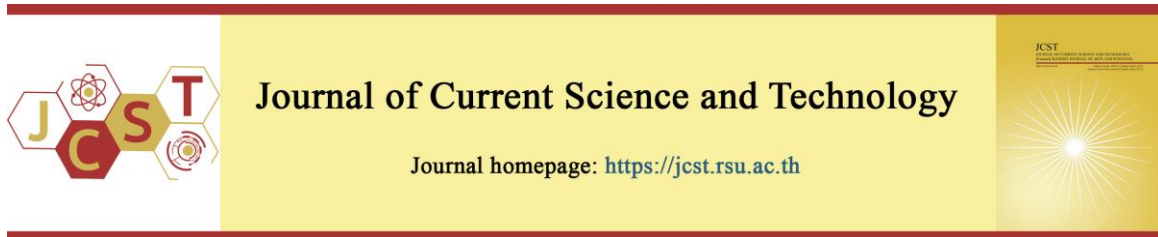


Cite this article: Pawadee, K., & Keeratipranon, N. (2026). Energy-efficient scheduling and control of aluminum melting processes using genetic algorithms and reinforcement learning. *Journal of Current Science and Technology*, 16(3), Article 205. <https://doi.org/10.59796/jcst.V16N3.2026.205>



## Energy-Efficient Scheduling and Control of Aluminum Melting Processes Using Genetic Algorithms and Reinforcement Learning

Kraiwitch Pawadee  
Narongdech Keeratipranon\*, ORCID: 0000-0001-9331-3716

Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand

\*Corresponding author; E-mail: [Narongdech.K@chula.ac.th](mailto:Narongdech.K@chula.ac.th)

Received 20 March 2026; Revised 4 April 2026; Accepted 30 April 2026; Published 25 June 2026

### Abstract

Manufacturing is a primary driver of global energy consumption, necessitating advanced production strategies to maintain industrial competitiveness. This study explores opportunities for thermal optimization within an industrial aluminum melting process in Thailand. While traditional manual furnace regulation and heuristic scheduling provide operational stability, the inherent complexity of multi-stage processes presents opportunities to further minimize holding and reheat losses through automated synchronization. However, existing studies typically address batch-level thermal control and system-level scheduling separately, and few explicitly target the holding and reheat losses caused by unsynchronized transfer between induction and melt-and-hold stages. To address this gap, we propose an integrated framework that combines deep reinforcement learning (RL) for batch-level power control with a genetic algorithm (GA) for system-level energy-aware scheduling. The RL component uses a deep Q-network (DQN) to learn an expert-guided power-control policy, while the GA incorporates a just-in-time (JIT) Gate, a synchronization mechanism that delays or releases melting batches to align induction-furnace output with downstream melt-and-hold (M&H) demand. Under repeated mixed-start evaluation, the revised RL controller achieved  $597.96 \pm 18.23$  kWh per batch, slightly improving upon the original DQN result ( $600.79 \pm 19.96$  kWh) and the fixed expert-profile baseline (601.73 kWh), while preserving 100% target attainment and zero executed safety violations. At the system level, the GA-based scheduler eliminated reheat energy from 2,833.3 kWh to 0 kWh and reduced peak demand by 23.9%. These results indicate that RL provides robust, near-expert batch-level control, while GA-based JIT synchronization delivers the major energy-saving benefit at the production-system level.

**Keywords:** *aluminum melting, deep Q-network, energy-aware scheduling, genetic algorithm, reinforcement learning, temperature control, industrial process control, thermal process optimization*

### 1. Introduction

Manufacturing industries account for approximately one-third of global energy demand (Fernandes et al., 2022), with high-temperature processes such as metal melting being among the most energy-intensive sectors. In the aluminum industry, electricity costs often represent a substantial portion of total production expenses, exceeding 30% in many industrial contexts (Braga & Netto, 2016), making energy efficiency a critical lever for both economic

competitiveness and environmental sustainability. Although energy optimization has been studied in both production scheduling and process control, these two decision layers are often treated separately. In aluminum melting, this separation creates a practical synchronization gap: melting may finish earlier than downstream demand, causing unnecessary holding and reheating losses, while power control decisions made at the batch level still affect energy use, peak demand, and process consistency.

This study investigates a real-world heating-plate production line in Thailand. The process involves melting aluminum ingots in induction furnaces (IFs) followed by temperature maintenance in melt-and-hold (M&H) furnaces before casting. Operational data reveal that temporal variations between melting completion and downstream casting readiness can result in molten metal being held at high temperatures longer than necessary or requiring reheating; addressing these variations represents a key area for enhancing efficiency. Furthermore, while batch-level power adjustments are guided by operator experience, maintaining absolute consistency across diverse conditions naturally results in variability in processing duration (87.0 – 129.0 minutes) and energy consumption (496.0 – 735.0 kWh). From 103 filtered historical batches, the average energy consumption and duration were  $587.31 \pm 34.85$  kWh and  $101.93 \pm 7.14$  minutes, respectively, with a strong positive correlation between duration and energy use ( $r = 0.745$ ). These observations indicate that both waiting-related thermal losses and batch-level control variability are meaningful contributors to inefficiency.

Existing research in this field typically addresses scheduling and control as separate layers; scheduling methods primarily focus on makespan minimization or meeting job due dates (Nayak et al., 2017), while control methods focus on process stability and temperature tracking (Gao et al., 2020). Although metaheuristics such as the genetic algorithm (GA) have shown success in improving energy efficiency in job-shop environments (Salido et al., 2016), they face limitations in addressing the dynamic fluctuations caused by system-level waiting. Conversely, Reinforcement Learning (RL) has emerged as a powerful tool for energy-aware control in multi-stage production systems (Loffredo et al., 2023), but most prior studies do not explicitly address the synchronization problem between induction-furnace output and downstream melt-and-hold demand in multi-furnace aluminum systems. This leads to two distinct research gaps. First, integrated RL+GA frameworks remain limited for industrial furnace operations that require both batch-level thermal control and system-level scheduling. Second, prior scheduling studies do not explicitly incorporate a just-in-time synchronization mechanism to reduce holding and reheat losses at the IF-to-M&H handoff. For example, Cui and Yuan (2024) studied hybrid RL+GA scheduling in photovoltaic glass production but did not address furnace-specific holding and reheat losses in a multi-stage aluminum melting process.

To enhance system performance, this paper proposes an integrated framework that combines GA for macro-level scheduling with deep RL for micro-level power control. The RL layer is used to learn an expert-guided batch-level power-control policy under safety constraints, while the GA layer evolves scheduling parameters that align melting completion with downstream demand. A just-in-time (JIT) Gate is introduced as a synchronization mechanism that delays or releases melting batches according to M&H inventory and demand timing, thereby reducing unnecessary holding and reheating. The framework aims to satisfy the 950°C quality constraint while improving batch-level control consistency and reducing system-level energy waste and peak demand under real-time electricity pricing. The main contributions of this study are as follows:

- **Integrated Control and Scheduling Framework:** We develop a two-layer framework that combines deep RL for batch-level power modulation with GA for plant-level production scheduling.

- **just-in-time (JIT) Gate Mechanism:** We introduce a JIT Gate that explicitly synchronizes induction-furnace completion with melt-and-hold (M&H) demand, reducing holding and reheat losses at the inter-stage handoff.

- **Empirical Evaluation in an Industrial Setting:** We evaluate the proposed framework using real plant operating data and simulation-based analysis, showing that the revised RL controller provides slight energy improvement over the original DQN and fixed expert-profile baseline at the batch level, while the GA-based scheduler delivers the major system-level benefit by eliminating reheat energy and reducing peak demand.

## 2. Objectives

1. To formulate a system-level model for an IF–M&H–casting production line, by defining energy metrics and holding/reheat losses under industrial constraints, with a quality target of 950°C.

2. To develop an expert-guided RL furnace controller that achieves the 950°C target with robust and energy-aware batch-level power control under safety constraints.

3. To develop a GA-based scheduling approach with JIT synchronization to eliminate system-level holding losses while maintaining throughput.

4. To evaluate the integrated framework against plant operating data, focusing on batch-level control consistency, system-level energy reduction, peak demand shaving, and operational robustness.

### 3. Related Work and Literature Review

The transition toward sustainable manufacturing has increased interest in computational methods for managing high-temperature, energy-intensive processes. In the aluminum industry, where electricity costs can exceed 30% of total production expenses, reducing energy use is both an economic and sustainability priority. Prior studies have examined scheduling, process control, and hybrid optimization separately; however, fewer studies explicitly address the coupling between batch-level furnace control and system-level synchronization in multi-stage aluminum melting lines. This section reviews the relevant literature in four areas: aluminum melting efficiency, metaheuristic scheduling, RL for process control, and integrated hybrid approaches.

