturing processes[29]. These systems must deal with activities of different execution times and sometimes, due to inevitably, sensor delays or other unexpected obstacles occur[30]. This overloads the server that is used in these settings to serve the task. Cause delays without the delay affecting stability of systems. In this case, tasks that overrun their execution time of allocated tasks, such as recalculation of a path for the robotic arm, are frozen and are redirected to an overloaded server for completion. This way, the heart of the scheduling mechanism focuses on processing other real-time tasks, ensuring critical operations (e.g., safety) all collision avoidance checks remain within their time bound. OSM keeps industrial robot systems operational in the case that critical safety tasks continue running, even if non-critical tasks are overrun. It improves Strength of the system, which consequently reduces the shock of unrelated failures that may trigger the assembly line. OSM makes sure all crucial tasks meet deadlines. The difficulty of managing overrun servers increases system overhead. Additionally, tuning the server parameters (e.g. priority levels) requires careful consideration to avoid inefficiencies during Routine activity.

### Case Study 4: Isolation Server Method in Healthcare Monitoring Systems

For example, real-time health monitoring systems such as those in a critical care unit continuously monitor the patient's vital signs for example, oxygen levels, heart rate, and must call medical staff when thresholds are violated. Such systems can afford to miss deadlines, since missing deadlines may delay responses to life-threatening messages conditions. This is utilized in the Isolation Server Method (ISM) of health monitoring systems to handle overloads and overruns Critical tasks such as keeping awake to important signs are kept strictly apart from ,for instance, data logging for retrospective analysis, which is less hazardous. If less hazardous tasks overruns, they are held in several isolated servers, which cancels the effect on important monitor activities. Through ISM, health monitoring systems produce critical alerts in real time, saving precious time without a moment's delay on background tasks. This feature makes the system capable effectively even during high loads, ensuring that medical staff are alerted to critical events with minimal delay. Even though multiple does give more system stability, overhead of having multiple Isolated servers increases the complexity of the system. Furthermore, getting proper isolation While the location is crucial for balancing performance and resource utilization.

## V. OPEN RESEARCH PROBLEMS

CPU Scheduling in Real-Time Systems As the technology advanced in real-time systems various problems were thus encountered with CPU scheduling. Unsolved, not least by the complexity of modern systems, as including multiprocessor environments, IoT, and energy constraints. Here is integrated open recent studies in literature for the issues

**1. Dynamic Scheduling in Multiprocessor and Distributed Systems-** Since most the real time systems rely on multiprocessor and distributed this architecture leads to lower scheduling and predictability[10]. The current algorithms used are Rate Monotonic Scheduling (RMS) and Earliest Deadline First (EDF), which cannot scale with multiprocessor systems. The existing scheduling algorithms for non-aborting faults do not consider the task migration, load balancing as well as synchronization across processors. Design dynamic and scalable scheduling algorithms that manage to effectively handle task migration and synchronization across several processors while reduce overhead .Examine hybrid dynamic /static algorithms for load-balanced in Distributed environments will become the norm.

### 2. Overload Handling and Resource Management:

Overload conditions, where system tasks go beyond the CPU capacity, can have the following consequences: missed deadlines and degraded system performance. The algorithms like EDF are prone to under severe loads. State-of-art techniques haven't considered the sudden occurrence of overloads or unpredictable task execution times. Adaptive Scheduling Algorithms: Investigate adaptive scheduling algorithms that adaptively Make adjustments on priorities or redistribute resources during overloads. Methods include

early overload detection, dynamic Task Management, and graceful degradation of non-critical tasks require Further modifications for the purpose of ensuring critical task execution once downloaded.

**3. Energy Aware Scheduling in Real-Time Systems:** Real-time systems in the environment of embedded and mobile often operate under tight energy constraints[11]. The algorithms reported here focus on timeliness but ignore Power consumption. Majority of the energy-efficient scheduling solutions do not balance the task with the best possible performance and energy optimization, particularly in multiprocessor systems. Develop power-aware scheduling algorithms that dynamically adjusts task execution and processor usage, based on real-time energy end availability, and may even encompass Dynamic Voltage and Frequency Scaling[20]. DVFS or energy-aware task allocation.