#### 3.1 Energy Efficiency in Aluminum Melting and Processing

Aluminum smelting and melting are sensitive to process timing, where unsynchronized operations lead to substantial thermal waste. Traditional approaches to furnace efficiency have relied on thermal state management and discretized thermal dynamic models to control power modulation (Wong et al., 2023). Despite these controls, manual adjustments driven by operator experience often result in high variability in processing time and energy consumption. Recent literature emphasizes that optimizing individual furnace parameters is insufficient without considering the broader production schedule.

#### 3.2 Metaheuristics for Industrial Production Scheduling

Genetic Algorithms (GAs) have been widely implemented in complex scheduling problems due to their robust capability to explore expansive search spaces. Historically, GAs have been utilized to minimize makespan in factory production sequences (Park et al., 2021). However, the focus has shifted toward energy-aware scheduling in job-shop and flow-shop environments (Dai et al., 2025).

Modern iterations, such as the Non-dominated Sorting Genetic Algorithm II (NSGA-II), have been adapted for multi-objective scheduling under flexible time-of-use (TOU) electricity pricing (Chen et al., 2023). Furthermore, research has explored managing peak load constraints to mitigate demand charges and optimizing multi-factory scheduling under real-time pricing (Ahmad Khan et al., 2026). While metaheuristics excel at macro-level sequencing, they often lack the

real-time flexibility required to manage microscopic thermal fluctuations during a melting batch.

#### 3.3 Reinforcement Learning in Manufacturing and Process Control

RL provides an inferential perspective for production management, allowing systems to learn optimal policies through interaction with environmental feedback. In multi-stage production lines, RL has demonstrated superior performance in managing parallel machine workstations and energy-efficient demand response (Lu et al., 2020; Modrak et al., 2024).

The adaptability of RL makes it ideal for industrial microgrid energy management, where it can be combined with other optimization algorithms to lower electricity costs (Loffredo et al., 2023). By formulating power regulation as a Markov decision process (MDP), RL agents can effectively track temperature targets while minimizing electricity consumption. However, in many industrial thermal processes, dense continuous sensing is unavailable, making direct online learning impractical. This motivates the use of simulation-based RL environments that act as control-oriented surrogates under limited sensor coverage, allowing policies to be trained safely before evaluation against plant operating data.

#### 3.4 Integrated and Hybrid Optimization Approaches

The integration of system-level scheduling and process-level control is an emerging trend for sustainable manufacturing (Li et al., 2024). Hybrid models that combine GAs with RL allow scheduling decisions to respond to real-time process volatility (Cui & Yuan, 2024). Nevertheless, the limitations of prior hybrid studies remain important in the context of aluminum melting. Li et al. (2024) addressed integrated system-process control from a broader manufacturing perspective, while Cui and Yuan (2024) investigated hybrid RL+GA scheduling in photovoltaic glass production. These studies support the value of hybrid optimization, but they do not explicitly address furnace-specific holding and reheat losses arising from unsynchronized transfer between induction furnaces and downstream melt-and-hold stages. Accordingly, this study focuses on two gaps: (i) the limited use of integrated RL+GA frameworks for industrial furnace operations requiring both batch-level thermal control and plant-level scheduling and (ii) the absence of an explicit just-in-time

synchronization mechanism for the IF-to-M&H handoff. The proposed JIT Gate is designed to address this second gap by linking batch release decisions to downstream inventory and demand timing, thereby bridging the macro-scheduling and micro-control layers.

#### 4. Materials and Methods

##### 4.1 Problem Definition and System Constraints

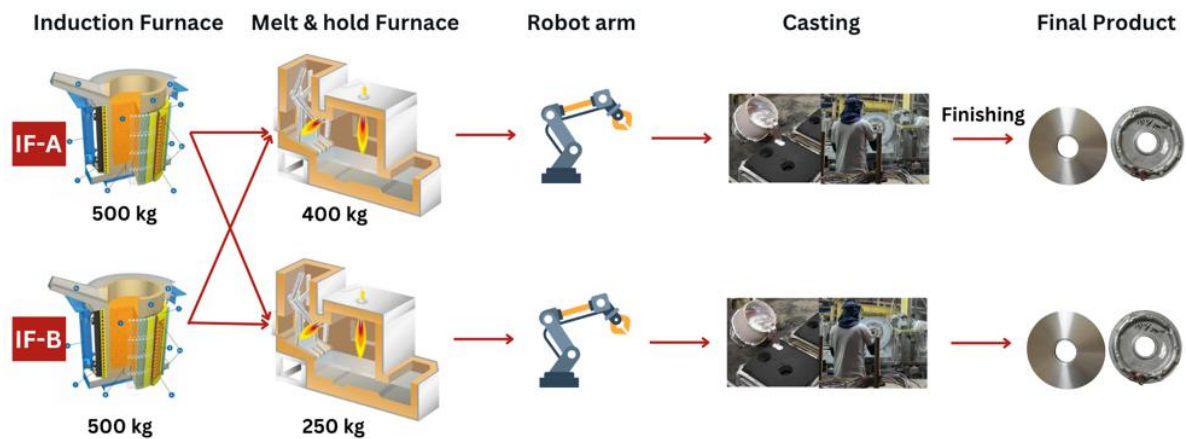
The production process follows a batch-based workflow, beginning with the melting of aluminum ingots in an induction furnace (IF) to produce a molten alloy. Once the target state is reached, the melt is transferred to a melt-and-hold (M&H) furnace to maintain thermal stability and prepare for the casting phase. This preparation includes flux stirring for melt purification before a robotic arm transfers the material to casting molds. The cast components then undergo secondary machining and finishing to meet final surface specifications (see Figure 1). The system comprises two IF units and two M&H units with unequal capacities: 400 kg for M&H-A and 250 kg for M&H-B. To maximize operational efficiency, each IF melting cycle processes a standardized 500 kg lot, which can be split poured into both M&H furnaces to accommodate capacity variations.

A significant operational constraint arises from the shared electrical infrastructure: both IF units utilize a common power source, limiting the system to one active melting cycle at any given time. However, to maintain high productivity, the system allows the preparation and loading of ingots in the idle IF while the other is active. This alternating pattern (IF-A/IF-B) is strategically employed to minimize idle time and

ensure a continuous supply of molten metal, particularly during peak production periods.

The withdrawal rate of molten metal from the M&H stage is inherently linked to the consumption requirements of the casting station, which fluctuate based on daily product dimensions and mold sizes. High-consumption scenarios (large molds) require frequent melting cycles, resulting in a tightly synchronized schedule. Conversely, in scenarios with smaller molds, the molten metal naturally remains in the M&H stage for longer durations. These temporal variations present an opportunity for advanced energy management, specifically in reducing the energy required for extended holding or reheating. Consequently, the proposed model focuses on two integrated decision layers: (i) system-level scheduling and lot allocation to optimize M&H inventory levels and (ii) batch-level power control to achieve quality requirements with high energy precision.

The governing system constraints for this study are summarized as follows: (1) simultaneous melting in both IF units is restricted by the shared power source, although pre-loading is permitted to support immediate alternation; (2) each IF batch is standardized to a 500 kg lot; (3) M&H inventory management must respect the furnace capacities of 400 kg (M&H-A) and 250 kg (M&H-B), with split-pouring allowed for flexibility; (4) the batch-level quality constraint is defined by an end-of-lot target temperature of 950°C; and (5) the framework must remain feasible across different production modes, ranging from high-consumption days requiring tight synchronization to lower-consumption days in which extended holding and reheating losses become more likely.

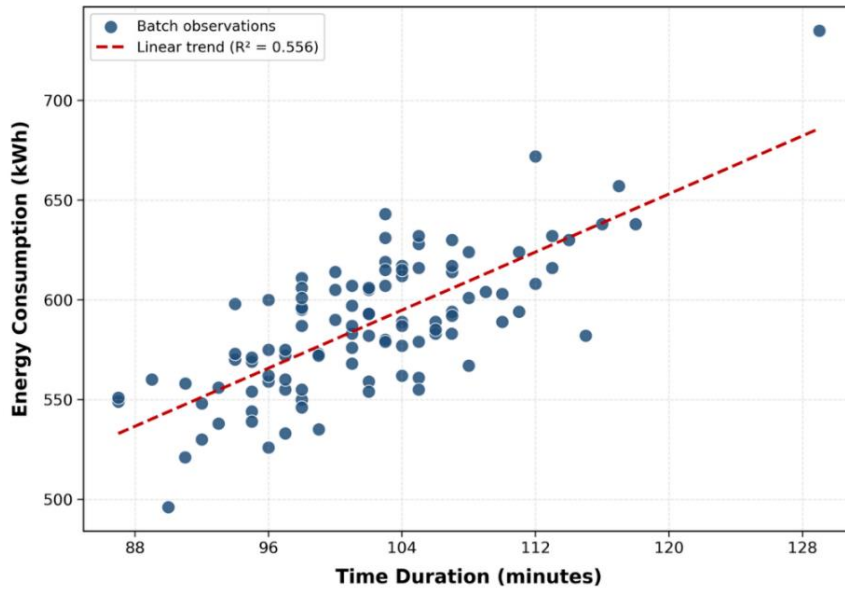


**Figure 1** Overall process flows from the induction furnace (IF) to finishing operations

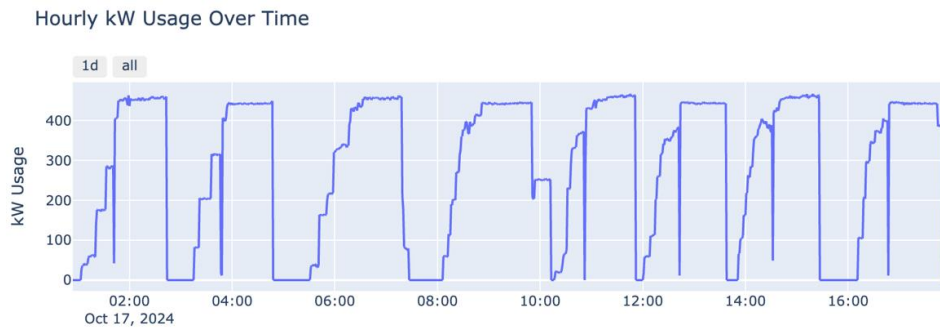
#### 4.2 Reinforcement Learning for Melting Power Control

The first operational focus of this work is the optimization of power-level regulation during each induction-furnace (IF) melting batch. Historical operational logs indicate that power adjustments are currently based on the nuanced expertise and experience of operators, which provides the flexibility needed to manage varying plant conditions. However, a retrospective analysis of 103 filtered batches reveals natural diversity in operational outcomes, with melting durations and energy consumption ranging from 87.0 to 129.0 minutes and from 496.0 to 735.0 kWh, respectively. The average melting duration was  $101.93 \pm 7.14$  minutes, with a strong correlation ( $\approx 0.745$ ) between duration and energy use. These results highlight a significant opportunity to assist operators with a standardized decision-support tool aimed at further enhancing energy efficiency and operational

consistency (Figures 2 and 3). The objective is to develop a robust and energy-aware melting policy that reliably achieves the end-of-batch quality target while optimizing the processing timeframe. The quality requirement, denoted as  $T_{\text{target}}$ , is specified as  $950^{\circ}\text{C}$  for all experimental trials. Given that direct online trial-and-error learning on an active furnace is not feasible due to the priority of maintaining safety and production stability, a simulation environment was developed. This environment captures the dominant thermal behavior of the melting process and serves as a robust training platform for the RL agent. The resulting policy was then evaluated alongside traditional operational patterns to demonstrate its potential for supporting energy precision, target attainment, and safe batch-level control.



**Figure 2** Relationship between batch energy consumption and processing time



**Figure 3** Baseline power regulation pattern

#### 4.2.1 Thermal Simulation Model

Because the real system lacks continuous temperature instrumentation, temperature observations are only available at limited time windows during a batch, after the melt becomes measurable. To enable safe learning while retaining physical plausibility, we adopt a lumped-parameter energy-balance model that links applied electrical power to melt temperature evolution, while accounting for the most influential heat-loss mechanisms. In its compact form, the thermal balance can be expressed as follows:

$$mc_p \frac{dT}{dt} = \eta P(t) - Q_{\text{loss}}(T) \quad (1)$$

where  $m$  is the melt mass,  $c_p$  is the effective specific heat,  $P(t)$  is the electrical power input, and  $\eta$  is an effective conversion factor capturing the portion of electrical power that contributes to useful heating. Heat loss is modeled as a combination of convection and radiation:

$$Q_{\text{loss}}(T) = hA(T - T_a) + \epsilon\sigma A[(T + 273.15)^4 - (T_a + 273.15)^4] \quad (2)$$

Equation (1) uses an effective sensible-heat formulation rather than an explicit enthalpy-based phase-change model. This simplification was adopted because continuous temperature measurements are unavailable during the early solid-to-liquid transition, making direct identification of latent-heat dynamics impractical from the available plant data. Accordingly, the model is intended as a control-oriented lumped surrogate rather than a full thermophysical melting model. The effective parameters absorb unmodeled phase-change effects into calibrated coefficients while preserving the main control trade-off: increasing power accelerates heating but increases energy consumption, whereas heat losses increase with temperature and penalize inefficient control patterns.

To improve reproducibility, calibration and validation were performed using a targeted dataset of 30 specialized melting batches collected under controlled operating conditions.

Instead of relying on direct continuous temperature tracking, the validation approach focuses on the total energy absorbed to reach the target temperature. These 30 batches were systematically categorized based on their applied maximum power limit ( $P_{\text{max}}$ ): 10 batches operating at 450 kW, 10 batches at 475 kW, and 10 batches at 500 kW. For each  $P_{\text{max}}$  level, a stratified 70:30 split was used,

resulting in seven calibration batches and three validation batches per power level, or 21 calibration batches and nine validation batches in total. The calibration subset was utilized to estimate and fine-tune essential system parameters, such as the effective specific heat ( $c_p$ ), electrical conversion efficiency ( $\eta$ ), and heat loss coefficients by matching the model's energy absorption to the actual recorded energy. The remaining independent validation subset was subsequently used to verify the model's predictive performance, ensuring that the simulation accurately replicates real-world processing times and total energy consumption across varying power limits. This validation strategy is based on batch-level energy and time-to-target agreement rather than dense temperature trajectories, consistent with the limited sensing available in the plant.

#### 4.2.2 MDP Formulation and Learning Objective

We formulate the melting of a single batch as a finite-horizon Markov Decision Process (MDP)  $(S, A, p, r, \gamma)$ . At each decision step, the agent observes a state  $s_t \in S$ , selects an action  $a_t \in A$  representing a discrete power adjustment level, and receives a reward  $r_t$ . The objective is to learn a policy  $\pi(a|s)$  that maximizes the expected discounted return:

$$J(\pi) = E_{\pi} \left[ \sum_{t=0}^T \gamma^t r_t \right] \quad (3)$$

to accurately capture real-world operational constraints and thermal dynamics, the MDP is defined as follows:

- **State Space (S):** The state vector is an 8-dimensional continuous representation of the furnace environment, defined as  $S_t = [\text{Temperature}, \text{Weight}, \text{Time}, \text{Power}, \text{Status}, \text{Energy\_Consumption}, \text{Scrap\_Added}, \text{Furnace\_Wall\_Temp}]$ . Here, `Scrap_Added` is modeled as an observable event flag in the simulation environment, representing operator-triggered scrap additions known at the decision step in the training setup.

- **Action Space (A):** The action space is discrete, consisting of 10 adjustable power levels ranging from 0 to 450 kW (i.e.,  $A = \{0, 50, 100, 150, 200, 250, 300, 350, 400, 450\}$  kW).

- **Reward Function (R):** Instead of relying solely on sparse terminal rewards, we introduce an expert-guided reward-shaping mechanism. This ensures that the agent satisfies safety protocols, minimizes energy usage, and closely mimics the heating profile established by experienced operators.

The step reward  $r_t$  is defined as the sum of multiple shaping components:

$$r_t = R_{\text{tracking}} + R_{\text{safety}} + R_{\text{progress}} + R_{\text{energy}} + R_{\text{time}} + R_{\text{terminal}} \quad (4)$$

- $R_{\text{tracking}}$ : A penalty for deviating from the expert operational profile, defined as

$$R_{\text{tracking}} = \alpha \cdot \frac{|P_{\text{target}} - P_{\text{expert}}|}{P_{\text{max}}} \quad (5)$$

- $R_{\text{safety}}$ : A strict safety constraint penalty ( $\beta = -5.0$ ) applied if any power is supplied ( $P > 0$ ) during the required alloying and fluxing time window.

- $R_{\text{progress}}$ : A dense reward for continuous temperature progression toward the target, defined as  $+2.0 \cdot \Delta P_{\text{progress}}$ .

- $R_{\text{energy}}$ : An energy consumption penalty forcing the agent to operate efficiently, defined as  $\gamma \cdot E_{\text{used\_kWh}}$ , where  $\gamma = -0.02$ .

- $R_{\text{time}}$ : A minor step penalty ( $-0.001$ ) to encourage faster batch completion.

- $R_{\text{terminal}}$ : A terminal reward given at the end of the episode. If the  $950^\circ\text{C}$  target is successfully reached, a  $+10.0$  bonus is awarded, augmented by an energy efficiency bonus ( $\delta \cdot \text{Normalized\_Efficiency}$ , where  $\delta = 5.0$ ). If the episode terminates by timeout without reaching the target, a penalty of  $-5.0$  is applied.