#### 4. Real-time scheduling cloud, IoT and heterogeneous systems:

Classic advancement of cloud computing and IoT has also come along with new challenges such as distributed task execution, resource heterogeneity, varying network conditions. Traditional algorithms are inappropriate for a real-time scenario of tasks for scheduling in dynamic, distributed or heterogeneous environments, high latency or resource-scarce networks, diversity in variability, and processor diversity make scheduling complicated. Design new scheduling algorithms targeting distributed hardware resource availability systems, incorporating delay in communication heterogeneity. Technologies are integrated with fog, edge computing and It provides real-time guarantees through heterogeneous processor coordination.

#### 5. Safety and Fault Tolerance in Real-Time Scheduling:

As more meshing and critical, the real-time systems are steadily at risk of Increase in cyber attack and system failure threats. Current algorithms are not Adequate explanation of security weaknesses or capacity to recover from faults. Research Gap: Very less attention towards incorporating fault tolerance and security measures incorporated into real-time scheduling algorithms, resulting in either system compromise or Failure. Fault-tolerant and security-aware scheduling. Algorithms capable of detecting and countering attacks in real-time while rescheduling tasks to Manage system failures. Solutions must be designed with methods for resource isolation and real-time regain performance without losing performance.

#### 6. AI and Machine Learning Enhanced Scheduling:

Real-time system scheduling is traditionally rule-based, which may not adapt well to dynamic and time-varying workloads. Research Gap: The integration of machine learning (ML) and artificial intelligence (AI) in Real-time scheduling is at its nascent stage with nearly no practical implementation. Develop ML and AI techniques for scheduling improvement. Decisions predicting the system load and behavior of tasks. Reinforcement learning and AI-driven adaptive scheduling algorithms can thus be applied for better real-time performance or optimize resource allocation dynamically.

## VI. CONCLUSIONS

The above real-time systems' CPU scheduling performance survey showed immense importance towards the scheduling algorithms in which strict timings of tasks are met with optimum usage of system resources. From the viewpoint of uniprocessor systems, RM and EDF are the dominant players. RM maintains “predictability” but deviates toward underutilization, while EDF excels as a CPU utilization “algorithm” but falls flat bad under overloads. Algorithms like Least Laxity First and Maximum Urgency First are “flexible” but consume high overheads in context switching. Some of the specialized algorithms developed for multiprocessor systems deal with complicated issues like task migration and synchronization. There are substantial research gaps indicated in dynamic scheduling, overload management, and the handling of mixed-criticality tasks, especially in multiprocessor, cloud, and IoT systems.

## VII. REFERENCES

- [1] A. Maurya and M. Tripathi, "Benchmarking Task Scheduling Algorithms in Heterogeneous Systems," *High-Performance Computing Journal*, vol. 12, no. 3, pp. 45-59, Jan, 2023.
- [2] R. Sirisah and R. Prasad, "MPEFT: Maximizing Parallelism for Minimizing Earliest Finish Time," in *Proc. Int. Conf. High-Perf. Comput.*, 2023, pp. 21-30.
- [3] M. Mangalampalli, S. Kumar, and R. Hayat, "Deep Reinforcement Learning for Real-Time CPU Scheduling," *IEEE Access*, vol. 11, pp. 45,663-45,675, Jul. 2023.
- [4] H. Zheng and Y. Mao, "Shockwave: Optimizing Scheduling for Dynamic Environments," in *Proc. ACM SIGMETRICS*, 2023, pp. 53-65.
- [5] K. Li, S. Wang, and D. Lee, "Adaptive Batch-Stream Scheduling in Multiprocessor Systems," *J. Comput. Arch. Eng.*, vol. 10, no. 4, pp. 110-121, Dec. 2022.
- [6] A. Hayat et al., "Work-Stealing Load Balancing in Real-Time Scheduling," in *Proc. IEEE Int. Conf. Computer. Sys.*, 2023, pp. 11-20.
- [7] G. Saifullah and R. Shah, "Energy-Aware Real-Time Scheduling with Deadline Guarantees," *Real-Time Systems*, vol. 19, no. 5, pp. 511-529, Aug. 2023.