The relative weighting of these terms was chosen empirically and was further examined through sensitivity analysis and controlled reward-retuning experiments reported in the Results section. In addition, safety-violation tracking was recorded during training and evaluation; while executed violations remained zero because of the environment-level safety override, attempted violations decreased substantially during training, indicating that the policy increasingly internalized the safety rule.

#### 4.2.3 RL Algorithm and Evaluation

We adopt a value-based RL approach suitable for discrete actions, specifically a deep Q-network (DQN), to approximate the action-value function  $Q(s,a)$  and derive the control policy. DQN was selected because the furnace control problem is modeled with a discrete power-action space, making value-based learning a natural fit. In addition, DQN offers implementation simplicity and stable training under limited process data and simulation-based interaction. More advanced continuous-control

algorithms, such as PPO or SAC may also be applicable, but comparative benchmarking across RL algorithms is left for future work. The neural network architecture and hyperparameters were empirically tuned for the high-inertia thermal environment to ensure stable learning, as summarized in Table 1.

**Table 1** Deep Q-network hyperparameters and architecture

| Parameter                       | Value / Description                                        |
|---------------------------------|------------------------------------------------------------|
| Network Architecture            | Multi-Layer Perceptron (MLP)                               |
| Hidden Layers                   | 3 Layers (512 $\rightarrow$ 256 $\rightarrow$ 128 neurons) |
| Activation Function             | ReLU                                                       |
| State Dimension                 | 8                                                          |
| Action Dimension                | 10 (Discrete power levels)                                 |
| Learning Rate                   | 0.0005                                                     |
| Discount Factor ( $\gamma$ )    | 0.95                                                       |
| Exploration Rate ( $\epsilon$ ) | Initial: 1.0, Minimum: 0.01, Decay: 0.995                  |
| Replay Buffer Size              | 10,000 transitions                                         |
| Batch Size                      | 64                                                         |
| Target Network                  | Every 100 steps                                            |
| Update Freq.                    |                                                            |

The policy is evaluated at the batch level using (1) total electrical energy per batch, (2) melting duration, (3) end-of-batch temperature accuracy, and (4) safety-compliance statistics. To improve robustness assessment, repeated evaluations are conducted under multiple start-state conditions, including cold-start, hot-start, and mixed-start scenarios. The revised RL controller is therefore assessed not by a single favorable episode but by repeated-episode averages and variability under a fixed evaluation protocol.

#### 4.3 Genetic Algorithm: Energy-Aware Batch Scheduling

This research introduces the just-in-time (JIT) Gate concept to synchronize batch release with the demand of the downstream melting and holding (M&H) station. This approach minimizes thermal energy loss during holding periods by ensuring that a melt cycle is completed exactly when the M&H inventory requires a refill.

We employ a hybrid genetic algorithm (HGA) in which the GA evolves plant-level scheduling policy parameters, while a greedy decoder converts each chromosome into a feasible operational schedule over a 24-hour simulation horizon. This combination enables the optimization of macro-level policy variables while preserving operational feasibility through simulation-based decoding.

#### 4.3.1 Problem Formulation: Decision Variables, Constraints, and Objectives

The scheduling problem is formulated as a policy-based optimization problem.

**Decision Variables ( $x$ ):** The decision vector represents a set of control policy parameters, such as the material level required to trigger a new melt (trigger fraction), the prediction horizon (lookahead window), and the weighting factor for avoiding high-tariff periods (TOU weight).

**Constraints:** The primary operational constraints include:

1. Inventory Integrity: Aluminum levels in the M&H furnace must remain above the critical threshold and must never reach an empty state.
2. Capacity Limits: Pouring volumes must not exceed the maximum capacity (overflow).
3. Power Limitations: The combined electricity consumption of the induction furnaces and the plant baseline must not exceed the contract demand.

**Objective:** The objective is to minimize a total-cost function composed of time-of-use (TOU) energy cost, demand-related penalties, and service penalties associated with operational violations, such as furnace starvation, overflow, or unmet production demand.

#### 4.3.2 Hybrid Genetic Algorithm Framework

The HGA framework evaluates each candidate scheduling policy through a discrete-event simulator that captures M&H inventory dynamics, induction-furnace availability, tariff conditions, and JIT synchronization behavior.

##### (1) Chromosome Representation

We utilize a real-valued vector encoding for the chromosomes. Instead of encoding fixed timestamps, each chromosome represents a “control policy”

consisting of various heuristic parameters. This allows the GA to evolve optimal rules that are applicable across diverse production scenarios. A chromosome  $x$  may be represented as follows:

$$x = [\text{trig\_a\_frac}, \text{lookahead\_min}, \text{tou\_weight}, \text{jit\_slack\_min}, \dots] \quad (6)$$

where each gene encodes a heuristic control-policy parameter. The complete set of genes, together with their physical meaning and permissible value ranges, is summarized in Table 2.

##### (2) JIT Gate and Greedy Decoder

The core of the HGA is the JIT Gate, which acts as a synchronization barrier during the decoding process:

- Demand & Energy Verification: When an induction furnace (IF) becomes idle, the JIT Gate evaluates the current demand (via  $\text{trig\_frac}$ ) against energy favorability (via  $\text{effective\_price}$ ).
- JIT Deferral: If the M&H inventory is sufficient, the gate defers the start of the melt cycle by a calculated slack period. This ensures that the melt is completed exactly when the M&H station requires a refill, significantly reducing holding time.

**Conflict Resolution:** The decoder employs an  $\text{alternation\_bias}$  to manage simultaneous requests from multiple furnaces, ensuring thermal stability and balanced machine utilization. (3) *Fitness Evaluation*

The fitness of each policy is determined by the simulator over a 24-hour horizon. The fitness value is calculated using a normalized weighted-sum objective:

$$\text{TotalCost} = \sum_i \left( w_i \cdot \frac{C_i}{R_i} \right) + \text{Penalty}_{\text{Service}} \quad (7)$$

where:

- $C_i$  represents specific cost components (e.g., energy cost, demand penalty, holding penalty).

**Table 2** GA chromosome representation and control policy parameters

| Gene                          | Physical Meaning                                                             | Range       |
|-------------------------------|------------------------------------------------------------------------------|-------------|
| $\text{trig\_a\_frac}$        | Threshold level in M&H A to trigger a new melt cycle.                        | [0.0, 1.0]  |
| $\text{trig\_b\_frac}$        | Threshold level in M&H B to trigger a new melt cycle.                        | [0.0, 1.0]  |
| $\text{lookahead\_min}$       | Duration used to forecast upcoming material demand.                          | 0 – 180 min |
| $\text{demand\_headroom\_kw}$ | Reserved power headroom used to reduce the risk of exceeding contract demand | 0–100 kW    |
| $\text{tou\_weight}$          | Importance factor for avoiding On-peak electricity rates.                    | [0.0, 2.0]  |
| $\text{jit\_slack\_min}$      | Maximum allowable delay for JIT synchronization                              | 0 – 60 min  |

---

Algorithm 1: Policy-based GA for Aluminum Smelting Optimization

---

**Input:**  $S_{\text{initial}}$  (initial plant state),  $P_{\text{TOU}}$  (Electricity Tariff Profile),  $x$  (Policy Chromosome)

**Output:**  $x^*$  (Optimal Control Policy)

1. **Initialize** population  $P$  with random policy vectors  $x$
  2. **For each chromosome**  $x$  in  $P$ , decode the policy parameters into scheduling rules.
  3. Simulate plant operation over the 24-hour horizon ( $t = 1, \dots, 1440$  min):
    - update M&H inventory levels,
    - check trigger conditions,
    - compute JIT start timing using demand forecast, batch duration, and slack allowance,
    - execute melting when constraints permit,
    - accumulate cost and penalty terms.
  4. Compute  $\text{Fitness}(x)$  from the normalized cost objective in Eq. (7).
  5. Apply tournament selection, simulated binary crossover (SBX), and polynomial mutation to generate the next population.
  6. Repeat until convergence and return the chromosome  $x^*$  with minimum total cost.
- 

- $R_i$  denotes the reference value used for normalization to bring different units into a comparable scale.

- $\text{Penalty}_{\text{Service}}$  represents a significant scalar penalty applied when hard constraints (e.g., furnace overflow or empty state) are violated, steering the GA toward feasible and safe operational policies.

In this study, the normalized cost components include electricity cost, demand-related penalty, and service-related penalty terms. The reference normalization constants  $R_i$  and weights  $w_i$  are reported explicitly in the revised implementation to improve reproducibility.

#### (4) Hybrid Genetic Algorithm Framework

To search efficiently over the 24-hour evaluation horizon, the HGA uses standard evolutionary operators that balance exploration and exploitation. The hyperparameter configurations are detailed in Table 3.