- [8] S. Lee and J. Kim, "RTGPU: Real-Time GPU Scheduling of Parallel Tasks," *IEEE Trans. Comput.*, vol. 21, no. 3, pp. 123–138, Mar. 2023.
- [9] S. Wu et al., "Multi-Core CPU Scheduling Optimization using AI Algorithms," *Future Generation Computer. Syst.*, vol. 140, pp. 44–56, Jun. 2023.
- [10] T. Nand and K. Roy, "Hybrid RMS-EDF for Mixed Criticality Systems," *IEEE Embedded Systems Mag.*, vol. 29, no. 2, pp. 34–42, 2022.
- [11] M. Parikh, "Real-Time Scheduling for Internet of Things Devices," in *Proc. IEEE IoT Conf.*, 2023, pp. 33–47.
- [12] S. Kumar, "Comparative Analysis of Rate-Monotonic and Earliest Deadline First Scheduling in Automotive Systems," *IEEE Access*, vol. 11, pp. 56,743–56,754, May 2023.
- [13] X. Zhou et al., "Machine Learning for CPU Scheduling Optimization," *J. Artif. Intell. Res.*, vol. 74, pp. 113–124, 2022.
- [14] H. Xu and W. Li, "Dynamic Scheduling Algorithms for Battery-Powered IoT Devices," *IEEE Embedded Systems Letters*, vol. 15, no. 3, pp. 31–40, 2022.
- [15] R. Singh et al., "Scalable Real-Time Scheduling for Cloud Systems," in *Proc. IEEE Int. Conf. Cloud Comput.*, 2023, pp. 23–34.
- [16] Zhang, "Adaptive Priority-Based Task Scheduling in Heterogeneous Environments," *J. Syst. Arch.*, vol. 116, pp. 33–42, Mar. 2023.
- [17] A. Patel et al., "Deadline-Aware Scheduling for High-Performance Embedded Systems," *IEEE Access*, vol. 10, pp. 73,456–73,469, Dec. 2022.
- [18] J. Brown et al., "Energy-Aware Real-Time Scheduling in Multiprocessor Systems," in *Proc. ACM SIGARCH Conf.*, 2023, pp. 98–110.
- [19] M. Gao and X. Yu, "Deep Learning for Adaptive CPU Scheduling," *Neural Comput.*, vol. 34, no. 7, pp. 1–10, Jul. 2023.
- [20] S. Lopez, "Task Overrun Management in Industrial Robotics Using OSM," *IEEE Robot. Autom. Lett.*, vol. 8, no. 3, pp. 42–50, Apr. 2023.
- [21] D. Kapoor et al., "Isolation Server Method for Healthcare Real-Time Systems," in *Proc. IEEE HealthTech Conf.*, 2022, pp. 19–30.
- [22] T. Chen, "Integrating Machine Learning into CPU Scheduling for Cloud Workloads," *IEEE Cloud Comput.*, vol. 9, no. 1, pp. 14–22, 2023.
- [23] P. Ravi and G. Swaminathan, "Overload Handling in Real-Time IoT Networks," *Int. J. Comput. Appl.*, vol. 175, no. 3, pp. 22–31, 2022.
- [24] A. Jain, "Scheduling Algorithms for Multi-Core Systems: A Survey," *Comput. Commun.*, vol. 189, pp. 56–71, May 2023.
- [25] M. Nakamura, "Energy-Conscious Scheduling with DVFS in Real-Time Systems," in *Proc. IEEE Int. Symp. Low-Power Electron.*, 2022, pp. 43–57.
- [26] J. Martinez, "Designing Robust Real-Time Schedulers for Autonomous Vehicles," *IEEE Trans. Autom. Veh.*, vol. 14, no. 2, pp. 210–220, Apr. 2023.
- [27] S. Gupta, "Performance Metrics for Mixed-Criticality Scheduling," *ACM Trans. Embedded Computer. Syst.*, vol. 20, no. 5, pp. 1–10, Sep. 2023.
- [28] D. Lee et al., "Advanced Schedulability Analysis for EDF-Based Systems," *Real-Time Systems Journal*, vol. 59, pp. 114–130, Oct. 2022.
- [29] H. Nguyen, "Machine Learning-Driven Task Scheduling for Energy Optimization," *ACM Trans. Cyber-Phys. Syst.*, vol. 14, no. 4, pp. 36–48, Jun. 2023.