**Table 3** Hybrid Genetic Algorithm Configuration Parameters

| Parameter                       | Value / Description                           |
|---------------------------------|-----------------------------------------------|
| Population Size                 | 50                                            |
| Maximum Generations             | 100                                           |
| Selection Method                | Tournament Selection<br>(Tournament Size = 3) |
| Crossover Method                | Simulated Binary Crossover<br>(SBX)           |
| Crossover Probability ( $P_c$ ) | 0.9                                           |
| Mutation Method                 | Polynomial Mutation                           |
| Mutation Probability ( $P_m$ )  | 0.1                                           |
| Evaluation Horizon              | 1,440 minutes (24 hours)                      |

#### 4.3.3 Optimization Procedure

To synthesize the components detailed in the previous sections, the overall execution of the Hybrid Genetic Algorithm is structured as a continuous evolutionary loop. The process integrates policy decoding, JIT synchronization, and fitness evaluation against the 24-hour simulation horizon. The complete step-by-step optimization process is summarized in Algorithm 1.

For clarity,  $t_{\text{demand}}$  denotes the predicted time at which M&H inventory reaches the refill threshold, duration denotes the estimated melting duration of one IF batch, and the solar window refers to the time intervals in which on-site solar generation offsets grid electricity demand.

## 5. Results and Discussion

This section presents the experimental results and discusses the performance of the proposed framework at both the batch-control and system-scheduling levels. The analysis is organized into three parts: (1) thermal-model validation as a control-oriented simulation environment, (2) RL-based batch-level power control, and (3) GA-based system-level scheduling.

### 5.1 Thermal Model Validation

To assess whether the lumped-parameter thermal model is sufficiently accurate for RL training and comparative evaluation, it was compared against plant observations from an industrial induction furnace. Because direct temperature measurement is only available after the melt becomes measurable, the validation focuses on the late-stage heating trajectory, beginning at approximately  $t = 70$  minutes and

continuing until the target temperature of 950°C is reached.

#### 5.1.1 Key-Point Temperature Validation

The model was evaluated under two operating scenarios: cold start, representing an initially cooler furnace, and hot start, representing residual heat from the previous batch. Accuracy was quantified using the Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{\text{real},i} - T_{\text{sim},i})^2} \quad (8)$$

For the cold-start scenario (Table 4), the model yielded an RMSE of 12.84°C. Although the maximum pointwise deviation reached 19.1°C at  $t = 80$  minutes, the model predicted the time to reach 950°C

exactly at 96 min. This indicates that, despite local temperature deviations, the model remains temporally consistent for control-oriented use.

In the hot-start scenario (Table 5), model performance improved, with the RMSE decreasing to 8.37°C. This suggests that the residual wall-temperature representation captures the dominant thermal inertia of a preheated furnace more accurately than in the cold-start case.

#### 5.1.2 Heating-Rate Validation

To examine whether the model reproduces the heating dynamics rather than only matching static checkpoints, the heating rate (°C/min) was compared over two-time windows, as shown in Table 6.

**Table 4** Key-point temperature validation (cold-start case)

| Key-point   | Definition                          | Real (°C) | Sim (°C) | Abs Error (°C) |
|-------------|-------------------------------------|-----------|----------|----------------|
| T70         | Temp at $t = 70$ min                | 649.0     | 659.5    | 10.5           |
| T72         | Temp at $t = 72$ min                | 668.0     | 679.0    | 11.0           |
| T74         | Temp at $t = 74$ min                | 685.0     | 673.5    | 11.5           |
| T80         | Temp at $t = 80$ min                | 757.0     | 737.9    | 19.1           |
| T86         | Temp at $t = 86$ min                | 823.0     | 840.3    | 17.3           |
| T90         | Temp at $t = 90$ min                | 881.0     | 894.9    | 13.9           |
| T92         | Temp at $t = 92$ min                | 907.0     | 919.3    | 12.3           |
| T94         | Temp at $t = 94$ min                | 932.0     | 942.2    | 10.2           |
| T96         | Temp at $t = 96$ min                | 956.0     | 953.1    | 2.9            |
| Time to 950 | first time $T \geq 950$ after $t_0$ | 96 min    | 96 min   | 0              |

**Table 5** Key-point temperature validation (hot-start case)

| Key-point   | Definition                          | Real (°C) | Sim (°C) | Abs Error (°C) |
|-------------|-------------------------------------|-----------|----------|----------------|
| T70         | Temp at $t = 70$ min                | 730.0     | 733.9    | 3.9            |
| T72         | Temp at $t = 72$ min                | 773.0     | 777.0    | 4.0            |
| T74         | Temp at $t = 74$ min                | 793.0     | 798.2    | 5.2            |
| T80         | Temp at $t = 80$ min                | 875.0     | 890.2    | 15.2           |
| T86         | Temp at $t = 86$ min                | 940.0     | 949.6    | 9.6            |
| T88         | Temp at $t = 88$ min                | 952.0     | 958.2    | 6.2            |
| Time to 950 | first time $T \geq 950$ after $t_0$ | 88 min    | 88 min   | 0              |

**Table 6** Heating rate comparison (70 – 90 min)

| Scenario   | Time Window (min) | Real slope (°C/min) | Sim slope (°C/min) | Error (°C/min) |
|------------|-------------------|---------------------|--------------------|----------------|
| Cold start | 70 – 80           | 10.80               | 7.84               | −2.96          |
| Cold start | 80 – 90           | 12.40               | 15.70              | +3.30          |
| Hot start  | 70 – 80           | 14.50               | 15.63              | +1.13          |
| Hot start  | 80 – 90           | 7.70                | 6.80               | −0.90          |

In the cold-start case, the model underestimates the heating slope during the 70 – 80 min window and then overestimates it during the 80 – 90 min window. This reversal indicates that the simplified thermal model does not fully capture the nonlinear behavior associated with early cold-furnace heating, including residual phase-change effects and changing heat-loss intensity. In contrast, the hot-start case shows much

closer agreement, with slope errors remaining within approximately  $\pm 1.2^\circ\text{C}/\text{min}$ . This difference is physically reasonable: hot-start batches begin from a thermally conditioned furnace state, so their dynamics are more consistent with the lumped-parameter approximation.

### 5.1.3 Energy and Processing Time Accuracy

Finally, the model was evaluated at the batch level by comparing total energy consumption (kWh) and processing time (min) under different maximum power limits ( $P_{\max}$ ).

The results in Table 7 show that the model maintains energy-estimation error within approximately 5% for cold-start scenarios and below 2% for hot-start scenarios. In addition, time-to-target accuracy is nearly exact in the hot-start cases. Taken together, these results indicate that the thermal model is adequate as a control-oriented simulation environment for RL training and comparative scheduling analysis, while still exhibiting reduced fidelity in the more nonlinear cold-start regime.

## 5.2 Reinforcement Learning-Based Batch Power Control Results

This section evaluates the performance of the RL controller for induction-furnace power regulation. The analysis focuses on training behavior, policy characteristics, and repeated batch-level evaluation under fixed start-state scenarios. Rather than relying on a single favorable run, the controller is assessed using repeated-episode statistics, comparison with the fixed expert-profile baseline, and safety-compliance analysis.

### 5.2.1 Training Performance and Convergence Analysis

The training process was conducted over 1,500 episodes using the validated thermal environment. The convergence behavior of the agent is illustrated using the following metrics:

- **Reward Convergence:** The total reward per episode demonstrates a steady increase, stabilizing after approximately 200 episodes. The Moving Average indicates that the agent successfully learned to balance the expert-tracking penalty with energy-saving rewards.

- **Loss Reduction:** The training loss decreased sharply within the first 100 episodes and remained near zero for the duration of the training, suggesting stable network parameter updates.

**Episode Stability:** The episode length (time to reach  $950^\circ\text{C}$ ) stabilized at approximately 87–95 steps (minutes), while energy consumption converged to a consistent range of 560–590 kWh.

Although individual episodes remain noisy, this variance is expected because the training environment includes stochastic operational events, such as scrap addition and start-state variation. The moving-average trend indicates policy convergence despite this episode-level variability.

### 5.2.2 Policy Behavior and Expert-Guided Logic

The learned policy was guided by an expert reference profile while retaining flexibility to adapt its power trajectory according to the furnace state and reward structure. A key feature of the environment is the safety override during the alloying period (minutes 30 – 31), during which the agent must cut power to 0 kW to allow the safe addition of Si and Fe.

**Table 7** Energy and time validation under power constraints

| Start | $P_{\max}$ (kW) | Energy model (kWh) | Time model (min) | Energy real (kWh) | Time real (min) | $\Delta\text{Energy}\%$ | $\Delta\text{Time}\%$ |
|-------|-----------------|--------------------|------------------|-------------------|-----------------|-------------------------|-----------------------|
| Cold  | 450             | 639.9              | 96               | 614.0             | 92              | +4.22%                  | +4.35%                |
| Cold  | 475             | 641.7              | 92               | 613.0             | 94              | +4.68%                  | -2.13%                |
| Cold  | 500             | 623.0              | 88               | N/A               | N/A             | —                       | —                     |
| Hot   | 475             | 575.8              | 88               | 576.0             | 88              | -0.03%                  | 0.00%                 |
| Hot   | 500             | 582.5              | 86               | 579.0             | 86              | +0.60%                  | 0.00%                 |
| Hot   | 450             | 582.5              | 86               | 579.0             | 86              | +0.60%                  | 0.00%                 |

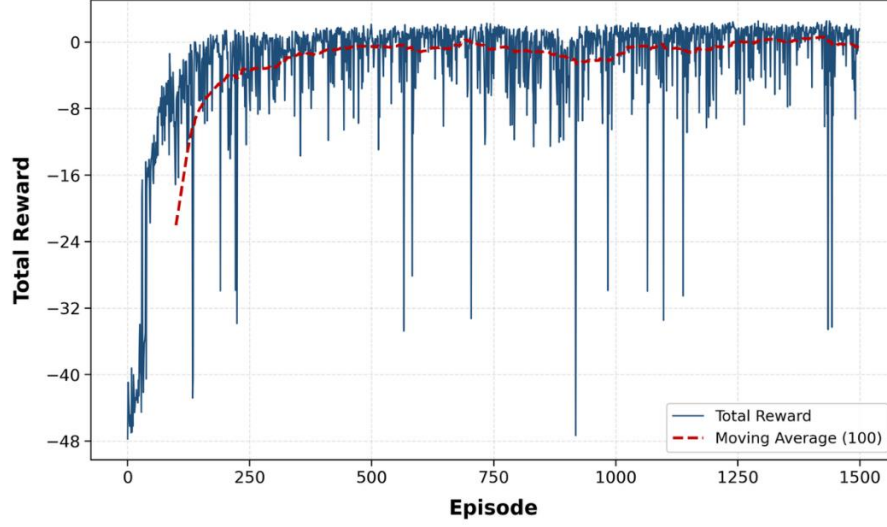


Figure 4 Training reward convergence with moving average (100 episodes)

Table 8 Action distribution of the evaluated agent (Episode 1500)

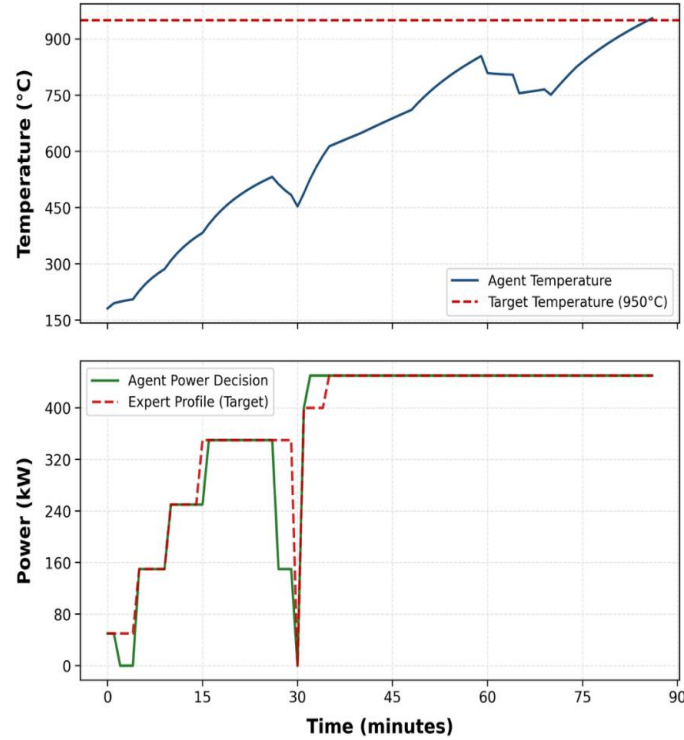
| Power Level (kW) | Frequency (%) |
|------------------|---------------|
| 0 (Off/Alloying) | 4.6%          |
| 50–250 (Low/Mid) | 18.4%         |
| 350–400 (High)   | 13.7%         |
| 450 (Maximum)    | 63.2%         |

As shown in Table 8, the agent predominantly utilizes the maximum power (450 kW) to reduce total processing time, which in turn minimizes cumulative thermal losses. However, the agent also selectively uses mid-range power levels to optimize the temperature ramp-up according to the reward function's efficiency component. In addition, safety analysis across training checkpoints showed that attempted violations during the alloying window decreased substantially as training progressed, while executed violations remained zero because of the environment-level safety override. This indicates that the agent increasingly internalized the safety rule rather than merely relying on hard enforcement.

### 5.2.3 Batch-Level Evaluation and Melt Profile Analysis

A single favorable episode can be informative for illustrating policy behavior, but it is not sufficient for manuscript-level performance claims. For this reason, the RL controller was evaluated under repeated episodes and multiple start-state scenarios. Under the mixed-start evaluation protocol (80% hot-start, 20% cold-start), the original DQN achieved  $600.79 \pm 19.96$  kWh with a mean duration of  $92.1 \pm 2.9$  min, while the revised RL controller achieved  $597.96 \pm 18.23$  kWh with a mean duration of  $90.8 \pm 2.6$  min. The revised controller preserved 100% target attainment and zero executed safety violations.

Compared with the fixed expert-profile baseline under the same protocol, the revised RL controller showed a slight energy advantage (597.96 vs. 601.73 kWh) while maintaining a similar batch duration and stable target attainment. These results indicate that the revised RL controller achieved a small but reproducible improvement over both the original DQN and the fixed expert-profile policy under the same evaluation conditions.



**Figure 5** Comparison between agent power decisions and the expert target profile

Analysis of the melting profile in Figure 5 reveals several critical behaviors:

1. **Safety Compliance:** The agent strictly follows the power-down command at the 30-minute mark, ensuring operational safety during material addition.

2. **Precise Tracking:** The temperature curve shows a steady progression with minimal overshoot, reaching the target accurately.

3. **Batch-level Performance:** Compared with the original DQN under the same mixed-start protocol, the revised RL controller reduced average batch energy by 2.83 kWh and reduced average duration by 1.3 min. Compared with the fixed expert-profile baseline, it achieved a slight energy improvement while maintaining similar control behavior.

4. **Robustness to Material Addition:** The noticeable temperature drops observed between the 60 and 75-minute marks represent the system's thermal response to the random addition of scrap metal (a total of 150 kg added across 3 separate instances). The RL agent demonstrates highly robust control by effectively recovering from these sudden thermal disturbances and smoothly resuming its heating trajectory toward the target.

Overall, the RL results support the interpretation that the controller provides robust, near-expert batch-

level power control with a slight efficiency improvement, rather than a large standalone energy-saving effect. The major system-level energy savings of the overall framework are produced by the GA-based scheduling layer discussed in the next section.

### 5.3 Genetic Algorithm Scheduling Results

While the RL agent optimizes power control at the batch level, the GA framework addresses system-level efficiency by synchronizing the melting cycles of multiple induction furnaces (IFs) with the downstream demand of melt-and-hold (M&H) stations. This section evaluates the GA-based scheduling approach against two comparators: a plant-inspired continuous baseline and a rule-based heuristic. In addition to single-scenario performance, repeated-run analysis was used to confirm the robustness of the GA results under stochastic optimization.

#### 5.3.1 Comparative Analysis of Scheduling Policies

The performance of different scheduling policies was evaluated over a 24-hour production horizon. The primary objectives were to minimize total operational costs, eliminate reheat/holding losses, and maintain a stable electrical load.

**Table 9** System-level performance comparison of scheduling policies

| Metric              | continuous baseline | Rule-based         | GA-Energy (Proposed) |
|---------------------|---------------------|--------------------|----------------------|
| total cost*         | 806,906             | 491,811            | 556,474              |
| energy cost (THB)   | 39,580              | 28,550             | 28,253               |
| Reheat Energy (kWh) | 2,833.30            | 1,108.30           | 0                    |
| Holding Time (min)  | 340                 | 140                | 0                    |
| Peak Demand (kW)    | 1,770.40            | 1,745.40           | 1,346.20             |
| Successful Batches  | 12/12               | 10/12              | 12/12                |
| Service Violations  | 0                   | <b>2 (Missing)</b> | 0                    |

Table 9 presents the principal system-level comparison across scheduling policies. Although total cost is the scalar objective used by the optimizer, it should be interpreted together with service feasibility and throughput-related metrics.

Although the rule-based approach exhibits a lower total cost (491,811), this is an artificial reduction resulting from its failure to meet the required production demand. Specifically, the Rule-based policy successfully completed only 10 of the 12 required batches, leading to two service violations (missing batches). By melting less than the necessary volume, the system inherently consumed less total energy, which skewed its objective score downward but failed to meet the fundamental production requirement.

In contrast, the GA-Energy policy registered a Total Cost of 556,474 while successfully fulfilling 100% of the production demand (12/12 batches) without any service violations. When examining the actual operational metrics for a fully completed schedule, the GA-Energy framework demonstrates superior real-world efficiency: it completely eliminated reheat energy and holding time and achieved the lowest actual Energy Cost (28,253 THB) alongside significantly reduced peak demand (1,346.20 kW). Therefore, the GA-Energy policy

provides a highly robust, cost-effective schedule that fully satisfies industrial production constraints.

To assess reproducibility, the GA was also executed across repeated independent runs. These experiments showed low variance in the objective value and no missing batches across repeated feasible runs, supporting the stability of the GA solution rather than a single favorable optimization outcome.

### 5.3.2 Elimination of Reheat and Holding Losses

A critical finding of this study is the total elimination of reheat and holding losses through the Just-in-Time (JIT) Gate logic within the GA framework.

- **Baseline Inefficiency:** In the Continuous Baseline (Figure 6), furnaces operate on a fixed cycle regardless of M&H levels, resulting in 2,833.3 kWh of energy wasted on reheating molten metal that reached the target temperature too early. This is visualized by the red blocks in the IF Gantt chart.

- **GA Optimization:** The GA-Energy policy (Figure 7) successfully eliminated these red blocks, achieving 0.0 kWh of reheat loss. By evolving optimal "trigger fractions" and "JIT slack" parameters, the GA ensures that a melt cycle completes exactly when the M&H inventory requires a refill, thus removing unnecessary thermal holding.

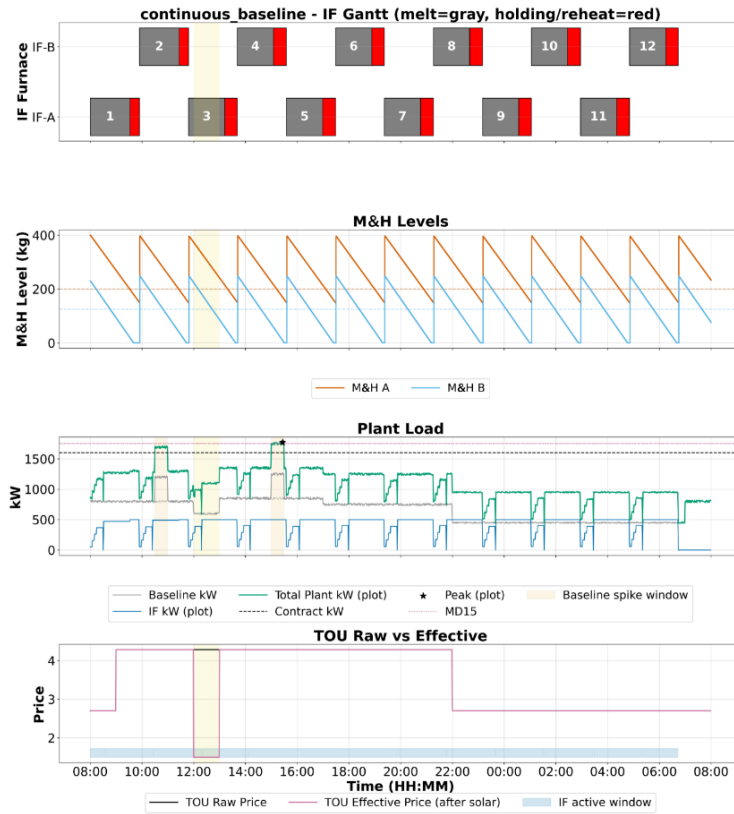


Figure 6 Baseline schedule showing significant holding/reheat losses (red blocks).

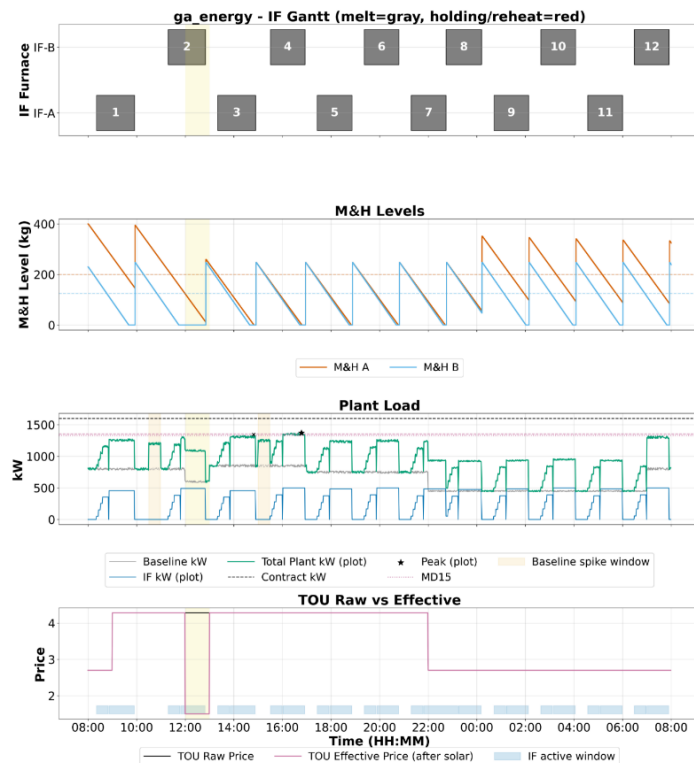


Figure 7 GA-optimized schedule showing the elimination of reheat losses and synchronized M&H levels

### 5.3.3 Peak Load Shaving and Electrical Stability

The GA-based approach also demonstrates superior performance in managing the plant's electrical load.

- **Peak Demand Reduction:** The GA-Energy policy reduced the peak power demand from 1,770.4 kW (Baseline) to 1,346.2 kW, a reduction of approximately 23.9%. This reduction is vital for industrial sites to stay within contract demand limits and avoid expensive demand charges.

- **Load Profiling:** As seen in the "Plant Load" subplots, the GA policy effectively "shaves" the peaks by avoiding the simultaneous operation of high-power auxiliary systems and furnace cycles during peak hours.

- **Robustness:** Unlike the rule-based logic, which failed to deliver 2 batches and caused inventory deficits (missing batches = 2), the GA policy maintained a 100% throughput rate (12 batches) while strictly adhering to inventory and power constraints. In addition, runtime measurements showed that one 24-hour GA optimization required approximately 68 s on the test machine, indicating practical feasibility for day-ahead or intraday re-optimization.

### 5.3.4 Policy Robustness (GA-Service vs. GA-Energy)

The GA was also tested under a "GA-Service" mode, which prioritizes inventory safety over cost. While this mode maintains similar energy efficiency to GA-Energy, it results in higher peak loads to ensure that M&H levels never approach critical thresholds. This flexibility allows plant managers to switch policies based on daily production priorities, prioritizing cost during low-demand periods or service stability during high-volume production.

Repeated-run analysis further showed that both GA-Service and GA-Energy maintained very low variance across independent seeds, confirming that the observed advantages are not attributable to a single stochastic run.

## 5.4 Integrated Discussion: Synergy Between RL-Based Control and GA-Based Scheduling

The proposed framework adopts a two-layer approach to improving energy performance in the aluminum melting process. Rather than assigning the same role to both optimization layers, the results show that RL and GA contribute at different decision scales: RL improves batch-level power control consistency, while GA delivers the major system-level savings through scheduling and synchronization.

### 5.4.1 Micro-Level vs. Macro-Level Impact

The integration of RL and GA bridges the gap between furnace-level thermal dynamics and plant-level production timing:

- **Batch-Level Control (RL):** The revised RL controller achieved robust target attainment, zero executed safety violations, and a slight improvement over both the original DQN and the fixed expert-profile baseline under the mixed-start evaluation protocol. This indicates that RL is valuable as a near-expert, state-aware power-control policy rather than as the sole source of large energy savings.

- **System-Level Synchronization (GA):** Even an efficient batch-level controller can create unnecessary holding or reheating if melting completion is poorly synchronized with downstream demand. The GA-based scheduler addresses this problem through the JIT Gate mechanism, ensuring that batches are released only when the M&H stations are ready to receive them.

### 5.4.2 Holistic Energy Savings

The combined framework improves energy performance through two complementary mechanisms:

1. **Batch-level Control Benefit:** The revised RL controller provided a small but reproducible batch-level energy improvement under the fixed mixed-start evaluation protocol, while preserving 100% success and safe operation.

2. **Eliminated Waste:** The GA eliminated the 2,833.3 kWh of reheat energy found in the continuous baseline, representing a 100% reduction in avoidable thermal waste.

3. **Peak Load Management:** The integrated framework reduced the peak electrical load from 1,770.4 kW to 1,346.2 kW. This "peak shaving" capability is critical for avoiding high-demand charges and maintaining grid stability in industrial settings.

Taken together, these findings indicate that the major system-level energy-saving benefit of the framework arises from GA-based synchronization, while RL contributes by making individual melting batches more consistent, safe, and slightly more efficient.

### 5.4.3 Practical Implications for Industry

The results indicate that the RL+GA framework provides a practical decision-support architecture for high-temperature manufacturing processes. The RL controller helps standardize batch-level power regulation under safety constraints, while

the GA-based scheduler ensures that these batches are delivered at operationally appropriate times. This synchronized approach reduces avoidable holding and reheating losses, lowers peak demand, and supports more stable plant operation.

Several practical limitations should also be noted. First, the thermal model is a control-oriented surrogate rather than a high-fidelity thermophysical model, and its accuracy is lower in the cold-start regime where nonlinear effects are more pronounced. Second, temperature measurements are unavailable during the early melting stage, so the controller must rely on estimated thermal states rather than direct sensing. Third, the RL policy was trained within a bounded operating regime; therefore, deployment under unseen alloy compositions, extreme ambient conditions, or furnace degradation should be approached cautiously.

Although the framework was evaluated on an aluminum induction-furnace system, the overall architecture is potentially transferable to other energy-intensive thermal processes, provided that the RL surrogate model, reward design, and GA scheduling constraints are recalibrated for the target facility. In addition, the computational burden appears practical for industrial use: GA optimization required approximately 68 s for a daily schedule, while the trained RL policy can be executed online at the batch-control level. These characteristics support the use of the framework for day-ahead planning and operational decision support.

## 6. Conclusion

This study presented an integrated RL+GA framework for improving energy performance in an aluminum melting process and evaluated it at both the batch-control and system-scheduling levels. The main conclusions, organized according to the research objectives, are as follows:

1. System Formulation: A system-level model of the IF–M&H production line was established, explicitly incorporating batch-level quality constraints, M&H inventory dynamics, holding/reheat losses, and shared electrical-demand limitations.

2. Batch-level Control: The revised RL controller achieved robust batch-level power control with 100% target attainment and zero executed safety violations. Under the mixed-start evaluation protocol, it slightly improved upon both the original DQN and the fixed expert-profile baseline, indicating that expert-guided RL can provide near-expert and reproducible batch-control performance.

3. System-level Scheduling: The GA-based scheduler, supported by the JIT Gate mechanism, eliminated reheat and holding losses in the evaluated scenario, thereby removing a major source of avoidable thermal waste.

4. Benchmarking and Operational Performance: Compared with the continuous baseline, the proposed framework reduced peak demand by 23.9% and maintained full throughput, while also outperforming the rule-based heuristic in terms of service feasibility.

5. Overall, the findings indicate that the major system-level energy-saving benefit arises from GA-based synchronization, while RL contributes robust and near-expert batch-level control under safety constraints. Future work should focus on richer sensing for early-stage melt observation, broader validation under out-of-distribution operating conditions, and extension of the framework to larger furnace networks, potentially including multi-agent RL and real-time renewable-energy forecasting.

## 7. Code and Data Availability

The code and processed datasets supporting the findings of this study are available from the corresponding author upon reasonable request. The raw industrial operating data used in this study are not publicly available due to confidentiality restrictions with the industrial partner.

## 8. Acknowledgements

The author would like to express their sincere gratitude to FEDERAL ELECTRIC CORP., LTD., the manufacturer and distributor of Sharp home appliances in Thailand, for providing access to its industrial facilities and for its invaluable support in making this research possible. The opportunity to study real-world production data and operational constraints was fundamental to the development and validation of the proposed framework.

## 9. Abbreviations

| Abbreviation | Full Term                                  |
|--------------|--------------------------------------------|
| IF           | Induction Furnace                          |
| M&H          | Melt-and-Hold Furnace                      |
| GA           | Genetic Algorithm                          |
| HGA          | Hybrid Genetic Algorithm                   |
| NSGA-II      | Non-dominated sorting genetic algorithm II |
| RL           | Reinforcement learning                     |
| DQN          | Deep Q-Network                             |
| MDP          | Markov decision process                    |
| JIT          | Just-In-Time                               |

| Abbreviation | Full Term                  |
|--------------|----------------------------|
| TOU          | Time-of-Use                |
| kW           | Kilowatt                   |
| kWh          | Kilowatt-hour              |
| MD           | Maximum Demand             |
| MD15         | 15-minute Maximum Demand   |
| DC           | Demand Charge              |
| RMSE         | Root mean square error     |
| MLP          | Multi-Layer Perceptron     |
| SBX          | Simulated Binary Crossover |

## 10. CRediT Statement

**Kraiwich Pawadee:** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft.

**Narongdech Keeratipranon:** Conceptualization, Supervision, Validation, Writing – Review & Editing, Project Administration.

## 11. References

- Ahmad Khan, S., Rehman, A. U., Arshad, A., Malik, F. H., & Ayadi, W. (2026). Metaheuristic optimizer-based segregated load scheduling approach for household energy consumption management. *Eng.* 7(2), Article 65. <https://doi.org/10.3390/eng7020065>
- Braga, C. A. P., & da Fonseca Netto, J. V. (2016). A dynamic state observer to control the energy consumption in aluminium production cells. *Systems Science & Control Engineering*, 4(1), 307-319. <https://doi.org/10.1080/21642583.2016.1238325>
- Chen, L., Cao, L. L., Wen, Y. M., Chen, H., & Jiang, S. L. (2023). A knowledge-based NSGA-II algorithm for multi-objective hot rolling production scheduling under flexible time-of-use electricity pricing. *Journal of Manufacturing Systems*, 69, 255-270. <https://doi.org/10.1016/j.jmsy.2023.06.009>
- Cui, W., & Yuan, B. (2024). A hybrid genetic algorithm based on reinforcement learning for the energy-aware production scheduling in the photovoltaic glass industry. *Computers & Operations Research*, 163, Article 106521. <https://doi.org/10.1016/j.cor.2023.106521>
- Dai, J., Liu, H., Zhang, Y., Shi, H., & Ling, P. (2025). An optimization design of energy consumption for aluminum smelting based on a multi-objective artificial vulture algorithm. *Metals*, 15(2), Article 105. <https://doi.org/10.3390/met15020105>
- Fernandes, J. M. R. C., Homayouni, S. M., & Fontes, D. B. M. M. (2022). Energy-efficient scheduling in job shop manufacturing systems: A literature review. *Sustainability*, 14(10), Article 6264. <https://doi.org/10.3390/su14106264>
- Gao, K., Huang, Y., Sadollah, A., & Wang, L. (2020). A review of energy-efficient scheduling in intelligent production systems. *Complex & Intelligent Systems*, 6(2), 237-249. <https://doi.org/10.1007/s40747-019-00122-6>
- Li, C., Chang, Q., & Fan, H. T. (2024). Multi-agent reinforcement learning for integrated manufacturing system-process control. *Journal of Manufacturing Systems*, 76, 585-598. <https://doi.org/10.1016/j.jmsy.2024.08.021>
- Loffredo, A., May, M. C., & Matta, A. (2023). *Reinforcement learning for energy-efficient control of multi-stage production lines with parallel machine workstations* [Conference presentation]. Italian Manufacturing Association Conference: XVI AITeM, Naples, Italy. <https://doi.org/10.21741/9781644902714-51>
- Lu, R., Li, Y. C., Li, Y., Jiang, J., & Ding, Y. (2020). Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management. *Applied Energy*, 276, Article 115473. <https://doi.org/10.1016/j.apenergy.2020.115473>
- Modrak, V., Sudhakarapandian, R., Balamurugan, A., & Soltysova, Z. (2024). A review on reinforcement learning in production scheduling: An inferential perspective. *Algorithms*, 17(8), Article 343. <https://doi.org/10.3390/a17080343>
- Nayak, A., Fang, K., & Lee, S. (2017). Demand response in flow shop with job due dates using genetic algorithm approach. *Smart and Sustainable Manufacturing Systems*, 1(1), 100-120. <https://doi.org/10.1520/SSMS20160006>
- Park, J. S., Ng, H. Y., Chua, T. J., Ng, Y. T., & Kim, J. W. (2021). Unified genetic algorithm approach for solving flexible job-shop scheduling problem. *Applied Sciences*, 11(14), Article 6454. <https://doi.org/10.3390/app11146454>
- Salido, M. A., Escamilla, J., Giret, A., & Barber, F. (2016). A genetic algorithm for energy-efficiency in job-shop scheduling. *The International Journal of Advanced*

- Manufacturing Technology*, 85(5), 1303-1314.  
<https://doi.org/10.1007/s00170-015-7987-0>
- Wong, C. J., Bao, J., Skyllas-Kazacos, M., Welch, B., Mahmoud, M., Arkhipov, A., & Ahli, N. (2023). Studies on power modulation of aluminum smelting cells based on a discretized mass and thermal dynamic model. *Metallurgical and Materials Transactions B*, 54(2), 562-577.  
<https://doi.org/10.1007/s11663-022-02709-w>